How to Create a Scalable and Automated Edge Strategy in the AWS Cloud

Dave Shackleford
The use cases for cloud services continue to expand rapidly, and organizations are realizing that many types of access scenarios are shifting as they move toward more PaaS use, cloud infrastructure deployments and changing interconnectivity models between users, remote offices and data centers, they are realizing that many types of access scenarios are shifting. Whether oriented toward end user access to applications and services, or traditional data center access to branch offices and other remote locations, the need to make traditional data centers the hub of connectivity is rapidly becoming a challenge. And the rapid maturation of cloud is driving a convergence of numerous elements of cloud services and security into a unified fabric.

The cloud as an edge service is oriented toward network access, control and architecture. Software-defined networking and security have come to include cloud-based networking platforms that allow for interconnectivity between on-premises environments and cloud provider infrastructure through a singular backbone service or vendor solution. These networking services often provide common networking capabilities, such as routing, bandwidth-shaping and quality of service (QoS), and core content delivery network (CDN) services that can set priorities on specific content and service access and transmission.

Cloud security-as-a-service (SecaaS) is the second convergence category in the cloud network edge. This extremely broad category includes services such as content filtering, malware detection and response, and traffic behavioral monitoring. In addition, the cloud network edge brings together offerings such as VPN connectivity, web application firewall (WAF) filtering, traditional firewall filtering, network intrusion detection and prevention, and others.
This newly defined cloud networking and security convergence will lead to a shift in how organizations design networks altogether. In a cloud edge model, cloud provider environments can now replace the traditional data center as the primary connectivity and traffic processing locations for end users, data center services and platforms, and IoT and other distinct devices through a combined networking and security fabric. Networking and security teams will jointly define and administer this model, likely with input from mobile, application development and systems administration teams as well.

In this paper, we will explore how to redefine and approach defense at the security perimeter. We will share best practices and real-world use cases to develop a layered control approach to perimeter security, implement a scalable security solution at the network’s edge and improve efficiency by automating manual security processes.

Protecting Applications in the Cloud

One of the major use cases for adopting cloud edge services and tools is improved and better-integrated protection for applications running in the cloud. In the cloud, where applications are often deeply integrated into numerous cloud fabric services, employing security and networking controls that are closer to the applications and integrate more natively with the cloud provider environment makes much more sense. In the following sections, we discuss using WAF controls and cloud-native services to secure applications in the cloud.

Web Application Firewalls

The first major control that can help to protect application workloads and access in the cloud are WAFs, which filter and monitor traffic to and from web application infrastructure. These tools are similar to application proxies in many ways, focusing only on web app traffic and applying application-layer filtering and rulesets to prevent security events. WAFs are often used to detect and block common web app events, such as cross-site scripting (XSS), SQL injection, command injection and directory traversal. Mature WAFs come with a starting ruleset that includes standard blocking and detection rules for events of these types, although many organizations modify these rules and add their own. More and more, organizations are using WAFs to “fingerprint” application traffic and identify anomalous behavior patterns. For example, an unexpected number of transaction attempts by a specific user could be considered atypical or certain users clicking links that they normally wouldn’t could be out of character. When setting up WAFs to perform this type of monitoring and filtering, it’s critical to involve application developers in order to reduce false positives and ensure that the monitoring effort is as accurate as possible.
AWS WAF is natively integrated into the Amazon Web Services (AWS) fabric and can protect web applications or APIs against common web exploits that might affect availability, impact security or consume excessive application resources. AWS WAF makes it relatively simple to create security rules that block potential events such as SQL injection, XSS and authentication attempts. In addition, AWS WAF supports custom rules that filter out specific traffic patterns that you define. Amazon provides a preconfigured set of managed rules for the AWS WAF, offered by AWS or AWS Marketplace Sellers. These managed rules address many common issues such as those in the OWASP Top 10. AWS WAF also includes a robust API that can help to automate the creation, deployment and maintenance of application security rules. AWS WAF monitors web traffic and can automate creation of new rules or alerts in Amazon CloudWatch (with granular control over Amazon CloudWatch metrics). In addition, AWS WAF offers comprehensive logging by capturing each inspected web request’s full header data for use in security automation and analytics or for the purpose of auditing. Many well-known third-party WAF platforms are available in AWS Marketplace, as well, making it easy to integrate these tools into the application infrastructure.

Cloud-Native Services

In addition to WAF controls, organizations can use numerous cloud-native services to protect their applications in the cloud. These include the following:

- Identity and access management (IAM)
- AWS Organizations
- Cloud-native network access controls
- Load balancing

We describe each of these services in the following sections.

Identity and Access Management (IAM)

IAM is one of the most important aspects of cloud security today. IAM is pivotal to controlling who and what can access application resources in the cloud. Defining roles, enabling strict access models and limiting the resources available to users and systems are critical steps in enabling a sound cloud security strategy overall. IAM roles are associated with privileges and permissions for making API calls to interact with cloud services and only exist within the cloud environment itself. New IAM accounts have no permissions (an implicit deny all policy), and all permissions must be explicitly granted. This approach can also help to prevent over-allocating privileges to services in the environment. Because IAM policies can be assigned to many assets and resources, IAM service accounts can be enabled for application access to AWS resources very flexibly.

WAFs are becoming a more critically important control in protecting web applications in the cloud, especially with increasing exposure of public-facing services. Having WAF products and services that are easy to deploy and deeply integrated into the cloud provider environment can greatly simplify operations for security and cloud engineering teams.

---

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.
AWS Organizations

AWS Organizations is a service that organizations can use to define policies to apply across multiple AWS accounts from a master control level. With AWS Organizations, it’s possible to 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 various BU accounts might 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).

Single Sign-On (SSO)

As more organizations look to unify and simplify their end-user access to cloud and web services, the use of federation to synchronize user directories with identity-as-a-service (IDaaS) and identity brokering solutions grows in importance, because this strategy can help to simplify and centralize user authentication in a single portal, enable strong multifactor authentication, and selectively associated users and user attributes with defined roles and privileges within a growing list of cloud services and applications. AWS offers an integrated single sign-on application, called AWS SSO. This service integrates with AWS Organizations or AWS Directory Service to accommodate Active Directory accounts or other federated user assertions, and then provides a portal providing identity brokering to AWS services, custom SAML apps and common SaaS apps.

Cloud-Native Network Access Controls

Cloud-native network access controls can help prevent bad actors from using unapproved network connections to access systems, moving laterally from an application or system to which unauthorized access has been gained, or performing any illegal network activity regardless of environment. The first focal area for any cloud-native network isolation and segmentation should be the core network zones associated with cloud accounts. In AWS, these are virtual private clouds (VPCs), and they can contain any number of distinct network subnets. VPCs can also be peered to one another and connected through AWS Transit Gateways and AWS Direct Connect circuits. Subnets within each VPC can be configured to communicate as needed through routing and cloud-native access controls.

Cloud-native access controls can be created and applied within the VPC and should be used to isolate and control traffic flow into the VPC subnets altogether, as well as to and from instance workloads running applications and services. Within AWS, there are two built-in types of network access and isolation controls: security groups and network access control lists (network ACLs or NACLs). Security groups and network ACLs can be used to control traffic into and out of network deployments. Security groups apply to instance workloads and are stateful, while network ACLs apply to VPC subnets and are stateless. Security groups start with a network access control policy of deny all, and enterprises can then add rules to allow only those types of network access needed.
Load Balancing

Load balancing is a critical element of network designs today. Load balancers aid in ensuring that availability and resiliency goals are met for all network and application traffic throughout the global cloud ecosystem. AWS Elastic Load Balancing distributes incoming app traffic across multiple Amazon Elastic Compute Cloud (EC2) instances. Cloud-native load balancers are much more capable and flexible than ever. Cloud providers’ native load balancers can route traffic based on simple application or network information, and these network-oriented approaches work best for standard network traffic or cloud environments that have more traditional application deployments. These are also good for internal load balancers distributing traffic to storage nodes.

Platforms such as AWS Elastic Load Balancing can also route traffic based on advanced application information that includes the content of the request and more granular microservices architecture. This is the preferred type of load balancer, especially for internet-facing and web application traffic. Tools such as AWS Elastic Load Balancing can also establish HTTPS sessions with clients, making the service highly valuable for mobile access and any secure data transmission. Because many web applications move to HTTPS by default, this becomes more and more relevant. You can easily upload your own certificates to cloud load balancers or use certificates from a cloud-native certificate authority (CA).

Between third-party products that are available in AWS Marketplace and the native controls and services that AWS offers, application environments can be readily provisioned, secured and monitored entirely within cloud edge environments.

Redefining the Security Perimeter for the Cloud

As the cloud emerges as a new edge computing environment, the classic concepts of a security perimeter have to change. In the core network, for internal users and services that need to access cloud services, cloud edge might replace cloud access security broker (CASB) solutions and other brokering options to access cloud securely. For newer and more specific use cases, IoT and edge network scenarios can be isolated and connected through network restricted policies assigned within a cloud edge brokering platform or service, or in tandem with software-defined wide area networking (SD-WAN) vendors.

Defense-in-Depth

The use of existing network perimeter technologies when defending against distributed denial-of-service (DDoS) attempts creates a challenge. Many on-premises tools and controls need greater capabilities to wholly protect applications and network services today, and the variety and types of security events are changing as well. While many attempts are still volume-based (primarily SYN floods and ICMP and UDP traffic), we are seeing more and more application-level traffic (primarily HTTP, HTTPS and DNS queries).
Some of these are much “slower” in nature and focus more on connection handling at
the application/service layer than pure volume. In addition, many DDoS attempts now
target stateful network devices, looking to fill connection queues and cause slowdown and
loss of availability. Additionally, we are seeing new types of illegal activity related to DDoS
events. For example, they might be used as a distraction mechanism while other events,
such as data exfiltration and privilege escalation attempts, are underway, making the need
to defend against DDoS efficiently and effectively even more pronounced.

**Protections Against DDoS**

Cloud-based DDoS defense products and services offer a variety of protection techniques
for enterprises. Common capabilities include:

- **Overprovisioning**—Overprovisioning is the ability to “open the pipes” and allow for
  far more traffic than usual. Many providers accomplish this by bursting traffic when
  needed to prevent the additional traffic from making its way to end-user customers.

- **“Clean pipe”/packet scrubbing**—Because cloud-based DDoS services will see traffic
  bound for customers before customers see it, these services provide an opportunity
  to filter anomalous traffic and replace anomalous attributes with “normal” ones,
  such as generic HTTP headers instead of those involved with bots.

- **DNS redirection and Domain Name System Security Extensions (DNSSEC)**—More
  and more DDoS attacks are taking the form of DNS requests, particularly extended
  DNS (EDNS) requests and responses of an extremely large size (responses can be
  up to 4,000 bytes in length). By controlling DNS requests that reach the target, and
  by limiting types and size of requests and responses, cloud-based DDoS services
  can prevent both unauthorized requests and reflected (or response-based) DDoS
  attempts. The use of DNSSEC, which allows for cryptographic signature of DNS
  requests and responses, is also emerging as a potential means to defend against
  these types of DNS-based threats.

- **User and source prioritization**—Many source address blocks and specific user
  organizations and individuals might reside on known blacklists or watchlists,
  enabling cloud providers to leverage reputation-based filtering and prioritization to
  defend against DDoS attempts.

- **Geographic blocking**—Often coupled with reputation-based filtering, filtering traffic
  based on geographic origin is a common technique that cloud providers and ISPs
  employ to effectively address a variety of threats, including DDoS attempts.

- **WAF/filtering**—For application-layer attacks, such as “low and slow” attacks
  that cause target web servers and services to hang, WAFs and filtering methods
  can prove effective at analyzing both protocol anomalies and attack signatures
  generated by automated tools.
• **Caching and redundancy**—Caching application and site content, as well as using redundant servers, sites and application instances, allows for a higher degree of availability. These techniques are commonly used to meet availability best practices regardless of the threat of DDoS, but cloud providers will often scale sites and services across numerous data centers to better balance load and potentially stay one step ahead of targeted DDoS traffic.

• **Flow monitoring**—One of the more recent varieties of DDoS events to emerge focuses on network platforms. With these types of events, DDoS traffic is generated that leaves numerous “half-open” TCP connections resident in firewall, router and IDS/IPS state tables. Depending on the type and configuration of these devices, they might become overwhelmed, throttling traffic and acting as bottlenecks that cause significant downtime and delays. Cloud-based DDoS services can monitor network flow data to determine whether typical traffic patterns are being seen, or whether anomalous “single direction” flows are taking place that could be the onset of a protocol attack targeting state tables.

• **Proxying**—Cloud-based DDoS providers and services can also act as a shield for customers’ sites and services by proxying their traffic in both directions. This is common today with some ISPs, and cloud providers can now also act as an intermediary for customers, in essence “hiding” their real workloads from the internet.

Using a cloud-based DDoS defense provider may prove to be an effective security control for preventing, detecting and responding to DDoS attempts, whether or not your organization has on-premises protection. For example, AWS Shield is a managed DDoS protection service that can help to protect applications running in the AWS cloud. AWS Shield provides always-on detection and automatic inline mitigations that minimize application downtime and latency. For more robust DDoS protection against attempts that target applications running on Amazon EC2, AWS Elastic Load Balancing, Amazon CloudFront, AWS Global Accelerator and Amazon Route 53 resources, AWS offers AWS Shield Advanced. In addition to the network and transport-layer protections that come with AWS Shield Standard (applicable to all AWS customers), AWS Shield Advanced provides additional detection and mitigation against large and sophisticated DDoS attempts. AWS Shield Advanced also provides real-time visibility into events along with integration with AWS WAF. AWS Shield Advanced also gives customers 24x7 access to the AWS DDoS Response Team (DRT) and protection against DDoS-related spikes in numerous AWS resources.

In order to develop and implement a robust cloud edge security strategy, a technology stack and architecture should include a defense-in-depth set of controls that help to achieve the goals of access control, network traffic protection, and availability and redundancy to meet performance requirements. Ensuring strong access controls and application-level attack prevention and detection is also critical in a best-in-class network security design.
Automation and Scalability

With the move toward cloud-integrated edge services and controls, automation is another key concept and skill set that security teams need for cloud security enablement. Information security needs to become a lot more automated and embedded in network access control and application protection policies now and in the future.

Security must be able to scale with business operations in the cloud, which means it should be tightly coupled to application components and workloads. As a simple example, imagine a case where a business needs to rapidly increase the number of application instances running in a cloud environment to support its business initiatives. Ideally, the network edge services would automatically scale at the same time, and dynamic policy updates for WAF, firewall, DDoS protection and other elements of a multilayered security stack would simultaneously be triggered as well.

Another example would be automated quarantine of an application workload upon detection of anomalous behavior. To facilitate network and application traffic isolation, some type of automated detection and response cycle would need to be initiated based on indicators of compromise (IoCs) or anomalous behavior that isolates the system at the host or virtual/cloud orchestration level, changing its network security groups designation or modifying WAF policies to limit access.

A number of cloud-native services can assist with automating large-scale deployments that make use of isolation and other cloud-native security control policies. Two of these services are:

• **AWS Control Tower**—This service simplifies the setup of new multi-account AWS environments and provides effective account and security governance at scale. Cloud administrators can provision an automated landing zone that deploys best practices blueprints and templates that can establish multi-account IAM policies with AWS Organizations, enable account provisioning with the AWS Service Catalog and create a centralized log store with services such as AWS CloudTrail and AWS Config.

• **AWS Firewall Manager**—This service enables security teams to centrally configure and manage numerous types of firewall rules for accounts and applications. As new applications are created, AWS Firewall Manager can automate enforcement of a common set of defined security rules. AWS Firewall Manager can manage and provision AWS WAF rules, AWS Shield Advanced policies for load balancers and other edge services and assets, and enable security groups for Amazon EC2 instances and other resource types in Amazon VPCs.

In addition to cloud-native services, many third-party security and networking partners can be easily implemented in the AWS cloud and can automatically provide scale and dynamic policy controls through API integration.
The biggest drivers in the development of cloud edge use cases are converged controls and capabilities that are deeply integrated into the cloud provider environment, scale and automation.

An example of the first driver in a simple use case would be an organization that is accelerating the move to the cloud software-defined data center and that also wants to reduce the cost of traditional data center edge services. These services could include application load balancing, VPN access for an increasingly mobile workforce, firewalls and WAFs, and DDoS protection for critical services with expensive on-premises equipment and third-party brokering services. All of these capabilities can be built in a single cloud platform using a combination of best-of-breed solution provider platforms and cloud-native services that the provider offers. The overall costs will likely decrease, along with highly dispersed management and operational oversight.

A second example use case would be oriented to scalability and anticipated or unplanned events that occur. For some businesses, significant increases or shifts in resource requirements may occur on special occasions or during specific periods of the calendar year. A large e-commerce provider, for example, might need to rapidly scale up application services and throughput (along with storage, additional compute and more) during holiday seasons and other peak purchasing periods. Scaling to this level on premises may prove to be a challenge for the workloads and underlying infrastructure alone, and security and networking edge services that provide access controls, DDoS protection and WAF services would need to scale simultaneously. In the cloud, all of these controls and services can easily be scaled as needed with a high degree of automation and orchestration based on triggered usage requirements.

Considerations for the Cloud Security Edge

For organizations considering whether a cloud edge option may be a good fit, there are some key considerations. First, decide whether a unified strategy with a single provider for numerous critical services makes sense. The primary benefit is operational simplification and a smaller list of vendors/providers, but another significant benefit is a much more cloud-native and adaptive model for access control, security services, network architecture and more. Table 1 (on the next page) should help enterprise networking and security teams gain a better understanding of the differences.
Wrap-Up: Improving Edge Security

As more core data center services shift toward the cloud, cloud edge architecture and deployment models offer the advantage of convergence and unification of numerous disparate network services into a single brokering fabric/platform for both edge environments and standalone users. This can help improve security by reducing the complexity and increasing interoperability of traditional approaches that often require numerous vendors and services to achieve the same control implementation.

<table>
<thead>
<tr>
<th>Features/Controls</th>
<th>Traditional Models</th>
<th>Cloud Edge Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remote access to data center</td>
<td>Many traditional models have relied largely on VPN technology through SSL/TLS browser access or a dedicated endpoint client.</td>
<td>Cloud edge services can act as a connectivity point where users connect to the cloud edge environment for access to both on-premises resources and cloud services. All policy is defined and applied in the cloud.</td>
</tr>
<tr>
<td>resources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to cloud resources</td>
<td>On-premises network access to cloud resources treats these like any other online properties, using traditional firewalls, proxies and routing controls.</td>
<td>Cloud edge services provide optimized and streamlined network access that is more &quot;cloud aware&quot; for numerous SaaS, PaaS and IaaS provider offerings. This often relies on API integration and request introspection for end-user requests.</td>
</tr>
<tr>
<td>Network access controls</td>
<td>Many on-premises environments rely on switching, routing, firewalls and proxies for access control to resources.</td>
<td>Cloud edge services aggregate a number of network security and access controls (including firewalls) into one unified fabric.</td>
</tr>
<tr>
<td>Web application security</td>
<td>WAFs are usually separate appliances or platforms, or are achieved through brokering to a separate CDN or in-cloud service.</td>
<td>Cloud edge platforms integrate WAF policies and services into the same brokered approach.</td>
</tr>
<tr>
<td>Network threat detection</td>
<td>In a traditional on-premises model, network threat detection is accomplished using next-generation firewall (NGFW) platforms, network malware detection sandboxes or CASB brokering for cloud threat detection.</td>
<td>Cloud edge environments can combine numerous network threat detection capabilities into one cloud-aware model.</td>
</tr>
</tbody>
</table>

Table 1. Traditional Models vs. a Cloud Edge Model
**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](https://aws.amazon.com/marketplace)
## Upcoming SANS Training

Click here to view a list of all SANS Courses

<table>
<thead>
<tr>
<th>Course</th>
<th>Location</th>
<th>Dates</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>SANS Sydney 2020</td>
<td>Sydney, AU</td>
<td>Nov 02, 2020 - Nov 14, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Secure Thailand</td>
<td>Bangkok, TH</td>
<td>Nov 09, 2020 - Nov 14, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Community CTF</td>
<td></td>
<td>Nov 19, 2020 - Nov 20, 2020</td>
<td>Self Paced</td>
</tr>
<tr>
<td>SANS OnDemand</td>
<td>OnlineUS</td>
<td>Anytime</td>
<td>Self Paced</td>
</tr>
<tr>
<td>SANS SelfStudy</td>
<td>Books &amp; MP3s OnlyUS</td>
<td>Anytime</td>
<td>Self Paced</td>
</tr>
</tbody>
</table>