AWS Shield

Threat Landscape Report – Q1 2020

AWS Shield is a managed threat protection service that safeguards applications running on AWS against exploitation of application vulnerabilities, bad bots, and Distributed Denial of Service (DDoS) attacks. This Threat Landscape Report provides a summary of threats detected and mitigated by AWS Shield. You can use this information to expand your knowledge of external threats and improve the security of your applications. The data in this report is derived from systems that AWS Shield uses to protect the availability of AWS, protect applications running on AWS, and alert AWS engineers to changes in the threat landscape. For example:

- Network volumetric events are detected by a system that monitors AWS network traffic and places mitigations as needed to protect the availability of AWS services and applications running on AWS.
- Web application-layer events are detected by systems that monitors traffic patterns on AWS WAF and alerts application owners of statistically significant anomalies.
- Malware is detected by a threat intelligence platform that monitors traffic on AWS and alerts AWS engineers to new threats and changes in botnet behavior.

At the end of this report, you can find additional references with steps you can take to help you protect your application against external threats, with emphasis on addressing recent trends in the threat landscape.

Volumetric Threat Analysis

Volumetric events, including network volumetric events and web application-layer events, are anomalies that are potential indicators of an external threat. Network volumetric events can include traffic that is not normally expected or is not expected at significant volume. This is informed by regular observations of traffic on AWS, knowledge of application use cases, and protocol specifications. DDoS attacks are one of the most common network volumetric events detected on AWS. Web application-layer events are statistically significant changes in both the volume and the composition of web requests. These events are detected when an Amazon CloudFront distribution or Application Load Balancer (ALB) is protected by AWS Shield Advanced.

The following tables summarize events detected by AWS Shield in Q1 2020, with comparisons against Q4 2019 and Q1 2019 to illustrate quarter-over-quarter and year-over-year changes in the volumetric threat landscape. For any network volumetric event that AWS Shield detected as a DDoS attack, a mitigation was automatically placed. Metrics are captured per vector, per resource. For example, a multi-vector DDoS attack against one resource may be detected as multiple events.

Metric	Prior Quarter (Q4 2019)	Most Recent Quarter (Q1 2020)	Change
Total number of events	282,582	310,954	+10%
Largest bit rate (Tbps)	0.6	2.3	+283%
Largest packet rate (Mpps)	282.2	293.1	+4%
Largest request rate (rps)	1,585,615	694,201	-56%
Days of elevated threat*	4	3	-25%

Table 1. Comparison of event frequency and volume since last quarter (Q4 2019 vs. Q1 2020).

Metric	Same Quarter, Prior Year (Q1 2019)	Most Recent Quarter (Q1 2020)	Change
Total number of events	253,231	310,954	+23%
Largest bit rate (Tbps)	0.8	2.3	+188%
Largest packet rate (Mpps)	260.1	293.1	+13%
Largest request rate (rps)	1,000,414	694,201	-31%
Days of elevated threat*	1	3	+200%

Table 2. Comparison of event frequency and volume since the same quarter last year (Q1 2019 vs. Q1 2020).

**"Days of elevated threat" in Tables 1 and 2 indicates the number of days in a given quarter that AWS determined the global threat level based on attack frequency, volume, or other attributes was "High" or "Critical". Under most circumstances, the threat level is "Normal." A threat level of "High" indicates that there is an ongoing, sustained, atypical increase in event activity. The threat level can also be set to "Critical" by an AWS Shield engineer, under exigent circumstances.*

The number of volumetric events detected by AWS Shield is influenced by growth in the number of applications hosted on AWS, improvements to AWS Shield detection, and the frequency at which applications are targeted by external threats. The number of detected events has increased by 23% since the same quarter in 2019. This was driven by an increase in the number of web application-layer events. Network volumetric events decreased by 20%.

Figure 1 shows a weekly profile of the number of volumetric events observed by AWS in Q1 2020.

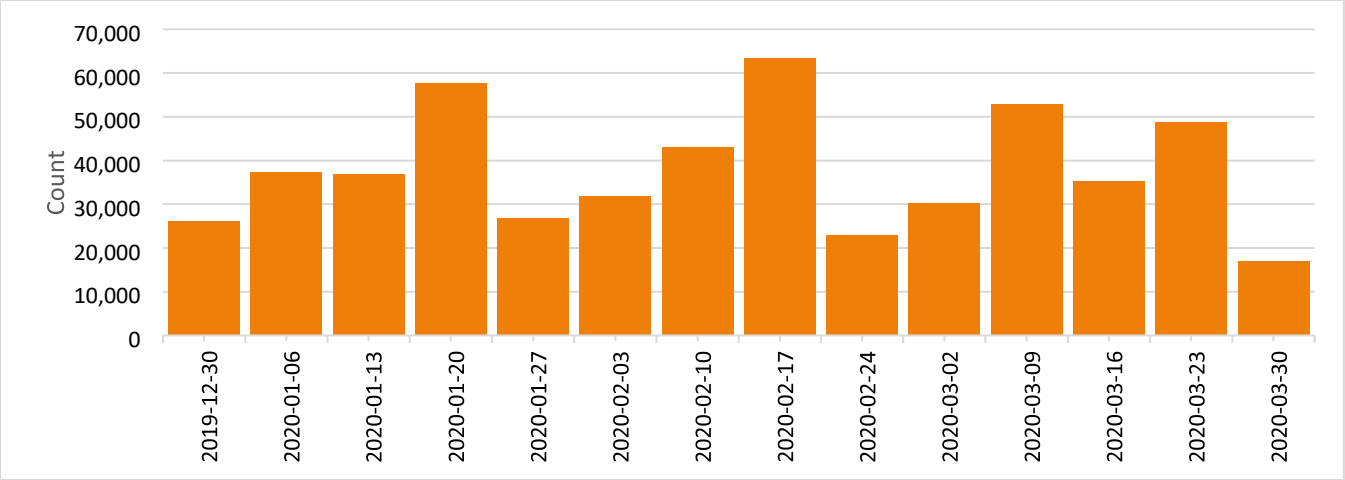

Figure 1. Weekly quantity of events observed for customer applications on AWS during Q1 2020.

Network Volumetric Events

DDoS attacks are the primary driver of larger network volumetric events. The most commonly observed network volumetric DDoS vectors are UDP reflection attacks. This includes attacks like DNS reflection, NTP reflection, SSDP reflection, and many others. Each of these vectors is similar in that an attacker spoofs the source IP of the victim application and floods legitimate UDP services on the Internet. Many of these services will unwittingly respond with one or more larger packets, resulting in a larger flood of traffic to the victim application. The largest known DDoS attacks are UDP reflection attacks. It is common to see the largest UDP reflection attacks shortly after a new vector is discovered. For example, in Q1 2018, a new memcached reflection vector was discovered and attacks greater than 1 Tbps in volume became common in the following days. From Q2 2018 to Q4 2019, the largest attacks observed on AWS were less than 1 Tbps.

In Q1 2020, a known UDP reflection vector, CLDAP reflection, was observed with a previously unseen volume of 2.3 Tbps. This is approximately 44% larger than any network volumetric event previously detected on AWS. CLDAP reflection attacks of this magnitude caused 3 days of elevated threat during a single week in February 2020 before subsiding. Despite this observation, smaller network volumetric events are far more common. The 99th percentile event in Q1 2020 was 43 Gbps.

Figure 2 shows the bit volume of the largest network volumetric events observed by AWS in Q1 2020.

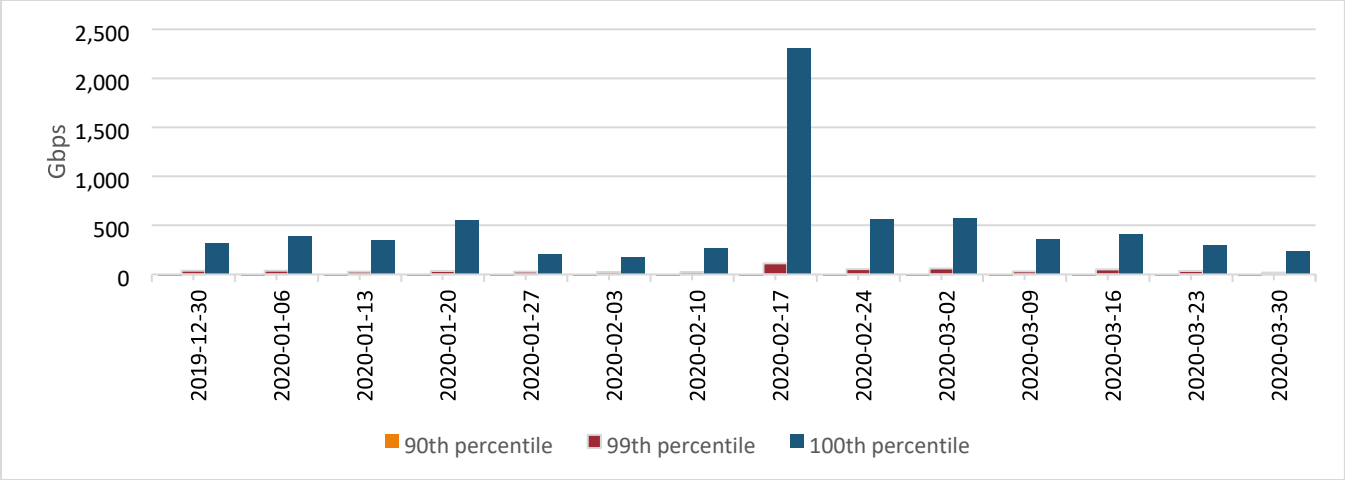

Figure 2. P90, P99, and P100 of volumetric events, measured in gigabits per second (Gbps), for resources on AWS during Q1 2020.

The second most common DDoS vector observed on AWS is a SYN flood. This vector allows an attacker to generate a large flood of traffic while also targeting the state tracking capability of devices like servers, load balancers, and firewalls. SYN floods are commonly spoofed, meaning that it is not usually helpful to block the attack by source IP address. SYN flood packets are often very small, which can make the attack more difficult to absorb. Additionally, each SYN packet may be interpreted by the application as a new connection attempt. This can prevent valid end users from connecting to the application if its resources are exhausted. These attacks are typically smaller than UDP reflection attacks when evaluated in bits, but larger when evaluated in packets. In Q1 2020, the SYN floods observed on AWS were consistent with those observed in prior quarters.

Figure 3 shows the packet volume of the largest network volumetric events observed by AWS in Q1 2020.

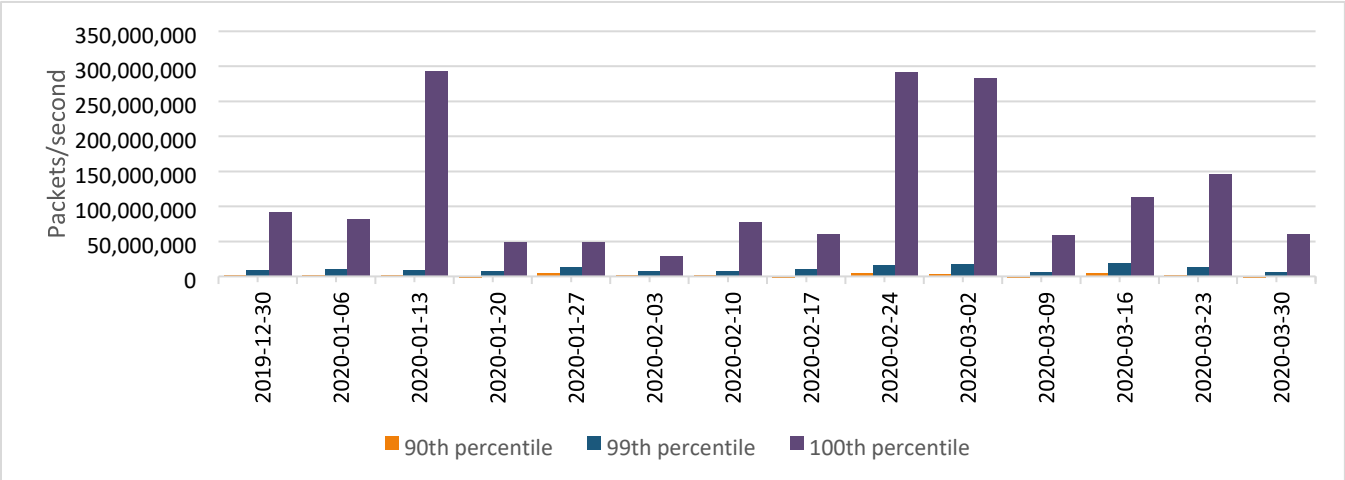

Figure 3. P90, P99, and P100 of network volumetric events, measured in packets per second (pps), for resources on AWS during Q1 2020.

TCP reflection is another known, less common vector that was observed during Q1 2020. It is similar to a SYN flood, except that it uses a flood of spoofed SYN packets sent to legitimate Internet services, which can elicit a larger flood of SYN/ACK packets toward the victim application. In a TCP reflection attack, one spoofed SYN packet can result in many SYN/ACK packets. This occurs when the legitimate service does not receive an ACK packet from the attack and retransmits the SYN/ACK packet many times.

Web Application-Layer Events

Applications protected by AWS Shield Advanced and AWS WAF are also monitored for web application-layer events, like web request floods and HTTP reflection attacks. These events can be DDoS attacks that target the ability of a web application to serve regular users by generating a flood of traffic that is difficult to identify with network-based detection systems. In other cases, they are floods of traffic that are incidental to web content scraping, account takeover bots, or other unauthorized, non-human traffic. Each event in this report is a statistically significant deviation from the baseline of web request traffic for an application running on AWS, where the composition of traffic, based on attributes like user agent, URI, or referrer, also changed.

The peak volume of the largest web application-layer events has declined, relative to the comparison quarters (Q4 2019 and Q1 2019). However, these events are unique in that a very large volume of requests is not necessarily required to threaten an application. In the case of bad bots, it is possible to scrape content or verify stolen user credentials without creating a large flood of traffic. Application-layer DDoS attacks may have a smaller request rate in an attempt to evade detection, while targeting more expensive resources on the application. For example, a request that invokes an API or results in a database query is more expensive than a request for a static object.

Web request floods, a common application-layer DDoS vector, can be mitigated by scaling your application to absorb the additional traffic or by using a web application firewall (WAF). Network-layer devices are generally unable to inspect encrypted requests. This allows an attacker to leverage expensive web requests or large volumes of web requests to generate a flood that is challenging to fingerprint, challenging to block or absorb, or both. The majority of these events are relatively small. In Q1 2020, the 90^{th} percentile of web application-layer events was 1,608 requests per second. It is also possible for these events to reach much larger peak request volumes. In Q4 2018, a web request flood peaked at 20.6 million requests per second. Because the application targeted by this flood was well-architected, using Amazon CloudFront, AWS Shield Advanced, and AWS WAF, the application's availability was not affected. This architecture distributes request handling across many AWS Edge Locations and temporarily blocks source IP addresses that exceed a pre-defined limit.

Figure 4 shows the request volume of the largest web application-layer events observed by AWS in Q1 2020.

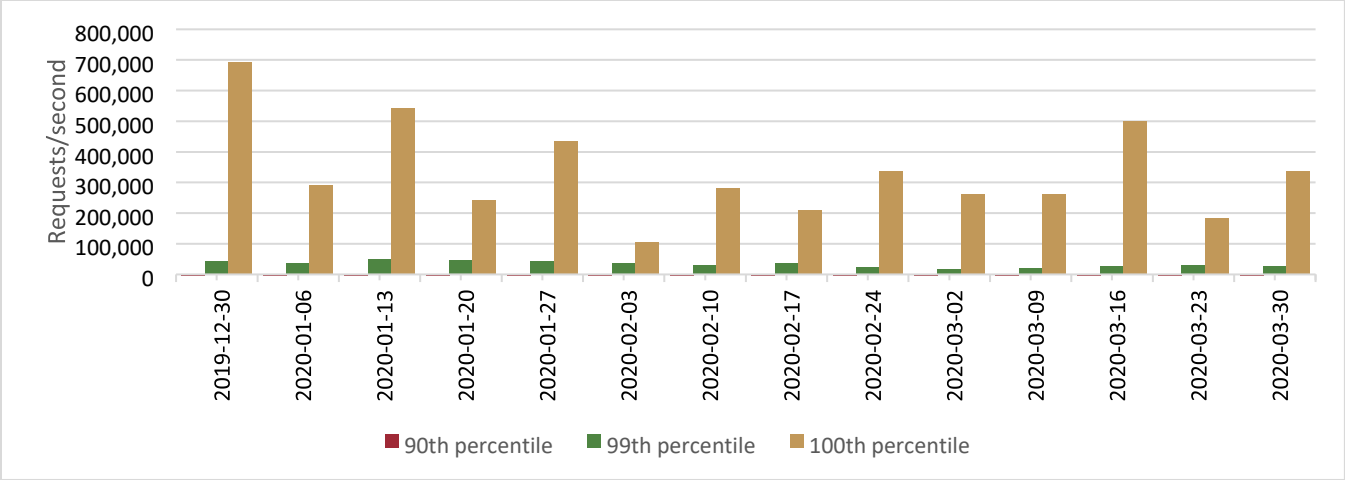

Figure 4. P90, P99, and P100 of web application-layer events for resources protected by AWS Shield Advanced and AWS WAF, measured in requests per second (rps) during Q1 2020.

Malware Threat Analysis

Malware is software that is used by an attacker in an attempt to compromise a targeted application. Many attackers use botnets to scan the Internet for applications that may be vulnerable to exploitation. This can occur through vectors like remote command execution (RCE) vulnerability, privilege escalation, or server side request forgery (SSRF). AWS operates a threat intelligence platform that monitors Internet traffic and evaluates potentially suspicious interactions. This information is used by AWS engineers to stay current on the malware threat environment and to advise application owners on best practices to mitigate the most significant threats.

The following table summarize events detected by AWS in Q1 2020, with comparisons against Q4 2019 to illustrate quarter-over-quarter changes in the malware threat landscape. The events in these tables are a subset of the total, for which AWS has higher confidence in the intent of the attacker. Unique suspects are individual sources, for which one or more events can be attributed.

Metric	Prior Quarter (Q4 2019)	Most Recent Quarter (Q1 2020)	Change
Total number of events (billion)	0.7	1.1	+57%
Unique suspects (million)	1.2	1.6	+33%

Table 3. Comparison of event frequency and interactions since last quarter (Q4 2019 vs. Q1 2020).

The growth in both events and suspects can imply a larger number of exploitation attempts from either a larger number of suspected attackers, or from suspected attackers who are employing more dynamic evasion techniques. For example, an attacker can use many different sources or change their interaction patterns in an attempt to evade detection. Overall, the number of events per suspect has increased 23%.

The interaction types that were most prevalent in Q1 2020 data accounted for 36% of events and 19% of suspects. These include:

- Docker unauthenticated RCE, where the suspect attempts to exploit a Docker engine API to build a container, without authorization.
- SSH intrusion attempts, where the suspect looks for ways to gain unauthorized access to the application using commonly used credentials or other exploits.
- Redis unauthenticated RCE, where the suspect attempts to exploit the API of a Redis database to gain remote access to the application, gain access to the contents of the database, or make it unavailable to end users.
- Apache Hadoop YARN RCE, where the suspect attempts to exploit the API of a Hadoop cluster's resource management system and execute code, without authorization. In March 2020, these interactions accounted for 31% of all events. During this time there were 134 million events, up from 88 million in each of the two prior months.

The motivation of an attacker can vary. Individual interactions may result from an attacker with a specific goal that related to the targeted application. The higher volume interactions are motivated by control of compute and network resources at scale for purposes like cryptocurrency mining, DDoS attacks, or data exfiltration. The frequency of interaction with an application depends on factors like its prevalence on the Internet, availability of unpatched RCE vulnerabilities, and the likelihood that application owners have effectively restricted access to those applications.

Key Takeaways and Recommendations

There are steps that you can take to protect the availability of your applications and protect them against exploitation. The volumetric events described in this report can be mitigated by building resilient architecture, as described in the AWS Best Practices for DDoS Resiliency whitepaper. This whitepaper provides an overview of DDoS attacks and choices that you can make when building on AWS to provide your application the greatest ability to absorb or mitigate volumetric attacks. Larger attacks, like the 2.3 Tbps CLDAP reflection attack or the 20.6 million requests per second request flood described in this report, are lower frequency, high severity threats that are best mitigated when you are closely adhering to the best practices. This can also help protect applications against complex or multi-vector attacks that are often lower volume, but are more challenging to identify or mitigate. These best practices include:

- ***Protect Internet-facing resources with AWS Shield Advanced***. You can use AWS Shield Advanced to better protect your applications running on AWS against vulnerabilities, bad bots, and Distributed Denial of Service (DDoS) attacks. When you add protected resources in AWS Shield Advanced, network volumetric attacks against those resources are detected and mitigated more quickly. You also receive attack visibility using the AWS Shield console, API, or Amazon CloudWatch metrics. If you require assistance during an attack, you can engage with AWS Shield experts or escalate to the AWS Shield Response Team (SRT).
- ***Use AWS Firewall Manager to centrally manage protection policies across multiple accounts***. With AWS Firewall Manager, you can define AWS Shield Advanced, AWS WAF, and Security Group policies spanning multiple accounts in an AWS Organization. AWS Firewall Manager

continuously monitors for resources not in compliance and can automatically apply protections, ensuring consistent adoption of organization-wide best practices for security.

- ***Access greater capacity with Amazon CloudFront and Amazon Route 53.*** You can use these services to serve static and dynamic web content as well as DNS answers using the global network of AWS Edge Locations. This provides you with greater capacity to help mitigate large volumetric attacks. Applications fronted by Amazon CloudFront and Amazon Route 53 also benefit from inline mitigation that continually inspects all traffic and mitigates most attacks in less than one second. SYN cookies are used by CloudFront and the AWS Shield DDoS mitigation systems to verify new connections, protecting against SYN floods and other traffic floods that are not valid for the application.

- ***Use AWS WAF and Rate-Based Rules to mitigate application-layer attacks.*** AWS Shield Advanced provides you with protection against volumetric attacks that can be mitigated with network-based DDoS mitigation systems. When you add AWS Shield Advanced protection to Amazon CloudFront or Application Load Balancer (ALB) for serving web content, you receive AWS WAF at no additional cost. AWS Managed Rules for AWS WAF makes it easy to select and apply pre-configured rules, depending on your specific requirements. You also receive web request flood detection and can mitigate attacks by configuring Rate-Based Rules to match and temporarily block IP addresses that are sending traffic above a rate that you define. For larger applications, or applications than span multiple AWS accounts, you can use AWS Firewall Manager to deploy rules across all of your resources.

It is also important to protect your application against external threats that may attempt to install malware or run other commands, without authorization. Steps you can take include:

- ***Reduce the attack surface***. You can reduce the application's attack surface by preventing access to the application, except as required to serve authorized end users. You can use Security Groups and Network ACLs in Amazon Virtual Private Cloud (VPC) to limit access to the application, or application components, by destination port, protocol, or IP address. For web applications, you can use AWS WAF to build a whitelist of rules that provides a wider range of attributes on which you can match, including header, method, query string, URI, body, source IP address or CIDR, and source country.

- ***Keep software current and follow vendor best practices***. Work with your software vendors to ensure that you are running the latest recommended version of any third-party software and that the application is patched any known vulnerabilities. Some RCE vulnerabilities may use a ServerSide Request Forgery (SSRF) vulnerability to interact with application components that were not directly exposed to the Internet. For example, a web application with a SSRF vulnerability may allow an attacker to craft a request that is executed by the application itself, bypassing access controls intended to prevent interaction with the backend. Software vendors may offer configuration options that prevent authenticated access.

- ***Manage remote access***. If possible, do not allow remote access to your hosts. If remote access is required, do not expose services like SSH or RDP to the Internet. Instead, you can use bastion hosts to act as the primary access point for administration of your application. Privileged users

must first access the bastion, which is protected by security groups to only allow access from a trusted network, like your VPN. Other security groups will only allow access from the bastion.

- ***Monitor your application with Amazon GuardDuty***. You can use Amazon GuardDuty to continuously monitor your application for malicious or unauthorized behavior. This can help you to identify malware behavior, including reconnaissance, cryptocurrency mining, and outbound DDoS attacks.

The following additional resources can help you implement these recommendations and improve the protection of your application against external threats:

- Getting started with AWS Shield
- Getting started with AWS WAF
- AWS Managed Rules for AWS WAF
- Getting started with AWS Firewall Manager
- Getting started with Amazon GuardDuty
- Controlling network access to EC2 instances using a bastion server (using RDP)
- Securely connect to Linux instances running in a private Amazon VPC (using SSH)
- How to help prepare for DDoS attacks by reducing your attack surface

Recent Announcements

AWS is constantly innovating and delivering new capabilities to make it easier for you to protect your applications against external threats. AWS Shield, AWS WAF, AWS Firewall Manager, and Amazon GuardDuty announced many new features during Q1 2020, including:

- AWS WAF and AWS Shield Advanced now available in Asia Pacific (Hong Kong) and Middle East (Bahrain)
- AWS Firewall Manager support for AWS WAF and AWS Managed Rules
- AWS WAF adds Anonymous IP List for AWS Managed Rules
- AWS Firewall Manager now supports AWS CloudFormation
- AWS Shield Advanced now supports Health Based Detection
- Amazon GuardDuty announces threat detection enhancements, reducing alert volume and increasing accuracy for common customer deployed architectures
- AWS Security Hub adds 15 new resources, increases resources limit, and adds RelatedRequirements field
- Amazon GuardDuty Price Reduction