# A Look Into Public Clouds From the Ransomware Actor's Perspective

By Jay Chen

Published: 2022-05-16 · Archived: 2026-04-05 16:45:50 UTC

## Executive Summary

Traditional ransomware mainly targets on-premises IT infrastructure but doesn't work well in cloud environments, which is one reason we haven't heard much about ransomware in public clouds. However, ransomware actors could adapt their tactics, techniques and procedures (TTPs) to be more cloud native, and now is a good time for organizations to get ahead of this possibility.

Ransomware incidents have severely disrupted business operations across all industries. In 2021, the average ransom demand was $2.2 million, and the average payment was $541,010. Since 2020, researchers have detected at least 130 different ransomware families. There is still no sign of a decrease in the frequency and severity of ransomware attacks.

While there are many known ransomware families, there is no known ransomware actor targeting cloud environments. With the increasing demands of the remote workforce, many organizations are migrating on-premises data centers to the public clouds. Cloud-native infrastructure is also the de facto way that startups can quickly build their business.

Here, we explore how ransomware threat actors might operate in cloud environments – what approaches they might use to attack and impact resources in public clouds. We ask questions including:

- Are cloud-hosted resources more resilient to ransomware attacks?
- What types of cloud resources are more vulnerable to ransomware?
- How might adversaries attack cloud resources?

Our goal is to help organizations prepare for this threat while it is still largely theoretical. The discussed scenarios are intended to be broad and not specific to any cloud service provider.

Knowing that threat actors such as Rocke and TeamTNT target unsecured cloud environments, it is sensible that ransomware actors will also turn to the cloud sooner or later. Due to the fundamental difference between cloud-native and on-premises IT infrastructure, existing ransomware will not be effective in attacking cloud environments. Ransomware actors will need new TTPs in order to target cloud workloads.

However, we have seen threat actors evolve in this way to target cloud workloads before. In our latest Cloud Threat Report, "IAM The First Line of Defense," we designated cloud threat actors as a new type of threat, defining a cloud threat actor as "**an individual or group posing a threat to organizations through directed and sustained access to cloud platform resources, services or embedded metadata.**"

Ransomware actors may also find ways to adapt to the cloud – especially if they begin to see the rewards as worth the effort. Unit 42 researchers identified the possible infection vectors, vertical/lateral movements, targeted resources and impact of such attacks. We incorporated our knowledge of existing ransomware groups and known security incidents in public clouds to derive possible TTPs of cloud-targeted ransomware attacks.

**Overall, we believe that cloud environments are more resilient to ransomware.** The shared responsibility model significantly reduces users' burden in securing the infrastructure, platform and software in the cloud. API-driven cloud services make monitoring, automation and centralized access control easier. Cloud-native backup services provide reliable ways to back up cloud resources. Nevertheless, it is the user's responsibility to securely configure, operate and monitor cloud workloads. As IT infrastructure grows with the business, securing thousands of dynamic workloads in a multi-cloud and hybrid cloud environment can become challenging.

Palo Alto Networks customers can get ahead of potential cloud-based ransomware through Prisma Cloud's threat detection capability, which can identify anomalies and zero-day attacks. Unit 42 offers a ransomware readiness assessment that organizations can use to enhance the ability to quickly and effectively respond to a ransomware attack. Cloud incident response services can help address and mitigate cloud security incidents if they do happen.

## Initial Access

**Phishing**, **exposed remote desktop protocol (RDP)**, **compromised credentials** and **unpatched vulnerabilities** are the most common attack vectors that ransomware actors exploit to gain initial access, as detailed in the Unit 42 Ransomware Threat Report. These techniques are simple but effective and can be carried out against any individual or organization. Some ransomware combines multiple techniques and utilizes commodity malware such as Emotet and TrickBot to breach a victim's environment. We believe attackers can also use the same techniques to gain initial access to cloud environments.

As with on-premises IT infrastructure, where hundreds of employees may work in the same network, hundreds of engineers may share and work in the same cloud infrastructure. If an attacker can compromise one engineer's laptop, that attacker can then potentially pivot from the laptop to the cloud infrastructure. This helps explain why attackers often see credentials as a prime target.

Another common attacker vector is unpatched vulnerabilities exposed to the public internet. Our previous research on exposed vulnerabilities found that 24% of exposed VM instances in public clouds have known vulnerabilities. If an attacker could find and compromise a vulnerable VM instance on the internet, the instance could serve as a gateway into the victim's cloud infrastructure.

Below are two hypothetical examples showing how an attacker could use these techniques to gain an initial foothold in a victim's cloud environment.

### Example 1: Compromised Laptop via Phishing Emails

If a developer's laptop is compromised – for example, by phishing emails – cloud credentials stored on the laptop could give attackers initial access to an organization's cloud infrastructure. Cloud Service Providers' (CSPs) command-line tools commonly store credentials in specific directories on the host. Malware that targets cloud environments often enumerates these directories.

If the developer also uses CSPs' web consoles, attackers may also find passwords in the password manager or active access tokens in the browsers. Any retrieved credential could be used to impersonate the developer and gain initial access to the victim's cloud infrastructure.

### Example 2: Compromised Server via Unpatched Vulnerabilities

Any remote code execution (RCE) vulnerability exposed to the internet may allow attackers to gain initial access. For example, an attacker can remotely exploit an RDP server with the BlueKeep vulnerability and gain full control of the server. Since the server is in the victim's cloud environment, the attacker may find cloud credentials on the server and use the credentials to pivot into the cloud.

Another way that attackers can harvest cloud credentials is through metadata services. Metadata services listen on a special IP address that applications on a VM instance can query to obtain the instance's information such as tags, region and cloud access tokens. All CSPs support metadata services. Any exposed vulnerability that allows attackers to reach this special IP address may allow the attacker to steal cloud access tokens. Vulnerabilities like server-side request forgery (SSRF) are commonly exploited to access metadata services.

### Best Practices to Limit Initial Access

These hypothetical examples both illustrate the value of using best practices to limit the potential for initial access. In particular, organizations can reduce the likelihood of cloud incidents – ransomware or otherwise – by educating employees about phishing, improving identity and access management policies and limiting what is exposed to the public internet. If possible, use the latest version of the metadata service to prevent attacks like SSRF.

## Execution

After an attacker gains the initial foothold, their next step is to plan and execute the attack.

The two most probable avenues for executing an attack in cloud environments are through control plane APIs or data plane APIs.

In an on-premises IT infrastructure, administrators typically directly connect to each endpoint – such as a router, a firewall or network-attached storage – to manage each device. In contrast, users rarely directly connect to resources in a cloud environment. Instead, they authenticate with a CSP's API gateway and call a set of control plane APIs to manage cloud resources. This centralized access control is a double-edged sword. It simplifies the access control management, but a leaked credential with high privilege can also be devastating, as every cloud resource may be accessed with the same key.

Control plane APIs are used for resource management, such as creating/configuring networks, creating/deleting VM instances and reading/writing resource policies. Data plane APIs are used for accessing data inside individual resources – for example, reading/writing files on a virtual machine instance or querying data from a database table.

Using the hypothetical examples in the Initial Access section, an attacker could gain access to the control plane if cloud credentials are found on a compromised laptop or RDP server. Once the attacker is in possession of a cloud

credential with sufficient privileges, the attacker can authenticate with the API gateway and access cloud resources.

The attacker may also compromise cloud resources through the data plane. On a compromised laptop, the attacker may find data plane credentials such as VMs' SSH keys or database passwords. From a compromised RDS server exposed to the internet, the attacker may use data plane APIs to search for other vulnerable servers in the same private network or harvest credentials in the environment variables or memory.

Attackers can move from the data plane to the control plane or vice versa. For example, an attacker who compromises a virtual machine instance through the data plane APIs may find control plane credentials in the instance's metadata. On the other hand, an attacker who can access virtual machine service through control plane APIs may use features like code execution[1][2] and remote access control to gain access into a VM instance and attack other hosts from the VM instance through the data plane APIs. In the next section, we will show it is common for attackers to move between the control plane and the data plane to escalate their privileges.

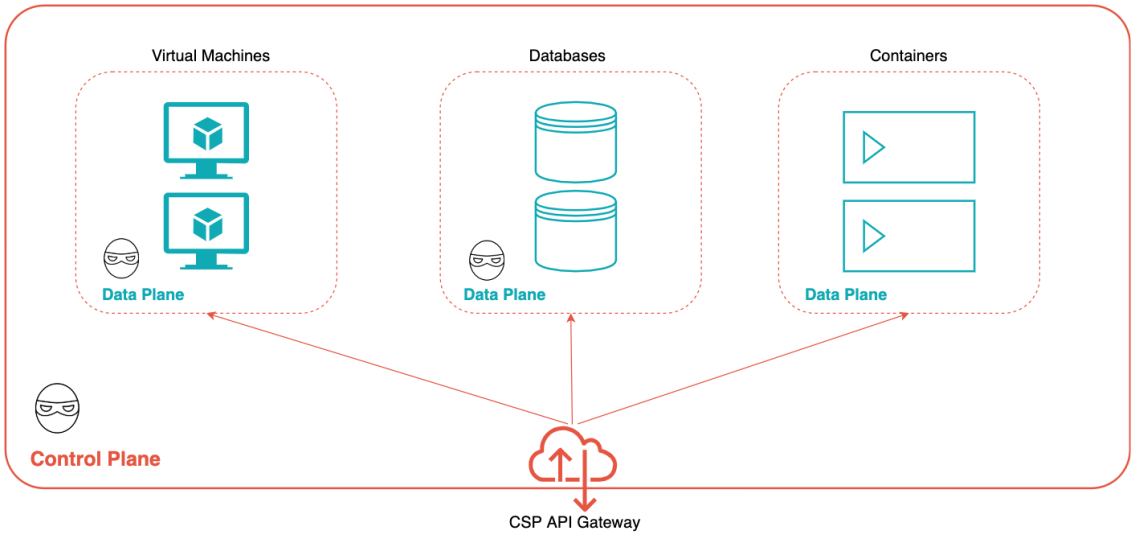


Figure 1. Control plane (red) and data plane (blue) in a cloud infrastructure.

### Best Practices to Limit Execution

Modern cloud-native infrastructure is built with layers of security boundaries, e.g., containers, namespaces, virtual machines, virtual private clouds and account isolation. When possible, create a boundary for each logically separated resource and workload to reduce the impact in case of security accidents.

## Privilege Escalation and Lateral Movement

The objective of a ransomware attack is to extort for ransom by disrupting a victim's business operations and causing loss of business revenue. Attackers accomplish this through privilege escalation and lateral movement – techniques that can allow them to gain access to valuable resources.

Traditional lateral movement techniques such as remote service hijacking or alternation bypass can still allow attackers to move laterally in the data plane. For example, attackers may use tools like mimikatz to dump credentials in the compromised VM instance, identify reachable endpoints in the same network and pivot to

another VM instance in the same virtual private clouds (VPC). The vulnerabilities that ransomware routinely exploits can allow attackers to escalate privileges or perform remote code execution on the compromised VM instances.

However, the cloud has some additional protections against these vulnerabilities – such as CSPs' hypervisors or VPC perimeters, which virtually isolate a group of cloud resources. At the time of writing, no known malware exploits vulnerabilities in CSPs' infrastructure to break out of the security boundary. CSPs are also more diligent in patching and securing their infrastructure than many individual organizations. This means that after attackers compromise a VM instance, they for the most part can't gain access to other cloud resources outside the security boundaries, unless they find ways into the control plane and pivot using control plane APIs.

As a result, the more concerning way of achieving privilege escalation and lateral movement is through control plane APIs. Attackers could potentially bypass all the security boundaries if they obtained the right credentials. The defense mechanism for securing control-plane APIs is identity and access management (IAM). Every request sent to the control plane needs to be first examined by IAM. Any unauthenticated or unauthorized request will be dropped immediately.

## Why Organizations Should Use Strong IAM Policies

IAM is the most critical component that governs the authentication and authorization of every resource in a cloud environment. Put simply: IAM is the first line of defense in most cloud environments. Therefore, **a more "cloud native" way of performing privilege escalation and lateral movement in the cloud is through IAM misconfigurations.**

Due to cloud infrastructure's complexity and dynamic nature, cloud identities are often overly permissive, meaning that they are granted more permissions than they actually need. We recently analyzed cloud accounts from 200 different organizations and found that nearly all lacked the proper IAM management policy controls to remain secure.

For example, a user who only needs to access one storage bucket may be granted permission to access all the buckets in the same account/subscription. Similarly, a container that only needs to write to a database may also be granted read and delete permissions. Our previous Cloud Threat Report described how misconfigured IAM controls allowed Unit 42 researchers to gain access to source code repositories and many private keys of a multi-million dollar SaaS company.

It is common that combining a set of permissions enables another unintended permission, leading to privilege escalation. For example, suppose a user has permission to:

1. Access a VM instance.
2. Modify the identity associated with the instance using actions that can update or impersonate the role associated with the VM.

In that case, the user may associate the VM instance with a more privileged identity and use this instance to gain more privileged permission.

Prior research such as IAM-Vulnerable, GCP Privilege Escalation, and Azure Privilege Escalation identified many vulnerable paths to achieve privilege escalation. Tools like pacu and skyArk can help identify privilege escalation vulnerabilities in IAM – and show organizations where to focus to improve security posture.

This blog focuses on the potential TTPs of ransomware actors in the cloud. The hypothetical attacks demonstrated in the next section assume that the attacker already performed various lateral movement and privilege escalation to gain access to the targeted resources.

## Impacts

Ransomware impacts the availability or confidentiality of a targeted system. Attacks like data encryption and system lockout affect availability, and attacks like eavesdropping and data exfiltration affect confidentiality. In a double extortion attack, both availability and confidentiality are impacted. 60% of the top ransomware in 2021 attempted to both encrypt the data and exfiltrate the data.

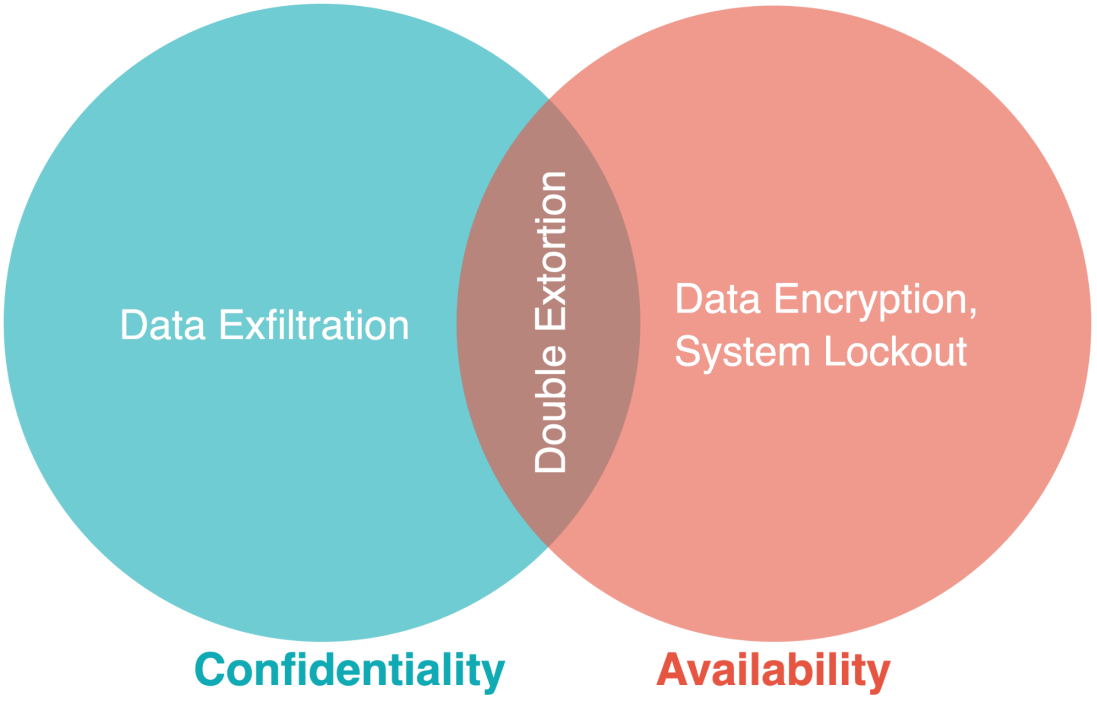


Figure 2. Impacts of ransomware attacks.

### Traditional Ransomware

We consider ransomware that targets Windows or Linux systems in on-premises infrastructure as traditional ransomware, e.g., Ryuk, Maze (ChaCha), Defray777. This type of ransomware attempts to infect many hosts within an organization and then encrypts files on the disk. Traditional ransomware uses file I/O operations to read and write files in the file systems. Any files that can be accessed through file I/O operations are vulnerable to traditional ransomware, including cloud-based file systems remotely mounted to a host such as Network File System (NFS) and Ceph File System (CephFS).

Fortunately, due to the differences in file access methods and architectural design, traditional ransomware is less effective in cloud environments.

**Most Cloud Storage Is Accessed Through APIs, Not File Systems**

More and more cloud-native applications rely on CSPs' APIs to access storage resources. Clients and servers can access the storage resources using the same APIs regardless of operating systems or platforms. Applications running on mobile devices, browsers or IoT devices can download/upload any size and type of data. These API-based cloud storage systems are not vulnerable to traditional ransomware because they are not exposed to file systems.

**Most Compute Resources Are Immutable and Ephemeral**

The design and architecture of cloud-native workloads make them more resilient to traditional ransomware. Compute resources such as VM instances and containers are usually dedicated to running applications, not storing data. Ransomware can still infect these compute resources, but there is no valuable data to encrypt or steal. Furthermore, compute resources are designed to be immutable and ephemeral, meaning that these resources do not change after deployment, and they only "live" for a short time. They are automatically created and deleted by the orchestrator based on the demands of the situation. If an application in a compute resource needs to be updated, a new VM instance or container is deployed to replace the old one. These characteristics make establishing persistence or exfiltrating data from these computer resources difficult for attackers, which contributes to why cloud environments have so far been more resilient against ransomware than on-premises environments.

### Cloud-Native Ransomware

Threat actors, however, may develop new TTPs to make it easier for them to launch ransomware attacks in cloud environments. By thinking ahead to what those TTPs might entail, organizations can better prepare to defend against them – and can identify key best practices for improving cloud security posture.

TTPs for cloud ransomware actors would likely focus on finding the cloud resources that contain persistent data such as object storage, block storage and databases, and then on using cloud APIs to access and encrypt that data.

Because the APIs for accessing every cloud service are very different, threat actors might choose to focus on specific services, or they might develop a different payload for each targeted service (as some traditional ransomware actors today have developed payloads targeting different operating systems).

For example, ransomware targeting one CSP's storage service will be different from ransomware targeting another CSP's storage service. Ransomware targeting an object storage service will be different from ransomware targeting a database service. Considering the number of different data storage services in each CSP, the good news for organizations is that threat actors would likely need more effort in order to launch a successful ransomware attack in the cloud.

## Hypothetical Attack

In this section, we hypothesize a high-level attack scenario. The attack scenario can help us understand the possible TTPs ransomware actors would use in the cloud and create protection strategies against these attacks. The next section will cover several protection and mitigation practices.

For this hypothetical attack, we assume that the attacker gains initial access from a leaked credential and executes the attack using the control plane APIs. Through lateral movement and privilege escalation, the attacker eventually gains access to the targeted data. The attacker then attempts to encrypt the storage resources using the CSP's data-at-rest encryption capabilities. The data-at-rest encryption is usually managed by the CSPs and transparent to the users, but if the attacker can control the encryption key, they may lock others out of accessing the resources. The attack scenario contains three stages: **Create Master Key**, **Modify Targeted Resource** and **Hide Master Key**, as shown in Figure 3.

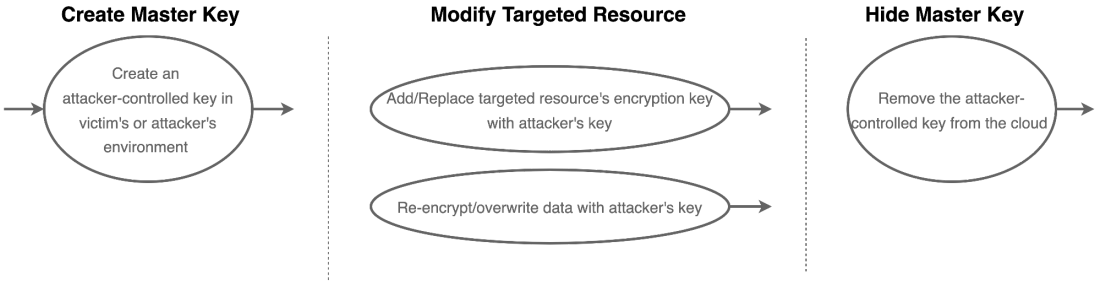


Figure 3. Hypothetical attack scenario.

1. **Create Master Key**: As data encryption in public clouds is often done by cloud-native key management services, an attacker might start by creating an attacker-controlled master key using the CSP's key management services. The key can be created in the victim's cloud environment or in the attacker's cloud environment. In the second case, the attacker must also have the "cross-account" permission to access cloud resources in one account from another account.
2. **Modify Targeted Resource**: In the second step, the attacker attempts to use the attacker-controlled key to encrypt the victim's cloud resources. Some cloud resources allow the data-at-rest encryption keys to be changed after the resources have been created, and the existing data will be automatically re-encrypted with the new keys. Some cloud resources do not allow data-at-rest encryption keys to be changed after the resources have been created. The attacker in this case would then create a decrypted data dump (e.g., snapshots), encrypt the data with the attacker-controlled key, and overwrite the original resource.
3. **Hide Master Key**: The last step is to hide the attacker-controlled keys from the victims or CSPs. An attacker might remove the keys from the cloud environments and locally store the raw key material that can be used to reconstruct the encryption key.

## Protection and Remediation

The root causes of the hypothetical attack scenario were a credential leak and an overly permissive identity. If the victim did not use long-lived credentials, or if the permissions granted to the credential had been more restricted, the attack would not be possible. Most cloud workloads do not need full access to the key management service, nor do they need to update the storage service's data-at-rest keys. As a result, **securing identities is the most important step to start securing a cloud environment**.

While a comprehensive protection strategy against ransomware attacks is out of the scope of this research, major CSPs have all published guidelines to protect their customers (for example, see guidelines from Azure and GCP). Below is a list of best practices that are CSP-agnostic.

**Secure IAM:**

- Minimize the usage of long-lived credentials such as password, access key and service-account key.
- Enforce multi-factor authentication (MFA) for APIs that modify business-critical resources such as database deletion, snapshot deletion and encryption key update.
- Use Federated Identity Management (FIM) to centrally manage access control.
- Grant each user and identity only the necessary permissions for their jobs. (In other words, follow the principle of least privilege). Tools such as AirIAM and IAM analyzer can help generate least-privilege permissions.

**Enable additional protection:**

- Enable version control such as blob versioning and object versioning.
- Enable delete protection such as database delete protection, object lock and resource lock.
- Enable logging on all cloud workloads.

**Create backups:**

- Create backups for business-critical data regularly and automatically.
- Create cross-region and cross-account backups to prevent having a single point of failure.
- Test and verify the backups regularly.

**Shift-left security:**

- Adopt shift-left security to identify vulnerabilities and misconfigurations as early as possible.
- Integrate vulnerability scanner and IaC scanner into every CI/CD pipeline.
- Tools such as Checkov can identify insecure configurations in infrastructure as code (IaC) across all major CSPs.

## Conclusion

Public clouds offer agile, reliable and scalable storage services that on-premises data centers would find almost impossible to keep up with. As cloud adoption is getting faster and more prevalent, cybercriminals will also find new ways to go after valuable data.

Although we have not seen ransomware groups targeting cloud environments, ransomware attacks are probable and concerning, and the time to consider and prepare for the possibility of ransomware in the cloud is now. Our hypothetical attack demonstrates that control plane APIs could be abused to encrypt data in cloud storage and database.

Overall, we think public clouds are more secure and resilient against ransomware attacks than traditional environments. Centralized APIs, immutable workloads and cloud native-backup services all make protecting cloud resources easier. In a shared responsibility model, users can focus more on the applications, and the CSPs take care of the infrastructure, platform or software, which significantly reduces the attack surface compared to an on-premises data center. With the assistance of Cloud Security Posture Management (CSPM), Cloud Workload

Protection Platform (CWPP) or Cloud Infrastructure Entitlement Management (CIEM) solutions, this also makes managing thousands of dynamic cloud workloads feasible.

Unit 42 offers a ransomware readiness assessment that organizations can use to enhance the ability to quickly and effectively respond to a ransomware attack. Cloud incident response services can help address and mitigate cloud security incidents if they do happen.

If you think you may have been compromised or have an urgent matter, get in touch with the Unit 42 Incident Response team or call:

- North America Toll-Free: 866.486.4842 (866.4.UNIT42)
- EMEA: +31.20.299.3130
- APAC: +65.6983.8730
- Japan: +81.50.1790.0200

---