APIs are increasingly powering mobile applications, websites, single-page applications, and data sharing. APIs connect across business and technology boundaries—meaning every enterprise has good reason to be developing, delivering, and consuming APIs. APIs also enable new combinations of hardware and software that change how products and services are delivered. The explosion of API use has left enterprises dealing with the fallout when APIs are built without appropriate consideration to infrastructure scale requirements, user experience, security, reliability, latency, and more.

What questions should enterprises be asking themselves as they review existing APIs and future APIs?

**Have we quantified the impact of API latency issues?**

Many websites, all native apps, and some single-page apps are powered by API calls. We know users, and mobile users in particular, have high expectations for performance. Yet it’s not unusual for an app to utilize 100+ API calls to gather the information requested by the end user—and if just one of those calls is delayed due to network latency, the content may not load.

The good news is that APIs can benefit from a variety of methods used for years to solve website performance challenges, including:

- **Caching**: APIs transmit XML or JSON data that typically contains commingled cacheable and non-cacheable content. Caching is often overlooked during the API development process since it is often considered “lightweight” traffic, but the fact remains that caching this content can help improve response times and offload traffic from API origin, improving user experience.

- **Compression**: By applying LMA compression to JSON and XML payloads, the number of bytes required may be reduced significantly, and fewer bytes translates into better API security and performance—especially when bandwidth is limited.

- **TLS Termination at the Edge**: If the content is secure, then the TLS handshakes can be negotiated at the edge of the Internet instead of at the API origin, which reduces the number of round trips required.
What does it mean to our business if our APIs are not available 100% of the time?

APIs are the bridge between resources for critical applications that are often revenue generating. Think about the APIs that power on-demand transportation mobile apps, real-time navigation systems, supply chain software to make real-time freight optimizations, or provide flight departure updates to those big boards in the airport. Disruptions in these information flows are not acceptable, and it is critical that these APIs are available 100% of the time.

Many enterprises use geographically diverse data centers or cloud providers to host their API infrastructure and redundant network architectures and components to try to combat inconsistent availability. However, traditional hardware load-balancing appliances alone can fall short, and data centers and cloud vendors are prone to unplanned downtime events, which can negatively affect an end user’s ability to access applications. There are several things that can be done to ensure availability, including:

- Utilizing a third party to deliver your APIs and serve as a distributed layer of protection
- Offloading API requests from your API origin infrastructure
- Using third-party services to route to multiple API origins (either your own data centers or a cloud provider’s) based on availability, performance, geographic location, etc., to manage delivery in the event of error or maintenance

How prepared is my API infrastructure to manage traffic spikes?

Like all origins, API servers need to be built to meet peak demand for the data they provide. Similar to web traffic, many APIs experience spikey and unpredictable traffic patterns. Over-provisioning for expected peak demand is an expensive proposition, even when using public cloud infrastructure. API traffic can be particularly difficult to predict as partners and third parties open unplanned new channels for consuming API data. When a server becomes overwhelmed with API requests, responses can be delayed or dropped altogether.

To overcome API challenges, companies are turning to solutions that automatically scale to address peak traffic and detect and avoid security threats.

How should I design the security stack for my APIs?

Just like web servers, API endpoints are exposed to attack risk from hackers and bots. In addition to malicious adversaries, APIs can be accidentally misused by code that doesn’t fit into the use case the publisher envisioned when creating the API, creating an unintentional DDoS attack.

Understanding what the API vulnerabilities are is only part of the solution. There are many different approaches to how to manage API security, but many enterprises leave themselves exposed by not having a comprehensive strategy.

A true defence-in-depth strategy—one that includes a positive security model by defining how the API should be consumed, a negative security model by protecting API endpoints from Layer-7 web application vulnerabilities, and API-specific reporting for increased visibility into how end users interact with available APIs—should be employed.

Once in place, you can ensure your data is secure and your API infrastructure won’t be overwhelmed, all without impacting performance.
Overcoming API Challenges

Most enterprises haven’t considered how a Content Delivery Network (CDN) can help with the delivery of APIs in the same way CDNs help websites and web applications. As the global leader in Content Delivery Network (CDN) services, Akamai has deployed the most pervasive, highly distributed CDN platform with hundreds of thousands of servers deployed in over 120 countries and within more than 1,500 networks around the world. This means that with Akamai, you can:

- Address API latency with our adaptive Internet routing based on real-time intelligence
- Improve performance and reduce origin burden with our caching services, pushing API responses closer to the end user and terminating TLS closer to the client than any other CDN
- Solve the rate-limiting problem with APIs for both traffic spikes that could be related to legitimate traffic (which should be controlled) or security issues (which should be blocked)
- Block DDoS attacks by leveraging WAF Rate Limiting
- Instantly and automatically switch data centers in the event of error or maintenance, and route to multiple data centers of cloud providers based on availability, performance, geographic location, and more with the Application Load Balancer Cloudlet
- Help Manage peak periods of API demand and manage requests during overflow periods with API Prioritization Cloudlet
- Mitigate delays and significantly reduce latency for calls to and from China by utilizing the China CDN to deliver content from Akamai’s network within China
- Control High-Traffic API activity related to bots by utilizing Akamai’s Bot Manager
- Limit requests based on geography, device type, and numerous other vectors with Akamai’s Access Control Cloudlet

As you continue to grow your use of APIs to drive your digital enterprise, make sure you’re addressing the availability, user experience, security, latency, and scaling problems that many enterprises have experienced—before they affect your API projects.

To learn more about how Akamai can help solve common challenges with APIs, visit the Akamai API page.