paloalto NETWORKS | **PRISMA**

Delivering Globally Consistent App Performance to the Hybrid Workforce

Introducing App Acceleration for Prisma SASE

App Acceleration Benefits

- **Drive up to 5X better than direct-to-internet performance (measured in throughput and latency)**. Optimize each user session individually to compensate for changing network conditions and cloud latency.

- **Enrich AI-powered Autonomous Digital Experience Management (ADEM) with Real User Metrics to accurately capture the user experience**, pinpoint network conditions and application performance issues, and reduce MTTR and enhance IT productivity.

To the Distributed Workforce, App Performance Equals Productivity

As hybrid workforces firmly establish themselves as the new normal, the profound cultural and lifestyle transformation this brings dramatically impacts cybersecurity imperatives across today's business landscape.

Hybrid workforces overwhelmingly utilize wireless networks to connect to modern business applications that rely heavily on dynamic content to provide rich and timely data.

Users often experience slow application performance on their devices, for example, in geographic regions with low-speed connectivity, using congested or poor-quality connections.

The overall user experience suffers greatly when using modern apps under these adverse network conditions. This causes a loss of productivity as well as diminished business and customer outcomes. A recent Bankrate.com Economic Indicator poll estimates that 68% of full-time workers support a hybrid work schedule,[1] making consistent application performance a top priority for CIOs and CISOs.

> According to a July 2023 Bankrate Economic Indicator poll, **68% of full-time workers** support a hybrid work schedule.

Problem/State of the Market

To ensure maximum productivity, meet the demands of their customers, and spur competitive differentiation, global enterprises must drive a consistent and secure app experience for their entire workforce. Enterprise SaaS applications are fundamental to productivity, but they don't fall under the control of the enterprise itself. Therefore, any performance or user experience problems associated with SaaS usage often result in negative business outcomes. Thus, businesses are now starting to explore solutions to help optimize SaaS performance while delivering consistent and secure experiences everywhere.

The first barrier to performance optimization is performance measurement. Existing infrastructure and application performance solutions don't provide a real-time measurement of the factors that influence user experience. This can lead to an incomplete understanding of the causes of poor app performance.

The two primary causes of performance degradation are adverse network conditions in wireless connectivity and dynamic content processing speed inside of SaaS apps.

Most users access the internet over a wireless connection, and latency caused by packet loss, packet corruption, and jitter significantly detract from the user experience. The conditions are highly variable based on connection (cellular/Wi-Fi), location (office/home/traveling), and wireless signal strength. Critically, these problems are intermittent and transient, so a one-time adjustment won't address these issues. Further, the device type and OS influence the behavior of the TCP stack, leading to variable performance under these conditions.

At the same time, modern apps are highly reliant on dynamic content, which isn't cacheable because it's different for every user. For example, different employees have different customer relationship management app (CRM) dashboards, and the content has to be processed for each user individually. Current network or application acceleration solutions, such as content delivery network (CDN) and WAN acceleration can't influence dynamic content processing speeds.

---

1. Lane Gillespie, "Survey: 89% of American workforce prefer 4-day workweeks, remote work or hybrid work," Bankrate, August 23, 2023.

Additionally, enterprises need to deliver a superior app experience without compromising security. End users expect a high-performing experience but often blame security controls for performance issues. When this happens, they typically elect to drop out of compliance with security, increasing the risk to the enterprise. This can be attributed to the consumerization of IT, where users expect SaaS apps to perform like native apps. Consumerization of IT is a critical factor in employee engagement, and users need to be able to use the devices they're most familiar with to be productive.

Ultimately, enterprises don't control the client "first mile," app location, or business logic so their workforce productivity suffers in the face of performance issues they can't resolve. This is multiplied across the tens or hundreds of SaaS vendors used to deliver value to customers.

The Causes of Poor Performance Can't Be Addressed by Existing Solutions

Digital organizations have begun exploring adding emerging technologies such as secure access service edge (SASE) to already deployed WAN acceleration or CDN technologies to help improve application performance for their hybrid workforces. Since the causes of poor performance occur outside of the security perimeter, common industry solutions aren't able to address the issues of adverse wireless connectivity or the latency caused by processing dynamic content.

SASE and SSE solutions often tout performance enhancements due to large network footprints or multiple points of presence (PoPs). These PoPs bring connectivity and security processing to local regions as opposed to backhauling remote traffic to centralized data or processing centers. However, adding more PoPs to the secure network doesn't significantly improve performance, as the PoP isn't optimizing the first mile.

WAN acceleration, on the other hand, only focuses on optimizing the network inside the security perimeter. And CDN technology can't accelerate the dynamic content powering modern enterprise apps. In addition to the physical barriers, different protocols and applications behave differently under adverse conditions, so a one-size-fits-all approach doesn't improve the user experience across all apps.

Any potential solution that improves the user experience must:

- Understand the device, network, and app context and how different types of apps behave under adverse network conditions.
- Adapt to network and app conditions in real time.
- Reduce latency caused by processing dynamic content.
- Act within the security perimeter.
- Require no modifications to apps or existing infrastructure.

The Solution: App Acceleration for Prisma SASE

In a world-first, App Acceleration for Prisma® Access combines faster-than-direct internet performance with best-in-class security. With App Acceleration, customers will get up to 5X better than direct-to-internet performance when connecting through Prisma SASE. This **boosts throughput** for all TCP traffic and **reduces cloud latency** when accessing enterprise SaaS apps.

App Acceleration works by directly mitigating the two major sources of performance issues in real time: network connectivity problems in the wireless mile and cloud latency caused by processing dynamic content.

App Acceleration **boosts throughput** via:

TCP optimization: App Acceleration performs aggressive tuning of the TCP window to boost initial throughput for all connections. It then creates a custom packet flow profile for every user session and updates that profile in real time. By taking into account network, device, and app context to create and update the profile, App Acceleration is able to send the maximum amount of data to the user at all times, mitigating the effects of conditions like packet loss, packet corruption, and jitter.

App Acceleration **reduces latency** with:

Cloud acceleration: Modern apps are highly personalized; the content presented to the user is dynamic and different for every user. Since apps are powered by many different domains, APIs and microservices, each browsing action results in many different API calls and requests. This often results in slow loading times for dashboards, analytics processing, and other heavy cloud workloads.

App Acceleration **reduces latency experienced by the user** through:

- **Understanding how real users interact with apps**: App Acceleration learns the different API calls required to load dynamic content for different users. Real user behavioral intelligence (RUBI) technology enables App Acceleration to not only understand what APIs are relevant but also which ones are long poles when it comes to app response time.

- **Intelligently prefetching dynamic content for that individual user**: By being session aware, App Acceleration technology can predictively model user-to-app interaction and dynamically prefetch the different APIs that the user will interact with next. By prefetching the APIs that are causing heavy workloads in the cloud, App Acceleration decouples the actual latency from the user interaction. Thus, App Acceleration prepares the dynamic content ahead of time so it's immediately delivered when the user clicks to load a new page.

Data

As noted previously, the first step toward improving performance is accurately measuring the baseline and changing conditions that affect performance. Measurements as well as modifications to the user's flow profile must occur in real time, and app-aware technology must also act in real time to improve the SaaS experience.

Packet Loss

To collect data on packet loss prevalence, Palo Alto Networks hosted a global entertainment provider's livestream event and monitored the network conditions experienced by users.

In the US and Canada, across multiple ISPs, **50% of all users** experienced packet loss.[2] The median packet loss was 5.48%.[3] The impact of packet loss on a livestreaming environment is experienced as increased **video start time**, increased **rebuffering ratio**, and **lower average bitrate**.

---

2. Real-world data collected during a global livestream event for a major media company.
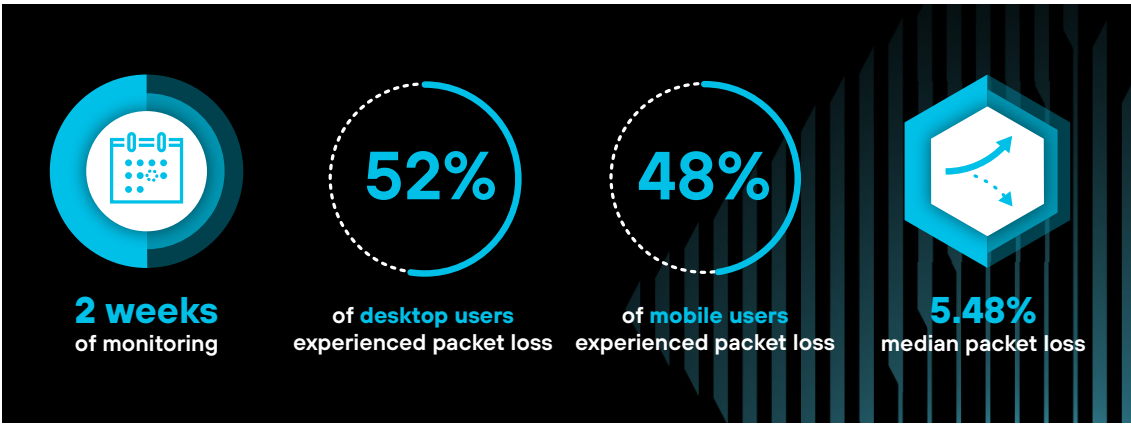
3. Ibid.

Figure 1: Latency landscape: packet loss is universal and intermittent

Utilizing Palo Alto Networks App Acceleration technology, the livestream provider was able to increase watch time across the board, improving a key engagement metric that other providers couldn't improve.

Throughput

To demonstrate the impact of App Acceleration on throughput, several common protocols and SaaS apps were measured: SMB, 1% packet loss, SaaS app downloads, and enterprise SaaS app performance improvement.

SMB

SMB is widely used to transfer large files in enterprise applications and workflows but is highly sensitive to packet loss. Even small amounts of packet loss dramatically reduce the throughput of SMB.

1% Packet Loss

With 1% packet loss, Palo Alto Networks App Acceleration technology significantly reduced the download time of files over SMB when compared with direct-to-internet access. The throughput boost is more pronounced for larger file downloads.
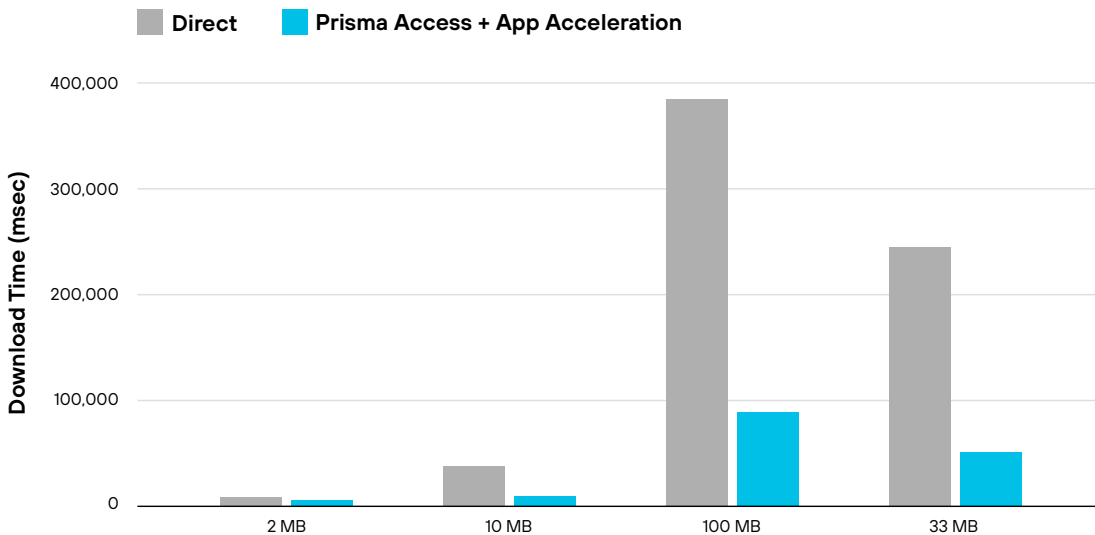

Figure 2: SMB file download speeds going direct versus leveraging Prisma Access with App Acceleration

SaaS App Downloads

Additionally, Palo Alto Networks performed 100 MB download tests using OneDrive, Google Drive, Amazon S3, and Box. In all cases, Prisma Access with App Acceleration was significantly faster than direct-to-internet access.
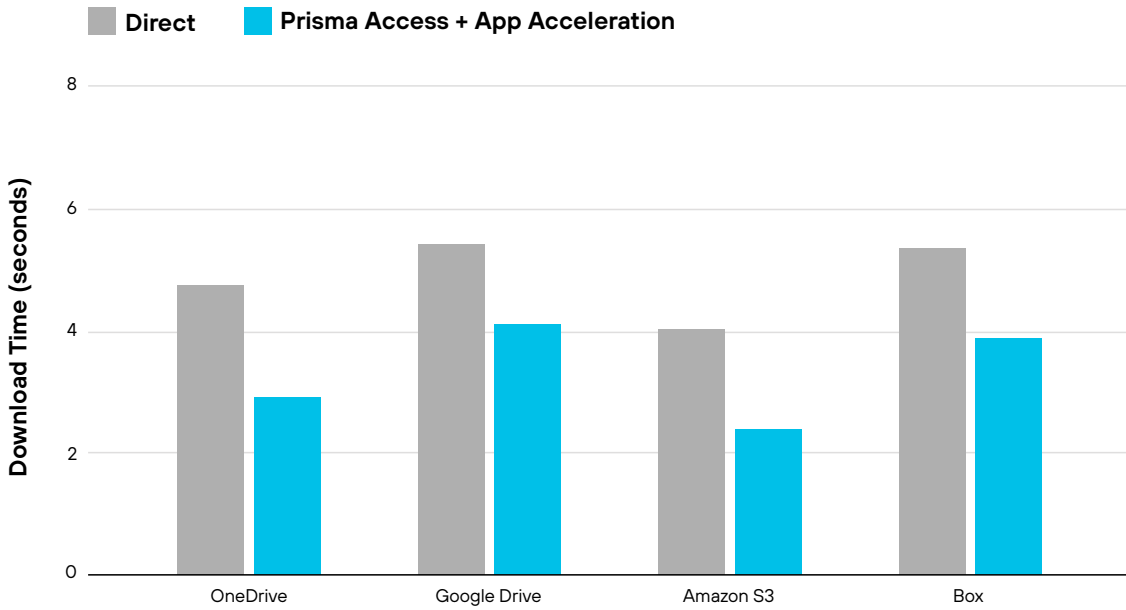

Figure 3: 100 MB file download speed from SaaS vendors going direct versus leveraging Prisma Access with App Acceleration

Enterprise SaaS App Performance Improvement

To show the impact of App Acceleration in SaaS apps, Palo Alto Networks accelerated Salesforce dashboards that load significant amounts of dynamic content. For an accurate representation of the real-world user experience, a Palo Alto Networks Salesforce dashboard was used.
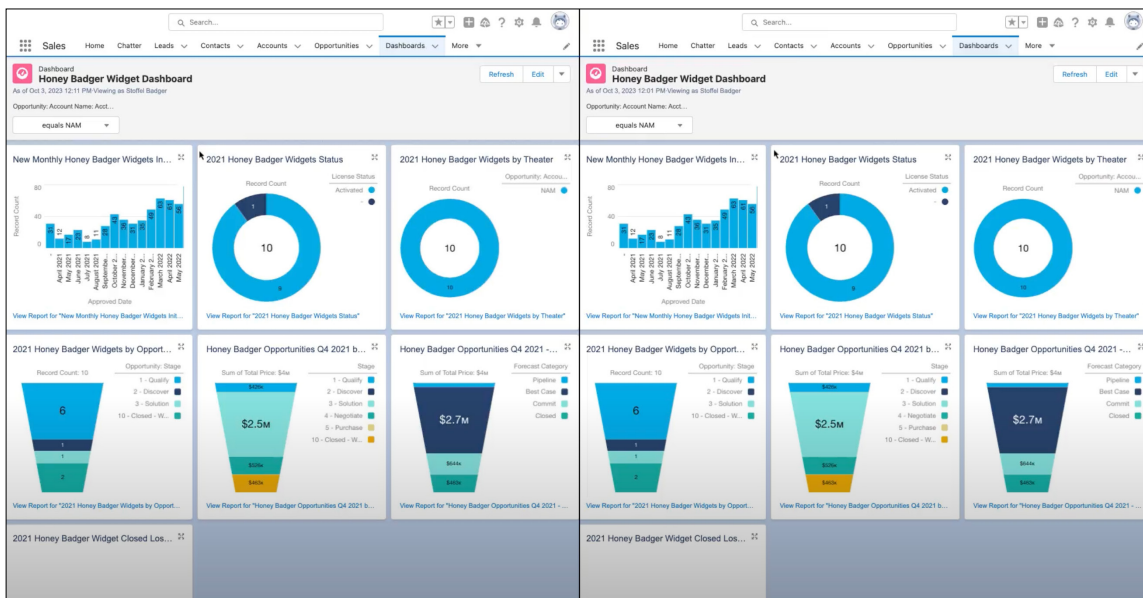
When connected directly to the internet without App Acceleration, the dashboard loaded in 16 seconds.

When using App Acceleration, the same dashboard with the same data loaded in 4 seconds. Since this is dynamic content that isn't cached in the browser or at the CDN layer, the performance boost came from App Acceleration's RUBI technology understanding what content the user needed ahead of time, and intelligently making those requests on behalf of the user.

This is demonstrated by the improvement in API response times when using Salesforce, as measured in the web browser.
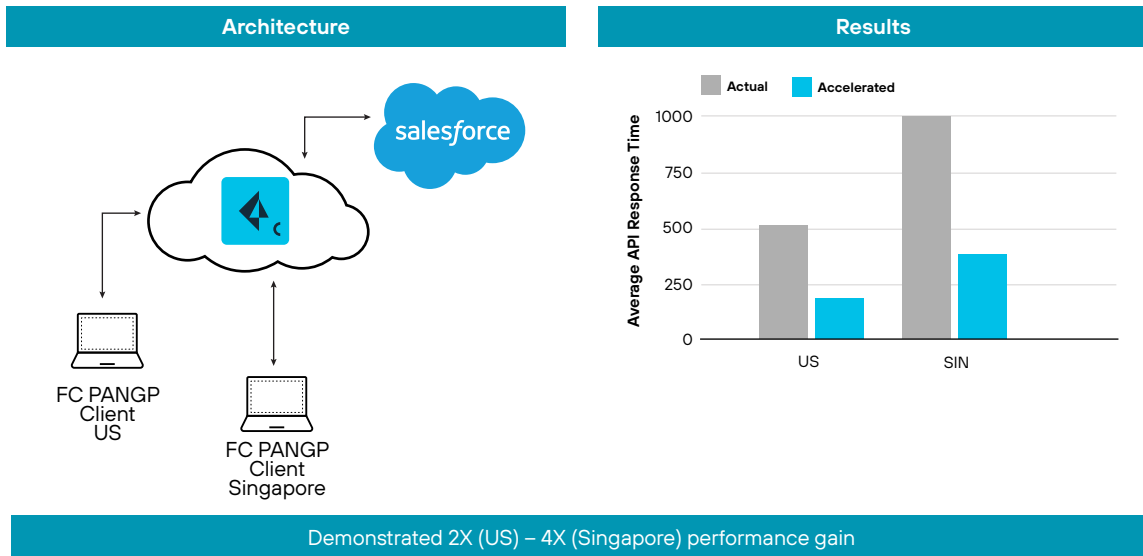
Architecture	Results

Demonstrated 2X (US) – 4X (Singapore) performance gain

Figure 4: App Acceleration PoC – Salesforce

Technical Deep Dive

How do the components of App Acceleration work?

Throughput Improvement

App Acceleration uses app-aware technology, meaning that it understands the individual user experience at the network and application level, and how that user interacts with various apps. Under the hood, App Acceleration uses bots—autonomous processes spawned for each user—to optimize every user session individually, taking into account connection, device, and app context.

First, App Acceleration starts with an aggressively tuned TCP congestion window to start downloads at a higher throughput than direct-to-internet access. Throughout the duration of the session, when App Acceleration encounters adverse network conditions, it's able to use its real-time understanding of the connection to recover throughput faster than the standard TCP stack.
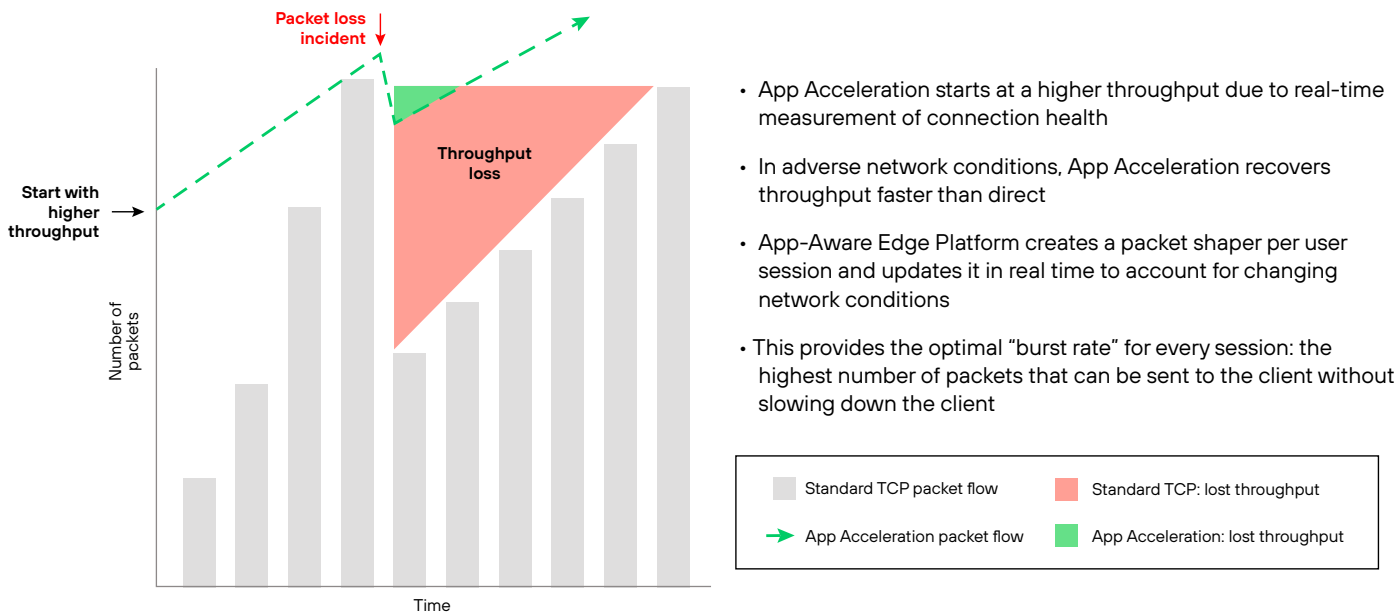
Figure 5: How an app-aware edge platform boosts throughput

- App Acceleration starts at a higher throughput due to real-time measurement of connection health

- In adverse network conditions, App Acceleration recovers throughput faster than direct

- App-Aware Edge Platform creates a packet shaper per user session and updates it in real time to account for changing network conditions

- This provides the optimal "burst rate" for every session: the highest number of packets that can be sent to the client without slowing down the client

Legend:
- Standard TCP packet flow
- Standard TCP: lost throughput
- App Acceleration packet flow
- App Acceleration: lost throughput

Common industry solutions for performance optimization don't have this view into real-time network performance, and as a result, can't mitigate the causes of poor performance. CDNs don't measure packet loss for every user and can't adapt on a per-user basis, and WAN accelerators only optimize the network inside the security perimeter.

Cloud Latency Reduction

SaaS apps' reliance on dynamic content requires a user-specific approach to acceleration. As noted previously, most SaaS app content isn't cached by the CDN, since it's richer than static content and must be timely and different for every user.

Instead of attempting to cache all content, App Acceleration learns the user journey inside of the SaaS app. It provides every user with a bot that accelerates that specific user journey. This reduces the time to first byte (TTFB) and content download times, providing a snappy experience for the accelerated SaaS app.

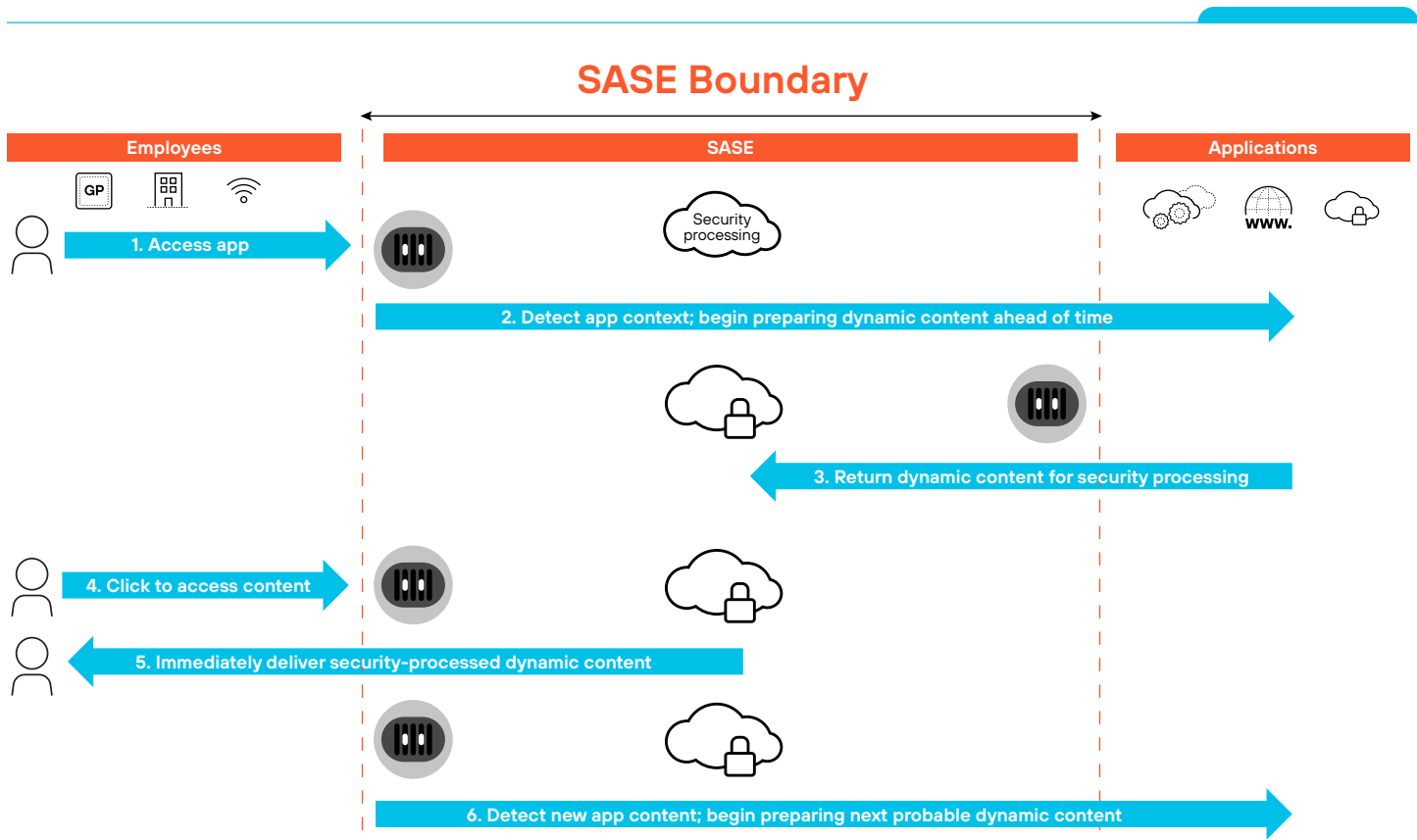The bots follow a predictive workflow depending on the user context within the app.

SASE Boundary

Employees	SASE	Applications

1. Access app

Security processing

2. Detect app context; begin preparing dynamic content ahead of time

3. Return dynamic content for security processing

4. Click to access content

5. Immediately deliver security-processed dynamic content

6. Detect new app content; begin preparing next probable dynamic content

Figure 6: How an app-aware edge platform reduces cloud response time

Outcomes

By accurately measuring the causes of poor performance and mitigating them on a per-user basis, Prisma SASE with App Acceleration drives improved productivity and employee engagement. By exposing the real-time data to admins in Strata™ Cloud Manager, App Acceleration helps improve MTTR through increased observability, pinpointing performance bottlenecks.

Zero code, infrastructure changes, or additional client configuration are required to use App Acceleration. Enable App Acceleration with a single switch inside of Strata Cloud Manager.

Conclusion

For the first time ever, adding a security product can boost performance beyond direct-to-internet access. Prisma SASE App Acceleration is available and being used by customers in real-world deployments today.

For more information, visit www.paloaltonetworks.com.