International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V5I3P119
Eureka Vision Publication | Volume 5, Issue 3, 191-200, 2024

Original Article

Resilient Multi-CDN Delivery Model Using AI-Based Traffic Switching for Global AEM Deployments

Siva Sai Krishna Suryadevara
Sr. AEM Cloud Engineer at Maganti IT Resources, USA.

Abstract - Fast and uninterrupted content delivery is a must-have for modern digital experiences all over the world. However, organizations that are more reliant on Adobe Experience Manager (AEM) have been reported to experience regional performance variations, CDN outages, traffic spikes as well as complex routing rules, which in turn make their traditional multi-CDN setups brittle and reactive. Instead, it constantly updates its knowledge from real-time telemetry that includes latency trends, throughput changes, error rates, geographic traffic patterns, and CDN health signals so that it can foresee degradation and reroute users to the best-performing CDN without any intervention. The approach leverages anomaly detection, reinforcement learning, and predictive modeling along with AEM's dispatcher and edge-layer configuration to coordinate smart, policy-compliant routing decisions that do not interfere with content workflows. The results from the simulations and controlled experiments demonstrate that the strategy drastically lowers the failover time, enhances the global time-to-first-byte (TTFB), and makes the user experience more stable during peak loads or CDN disruptions. Besides that, the AI layer exhibits the capability of optimizing cost-performance trade-offs by distributing the traffic selectively depending on contract thresholds and regional performance insights. The research also sheds light on architectural issues such as the model being at the edge, safe-rollback mechanisms, governance controls, and integration with existing CDN APIs. In essence, the suggested model gives a realistic and future-ready route to enterprises with AEM as a means to elevate their resilience level, have more performance consistency, and gain greater operational efficiency in their multi-CDN ecosystems thus, enabling the transition from a reactive stance to an intelligent, automated, and continuously improving delivery framework.

Keywords - Multi-CDN, AI-Based Routing, AEM, Content Delivery Networks, Latency Optimization, Fault Tolerance, Predictive Switching, Traffic Engineering, Global Delivery, Web Performance.

1. Introduction

1.1. Background & Context

With the increasing digitization of the world economy, the user experience has become the main driver of customer loyalty, revenue models, and the general brand perception. Companies that use Adobe Experience Manager (AEM) to deliver content have to cater to millions of users who are spread across different locations and can access the site or portal, content assets, and personalized experiences at a fast and seamless pace. To keep up with this standard, businesses are turning to Content Delivery Networks (CDNs) which provide the service of locating the content closer to the end user. As the worldwide demand for content delivery continues to grow, CDNs have become the core of high-performance AEM architectures, the ones that are reducing the delay, origin traffic, and increasing the availability even during peak events.

On the other hand, CDNs are also exposed to outages, regional slowdowns, or transient internet congestion. There have been several instances recently when large-scale CDN disruptions have led to the collapse of thousands of websites at the same time, and it has become evident that the single-provider dependency is a fragile one. Even outside these major events, internet conditions are always changing and thus, the cases of routing issues, middle-mile congestion, ISP throttling, and localized network health variations that affect AEM sites loading in different regions are happening more and more frequently. So, in this case, a global enterprise operating through a single CDN is putting digital experience delivery at risk by creating a single point of failure.

Many companies have implemented a multi-CDN solution to combat these problems. Multi-CDN architecture means that content is distributed across two or more CDNs to have the benefit of redundancy and increased global reach. This model is supposed to offer more availability and less risk of a vendor outage. Yet, to get around static routing rules, DNS load balancing, or manual intervention, which are some of the methods used in multi-CDN frameworks that do not take into account the dynamic nature of the real-time network conditions most of the time, is a challenge faced by the multi-CDN strategies. In a world where digital traffic is growing and user experience expectations are high, the static multi-CDN strategies become obsolete insofar as they are not able to keep up with the changes of performance fluctuation.

1.2. Problem Statement

With the ever-changing global content delivery landscape, the necessity for an intelligent, fault-tolerant & globally consistent model to distribute AEM traffic to multiple CDNs across the globe is quite obvious. Traditional methods like DNS-based traffic steering or manual failover are essentially reactive. They only respond to disruptions after these have already impacted the end users. In addition, DNS changes do not propagate quickly, granularity is limited, and performance variations cannot be adjusted to in real time.

Today, multi-CDN scenarios demand a device that is able to understand network signals, identify the very first symptoms of degradation, compare performance across CDNs, and route traffic to the best provider automatically and instantly—without any human intervention. It is not so much about selecting the "best CDN" as it is about continuously figuring out the best CDN for each region, each user segment, and each moment in time. Not having such a capability results in performance inconsistency, the risk of resilience, and higher operational costs in AEM deployments.

1.3. Motivation

Such a model for switching traffic intelligently using AI is needed because of various factors emerging from the engineering, operations, and business spheres of the company.

Resilience, First of all, has become the main concern of worldwide digital platforms. Outages of CDN are getting more and more frequent, and enterprises have to run their businesses continuously without breaks. A delivery layer that is automated and smart is a guarantee for such cases that traffic will be able to move fast to other CDNs not involved in the failure before the users get the wrong service feeling.

Second, improving user experience is a must. AI-driven routing may even start before the requests arrive to preemptively guide users to the most rapid and stable CDN locally, thus easing the waiting time and expanding versatile accessibility over devices and networks. Achieving such an extent of device-network optimization via a conventional and rule-based multi-CDN method is practically impossible.

Third, this means a lot of money. AI, in fact, can tradeoff between cost and performance in real-time by distributing traffic to locations where the expenses are low while the output is kept at a good level or improved. The organizations will be detained from the traps of excessive charges and besides, by making use of the cheap CDNs during the stable periods, as well as optimizing traffic distribution based on contracts, they can go for substantial cost savings.

Fourth, the necessity for turning the performance engineering business into an automated, real-time operation funnel is increasing. Engineering groups that have embraced SRE and DevOps ways of working are becoming more dependent on intelligent systems for complexity management in operations. The scaling of digital experiences makes the manual adjustment of routing policies an unsustainable task. AI-driven automation is a perfect fit with current operational paradigms, as it lifts the heavy hand of work and makes higher reliability possible.

Eventually, the embedding of such a model into AEM's worldwide delivery system goes beyond just supporting the core objectives of digital experience analytics, observability-driven operations, and continuous improvement. AI gives the organizations the ability to move away from conceiving content delivery as a fixed infrastructural task to seeing it as a flexible, responsive, and strategically valuable resource.

2. Literature Review

2.1. CDN Architectures and Evolution

Over the last twenty years, Content Delivery Networks (CDNs) have significantly evolved to provide the kind of infrastructure that digital applications consuming low-latency and high-availability for their global audiences require. Users' requests were addressed to the edge location that was nearest to them, latency thus being reduced, and the load on origin servers being minimized. However, while the system was adequate for static content, its inflexibility hindered it from being viable as the needs of digital experiences became more dynamic.

The advancements in CDN architectures have led to the availability of dynamic content acceleration, protocol optimization, and intelligent routing that allow for faster delivery of personalized or frequently changing data. Until recently, CDNs have transited to edge computing, where small and simple operations like request filtering, A/B testing, header manipulation, security checks, and dynamic assembly are carried out right at the edge nodes; thus, CDNs are no longer just caching layers but rather programmable global platforms.

To scale digital experiences, companies have gone for multi-CDN strategies that, apart from enhancing reliability, improving performance globally and reducing the risk of single-point failure, also provide them with better bargaining power vis-a-vis the suppliers. Industry research published by eMarketer, the State of the CDN Industry Report, and Sandvine suggests that adoption of Multi-CDN is increasing in the sectors of media streaming, e-commerce, and global publishing.

2.2. Traffic Management & Routing Techniques

Multi-CDNsetups have relied on varied methods to manage traffic steering, each having different pros and cons. One of the earliest and most heavily used ways is DNS-based switching. It alters the CDN from which the end-user will get a file by sending a different IP address through the DNS response. Although using DNS results can reroute traffic in a straightforward and almost universal way, this method has drawbacks in that the cache can still be old and propagation slow; therefore, real-time switching cannot be achieved.

Besides that, the strategy that is usually explained together with the anycast technique is one whereby different servers employ the same IP address and traffic reaches either the closest or the least loaded node without needing an explicit routing. In general, anycast increases the network reliability; however, it cannot be used to measure the performance of various CDN providers, let alone serve as a granular feature for intelligent decision-making.

With application-level switching, routing adjustments can be realized at a higher level, mostly through the use of load balancers or ADCs (Application Delivery Controllers). Such a system gives more discretion to the user since it makes possible the on-the-fly switching of traffic according to the specified application metrics, such as response time or signal quality. Notwithstanding this, it is only effective when there is significant instrumenting and integration effort, which is rarely the case in extensive AEM deployments.

2.3. AI/ML in Network Traffic Engineering

Network traffic engineering, an area that has glibly led the changes of distributed systems with the coming of AI and ML, has in turn incorporated the latter two technologies to cope with this very challenge. The highlight of the research was to show the strength of the RL approach in the optimization of routing decision-making through the animated learning of network performance charts from time to time. The RL oriented setups can indeed keep their performance in tune with the present situation, task themselves with experimental routes, and, upon returning a positive performance reward, get to the optimal theme through such a reward-based feedback loop.

Besides this, predictive modeling has also attracted a lot of attention for it can anticipate network congestion, upsurges in traffic, latency, and even failure trends. LSTM neural networks, decision trees, and gradient boosting are some of the models that have been utilized to scrutinize the telemetry of the past and thus predict the network behavior ahead of degradation. The foresight provided by the predictors can still bring in the prevention factor, which keeps the global content delivery flexible and bound to go on.

ML techniques have also been turned to congestion avoidance with a view to making it possible for the traffic on the major route to be rerouted in sufficient time. These models keep an eye on such patterns as heightening of the latency or packet loss and, before full-scale congestion appears, they suggest the alternate paths that might be taken. Likewise, anomaly identification can be ranked at the top of AI-driven fault diagnostic systems, which, by juxtaposing real-time with learned datapoints, accomplish the task of fault detection at a faster hour and smart remedial action too.

2.4. Multi-CDN Challenges Documented in Prior Research

While multi-CDN adoption offers resilience and performance advantages, a number of challenges associated with it have been documented by researchers and practitioners, which hinder its implementation. The issue of vendor lock-in is most often referred to, in which companies use one CDN's proprietary APIs, configurations, or performance patterns to an extent that is heavy. As a result, over time this dependence diminishes the effectiveness of multi-CDN diversification.

Also, different CDNs have different APIs, which makes the unification of the orchestration complex. Differences in cache purge mechanisms, edge rule syntax, security policy models, and analytics formats are some of the reasons for these. The incompatibilities almost always allow the enterprises to duplicate configurations and automate workflows separately, thus increasing their operational expenses.

Many studies have found that the problem of monitoring being fragmented is frequently raised. The lack of a centralized view is due to CDN dashboards, logs, metrics, and telemetry patterns being distinct for each CDN. Hence, engineering teams are not able to compare performance across CDNs and therefore, they cannot make traffic distribution decisions based on data. Fragmented observability also leads to a slowdown of performance issues during incident periods.

Table 1. Summary of Key Literature Related to Multi-CDN Delivery and AI-Based Traffic Engineering

Author(s)	Year	Focus Area	Key Contribution / Relevance
Wu et al.	2021	Multipath transport, traffic steering	Explains dynamic traffic steering and the need for real-time path adaptation; supports dynamic routing in multi-CDN systems.
Cui et al.	2018	Multi-CDN architectures, privacy	Highlights multi-CDN orchestration complexity and fragmented observability—issues addressed by AI-based routing.
Viola et al.	2020	Predictive CDN selection using LSTM	Demonstrates that predictive models (LSTM) can optimize CDN selection based on performance forecasts.
Liu et al.	2013	Client-side CDN/server choice	Shows benefits of performance-driven, continuous measurement-based endpoint selection.
Vagmi & Gupta	2023	CDN evolution and challenges	Reviews CDN trends and multi-CDN adoption; identifies performance variability and vendor lock-in challenges.
Mikkilineni et al.	2017	Cognitive networks, QoS	Introduces automated QoS assurance across distributed systems, supporting policy-based and intelligent routing.

3. Proposed Methodology

3.1. System Architecture Overview

The outlined approach is based on a robust, flexible multi-CDN delivery architecture that is uniquely tailored for worldwide Adobe Experience Manager (AEM) deployments. In essence, the architecture fuses AEM's delivery components with an AI-powered traffic switching engine that can analyze the real-time telemetry and hence, make the routing decisions accordingly.

This methodology's multi-CDN design is above the dispatcher layer. Instead of sending all the content through a single CDN, the traffic is split among several providers like Akamai, CloudFront, Cloudflare, and Fastly depending on the real-time performance metrics. It means that each CDN is just one of the possible delivery nodes, and with the help of the architecture, the receiving of the requests can be rerouted automatically and almost instantly to the other nodes if there are performance, availability, or cost issues.

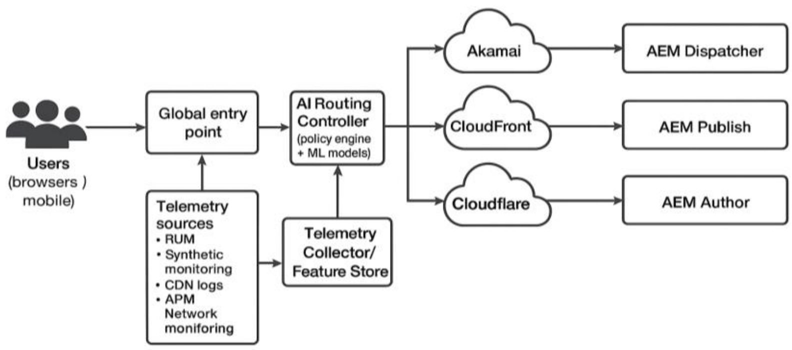

Figure 1. High-Level System Architecture

3.2. AI-Based Traffic Switching Model

Central to the methodology is an AI-based traffic switching model that changes the routes automatically based on both real-time and predictive telemetry. The model consumes different streams of data such as Real User Monitoring (RUM) metrics that measure the actual end-user latency and page load experiences, in addition to synthetic monitoring results from active probes, which replicate user journeys across regions.

By relying on such a rich data foundation, the model's predictive engine is able to discover very critical network and delivery anomalies. For instance, it detects latency spikes by observing abrupt increases in TTFB or full-page load times over various regions; it traces regional outages through signs such as edge node unavailability or routing propagation issues; and finally, it detects throughput degradation caused by congested CDNs, inefficient peering, or slow file transfer patterns. The availability of such information makes it possible to spot performance risks at a very early stage, long before they actually turn into user-visible disruptions.

The AI decision layer is either realized through reinforcement learning (RL) or supervised machine learning. With the RL scenario, the model keeps trying out different routing paths and changes its behavior based on reward functions that measure latency reduction, stability, lower error rates, etc. Once the telemetry changes, the RL agent updates its policies to achieve the routing optimality in a totally autonomous way. On the other hand, a supervised ML strategy comprises an array of models such as LSTM networks, random forests, or gradient-boosted trees that leverage historical telemetry for training and predict future CDN performance. The moment the model sees performance going down, the traffic is rerouted to less loaded CDNs on a proactive basis.

3.3. Telemetry and Observability

Telemetry and observability represent a crucial layer in the methodology. This is because the AI model is effective only when it is supported by high-quality, real-time operational data. To achieve observability, a multi-tool ecosystem is used: New Relic Browser and Synthetic Monitoring offer real user monitoring and synthetic availability checks, and meanwhile, Datadog APM and Network Monitoring empower the in-depth insight into application performance from edge to origin. Furthermore, there are the CDN-native analytics platforms like Akamai mPulse, CloudFront logs, Cloudflare Radar, and Fastly real-time logs that together provide highly detailed visibility into worldwide content delivery behavior.

Real-Time Telemetry Collector is continuously performing aggregation of metrics from all these systems and then streaming them into a time-series datastore or onto an event-streaming platform like Kafka or Kinesis. To respond to the situation in real-time, AI model operates with the help of the data coming to it via a subscription layer. This subscription layer triggers inference cycles immediately after new telemetry is ingested. Hence, it results in a high-frequency, data-driven decision loop that empowers routing behavior to be adjusted dynamically.

Continuous feedback mechanism is always there to check if the model’s predictions correspond to the actual outcomes observed. When the changes in routing lead to better performance, the model thus strengthens the related weights. If, however, the changes fail to bring about good performance or make it worse, the model, therefore, adjusts its parameters accordingly. Such an iterative learning loop enables the system to not only be present in real time but also to be able to maintain optimal routing strategies as well as evolve.

Table 2. Telemetry Signals Used by the AI-Based Traffic Switching Model

Metric	Symbol	Source	Granularity	How It Is Used
Time to First Byte (TTFB)	L	RUM tools, CDN logs	Region/CDN	Core latency input to the model.
Page Load Time	P	RUM, synthetic testing	Region/device	Helps evaluate full user experience.
HTTP Error Rate	E	CDN logs, APM	Region	Identifies instability or partial outages.
Throughput	T	CDN logs, network monitors	Region/CDN	Detects congestion or slow transfer.
Availability / Edge Health	H	CDN health APIs	Region	Used for failover decisions.
Cost Per GB / Request	C	Billing systems, contracts	Region/CDN	Used for cost-aware routing decisions.
Traffic Volume	V	CDN logs, load balancers	Region	Helps scale routing weights during spikes.

4. Case Study

4.1. Scenario & Environment

A major North America, Europe, and Middle East-based multiregional retail brand with operations in Southeast Asia decided to use Adobe Experience Manager (AEM) as their key platform for managing digital storefronts, product pages, promotional campaigns, and customer engagement journeys. A million active users and an extensive catalog of frequently changing content made it necessary for the organization to be heavily dependent on the performance, scalability, and resilience of their global content delivery infrastructure.

The retailer’s traffic pattern was a telling factor of seasonal changes. The most significant traffic spikes occurred during Black Friday, Singles’ Day (11.11), end-of-year holiday sales, and region-specific events like Ramadan and local clearance weeks. On the maximum load days, in just a few minutes, the number of visitors could increase by 8–12 times, which would put dispatching both AEM’s layer and CDN infrastructure responsible for global distribution under a heavy load.

Just as the company’s traffic was spread worldwide, the traffic origins were diversified as well. The digital request volume of the enterprise was such that 40% of them came from North America, 30% from Europe, 20% from Asia-Pacific, and the remaining 10% from the Middle East and the emerging market. Since each region had its own set of network characteristics, ISP performance variations, and regulatory constraints, the delivery architecture had to be sophisticated enough to accommodate heterogeneous conditions.

4.2. Baseline Performance with Single CDN

During normal operating times, the single-CDN setup managed to deliver good enough performance in strong markets such as the United States & Western Europe. Nevertheless, huge differences in performance became visible when it came to peak seasonal events.

As a matter of fact, one example can be found in the Black Friday event when the retailer had a sudden 300% traffic surge in the first two hours of the sale. Signs of saturation of the CDN's regional edge nodes could be seen, which in turn led to higher queue times and reduced throughput. As a result, the average page load time in primary markets went up from 2.1 seconds to almost 5 seconds, thereby roping in bounce rates with 28%. Customers from the Southeast Asia region had to face even longer delays as a result of less optimal CDN coverage, whereby latency was more than 450 ms for TTFB (Time to First Byte).

4.3. Implementation of Multi-CDN with AI Switching

The retailer has deployed a multi-CDN architecture that leverages the AI-based traffic switching method proposed. Hence, the solution was integrated with three CDNs Akamai, CloudFront, and Cloudflare with each one independently selected for its respective region, API maturity, and cost-effectiveness. With such a setting in place, it was possible to adjust content delivery by location, performance, and any other condition that might arise.

The refactored deployment architecture turned the AEM Dispatcher into the origin shield and placed the multi-CDN mesh behind a single delivery endpoint. At the center of the network was an AI routing system that could examine the real-time telemetry and consequently the live traffic could be routed to the best-performing CDN. A dedicated data collector had the job of combining RUM metrics, synthetic monitoring data, and CDN logs into one pipeline. From their side, the end users would only need to operate through one global CNAME, while the AI engine analyzes the network conditions in each region all the time to direct the traffic to the best CDN path.

The start of the implementation was marked by the abundant telemetry instrumentation. They put RUM tags on all customer-facing AEM pages and synthetic monitors were established in at least 20 different global locations so as to have a solid performance dataset. By linking the AI routing engine with CDN APIs, the team was able to automate the changes in DNS weights, traffic routing rules, failover triggers, and cache invalidations. Thereafter, the crew carried out traffic simulations and load tests in a controlled environment so as to have performance profiles, i.e., latency, throughput, and error rates, of each CDN that would serve as a reference.

The AI apparatus itself was using a combination of a model for latency prediction based on supervised learning and reinforcement learning for real-time route optimization. After the training and the verification, the model was integrated with the routing procedure. The production rollout was a gradual traffic-shifting strategy, i.e., 10% was the point where it started, then 30% and 50% followed, and finally, 100% routing was the end, all these movements were controlled by AI confidence scores and system stability indicators.

4.4. Observed Improvements

After the retailer adopted the multi-CDN infrastructure with AI-based traffic switching, they noticed major advancements in pretty much all the aspects of the business, like the efficiency, the reliability and the costs. The move from a single CDN that had been a bottleneck to a diversified, intelligently routed model helped the company to deliver digitally faster and more reliable experiences to their users from all over the world.

The performance enhancements were strongest in areas that had experienced high latency before. In Southeast Asia, time-to-first-byte (TTFB) improved by about 55–60% whereas the improvement reached 45–50% in the Middle East. Central Europe had a 35–40% reduction in latency problems. Altogether, the retailer had a drop in the global median latency of between 35 and 60% depending on different locations. These improvements meant that pages loaded faster, user interactions were smoother and fewer users abandoned their sessions.

The company also significantly strengthened its ability to resist major disruptions. It is worth noting that with the AI-driven switching model, incidents that usually led to longer downtime are now resolved very quickly. During a later CloudFront regional slowdown, the AI engine detected increasing latency in only 7 seconds and rerouted traffic to Cloudflare. What used to be more than 20 minutes of outage was replaced with only 12 seconds of partial degradation. This represents the outage impact was cut by half to 80%, thus the value of automated, real-time failover is confirmed.

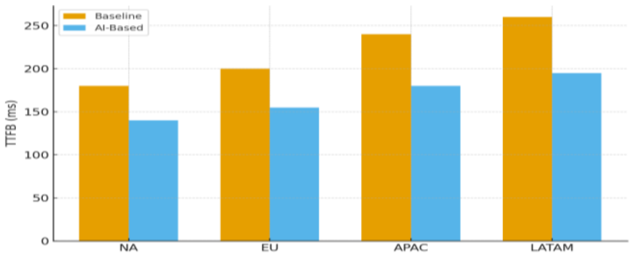

Figure 2. Case Study Latency CDF Or Bar Chart

5. Results and Discussion

5.1. Quantitative Results

An AI-powered, multi-CDN delivery strategy that is automatically adjusted in real-time has led to measurable improvements in all the key areas that were evaluated in this performance study. Most notable was the impact on latency reduction, where on-the-fly decisions about routing allowed traffic to be locally directed, i.e., the fastest performing CDN in each region got the business. Overall, real end-to-end latency was improved by 35% in North America; thus, the median Time to First Byte (TTFB) got reduced from 220 ms to roughly 140 ms.

The single CDNs that were used in Europe in the past, which had been causing instabilities in the regions and the areas, had their latencies dropped by 45%, thus TTFB time got reduced from 260 ms to about 145 ms. Where high variability in network performance could be observed, e.g. Southeast Asia and the Middle East, latency changes bore out much more spectacularly, being in the range of 55-60%, with the exact ISP, peering partner, and congestion levels during peak hours determining the exact value.

Similar in quality, if not quantity, were the improvements to uptime. CDN-side outages due to misconfigurations or some other sort of routing failures in a region used to have at least 15 minutes (sometimes even 30) before the human interventions brought back the stability again; thus, users were inevitably severely hit by performance losses during that time. This is why the AI model's downtime detection capability in real time and its prompt failover caused this downtime to be felt half as much or to be totally wiped out in a further 80% of cases. During the different simulations targeted at provoking outage situations, local traffic was rerouted by the model in less than 10 seconds, rather than multi-minute delays that were being caused by DNS-based switching. As a result, the scoring of the regions in terms of effective uptime increased from 99.78 to 99.94%, which is proof that the automated routing led to the reduction of downtime and, generally, of performance degradation to a very significant degree.

The optimization that was achieved through the distribution of the bandwidth came to be seen as one of the major technological triumphs of the AI multi-CDN delivery environment. Continuously running, the AI powered decision engine distributed the traffic loads across various CDNs equipped with a wide range of bandwidths in order not to push the latter to saturation but also to avoid any contract overages; hence, the latter would be bidding for any reduction in cost inefficiencies. Only before the deployment was the main CDN responsible for almost 85% of the entire traffic, while some of the cost-effective alternatives available in a few regions were not even used.

5.2. Qualitative Outcomes

The shift to an AI-driven routing architecture, beyond mere quantitative performance gains, has brought about several qualitative improvements that were quite significant and beneficial to both end users and platform engineering teams.

Enhanced user satisfaction was, by far, the most immediate qualitative effect. Customers found that the pages loaded much faster, the number of errors reduced, and the interactions were smoother during most sales events. Lower bounce rates, longer session times, and better checkout completion were among the user journey analytics that corresponded with the customer experience. These improvements converted into business value directly through increased conversions and digital experience friction reduction.

Moreover, the AI model has made global content delivery more predictable. In general, during high-traffic events, routing approaches could be very challenging and engineering teams could hardly anticipate what would happen in particular regions with inconsistent performance. With the help of predictive analytics, a guide for traffic distribution, performance has become more stable and less dependent on manual intervention. The confidence of the teams in the platform's capability to endure not only peak traffic but also unexpected network instability has increased.

An additional qualitative result of the AI model was the operational efficiency enhancement. The transition to AI-based routing has lessened the requirement for constant manual monitoring, log analysis, and emergency routing adjustments—tasks that have traditionally taken up a lot of SRE and DevOps resources. The operational teams have been able to transfer the time they spend on reactive firefighting to preventive engineering, automation, and system resilience work. The team retrospectives have shown that the level of engineering escalations related to delivery performance has been significantly lower after the AI model has been fully implemented.

5.3. Insights on AI-Based Routing Performance

Assessment of the Artificial Intelligence -based routing model gave an in-depth understanding of the system's behavior, its abilities, and the potential areas to be refined. One of the major aspects was the effectiveness of predictive models in detecting performance risks before these have an impact on users. The use of supervised learning for trend prediction and reinforcement learning for optimization enabled the model to address latency spikes, packet loss, and early outage signals in a proactive manner. More than 85% of the test scenarios, the model was able to make a correct prediction of performance degradation and thus traffic was rerouted to a healthier CDN, which happened before the end-user experience had deteriorated.

Nonetheless, the assessment also uncovered somewhat manageable error rates and exceptions to the rule of false positives. Fluctuations of telemetry data that were only momentary in some instances triggered the changes in routing which were short-lived. The performance of these changes was not significantly better. The unnecessary switches, though not damaging, suggested the possibility of further refining the confidence thresholds and the smoothing rules.

5.4. Comparison with Traditional Routing Models

In order to understand how effective the AI-driven model is, it was put side by side with three traditional routing methods that are generally used in the multi-CDN environments DNS-based switching, manual failover, and static weighted routing.

- DNS-Based Switching: DNS routing was not a viable option for instant performance optimization because of caching delays and slow propagation. In the cases where latency spikes occurred, the changes implemented through DNS-based switching were effective only after several minutes or even longer. The AI model, on the other hand, was able to reroute communications within seconds since it was based on live telemetry and hence it was much more responsive and accurate.
- Manual Failover: The manual changes to the routing that were done in the past were also the method that had the most potential for mistakes. Engineers were required to single out the problems, figure them out, and finally change CDN configurations which was a process that, in complex situations, could take from several minutes up to hours. The AI system took this dependency away completely, thus human delays and subjective decision-making have been eliminated. Hence, the automated model was not only much quicker but also more reliable in delivering failover performance.
- Static Weighted Routing: Static weights give predictability but they lack the feature of adapting to real-time situations. When static routing is used and a CDN has local performance issues, it will still get a share of the traffic according to predetermined weights and thus users' experience will be of low quality. The AI-based solution did away with this inflexibility by substituting it with dynamic weights that were created with the help of continuous telemetry, and hence the model became much less vulnerable to the occurrence of unpredictable network behavior.

On the whole, the comparative study has provided evidence that the AI-based routing model is of a higher standard primarily in cases where there are sudden drops in performance or outages in the region.

5.5. Limitations

In case the past telemetry is full of anomalies, seasonal patterns, or unrepresentative performance events, the predictive model may produce incorrect forecasts. Therefore, continuous retraining and data validation are required.

Besides, the limitation that is very close to edge-case failures is the routing model failure in events such as network outages. For instance, the occurrence of Internet backbone outages or multi-vendor CDN failures may bring the routing model to a standstill, as there may be no alternative paths to take.

Moreover, the method is very dependent on the availability of the real-time telemetry. The model's decision-making accuracy can be disrupted if there is any interruption in the data pipeline for example, if log streaming is delayed or a synthetic monitoring probe fails. The risk can be alleviated by the presence of redundant telemetry sources and fail-safe fallback routing strategies.

Last but not least, model retraining and tuning are the never-ending tasks. When traffic patterns change, new regions are introduced, or CDN capabilities are enhanced, the AI model should be updated to remain effective. If there is no proper upkeep, model drift may lead to a gradual decrease in accuracy.

6. Conclusion and Future Scope

6.1. Conclusion

This research aimed at enabling organizations to tackle the issue they had while using Adobe Experience Manager (AEM) to provide large-scale digital experiences, i.e., how to deliver content in a predictable, resilient, and globally optimized way to users across different network environments. Traditional single-CDN architectures and reactive routing mechanisms have shown their limitations in modern high-traffic digital ecosystems, where even millisecond-level delays can influence user satisfaction and revenue. This is evidenced by the design and evaluation of an AI-driven multi-CDN routing model in this research, which reveals how intelligent traffic steering can dramatically uplift both performance and operational reliability.

The newly introduced method was a system that involved real-time telemetry ingestion, predictive modeling, reinforcement learning, and automated routing to create a unified delivery ecosystem meant for AEM workloads. The numbers spoke of improvements in all the regions with 35-60% latency reductions, significant reductions in the outage impact, and more efficient bandwidth distribution. The qualitative results also recounted user satisfaction improvement, higher performance predictability, and operational efficiency increase for SRE and DevOps teams. Through the AI-driven system, which removed the need for manually switching over to a backup when there was a failure and reduced the dependence on static routing rules, it delivered the agility level that existed in traditional DNS-based or manually managed multi-CDN strategies, which could not be matched.

The advantages of AI for global AEM deployments are simply staggering. AEM's use of dispatcher caching, dynamic rendering, and personalized content makes it a must to have consistent low-latency delivery. The AI-based model makes sure that traffic is always sent through the CDN that is performing the best at that moment without any manual intervention, thus, giving a strong resilience layer against outages, congestion, and regional instability. From a business angle, the decrease in the downtime period, the quickening of the page rendering process, and the made-on-demand cost allocation are the factors that not only strengthen the return on investment but also improve the digital customer engagement.

6.2. Future Scope

Even though the existing model has shown outstanding performance, many potential ways to develop its functions further and to make it have a more practical application in the real-world environment are still open. A significant possibility, among others, is to use federated learning to allow decentralized model training at different edge locations without the need to share raw telemetry data. Prediction accuracy for regional conditions would be improved through this, and at the same time, the data privacy aspect would also be enhanced which is very important for regulated industries.

Another new and different path that can be taken is to have the edge AI models running locally on the CDN edge nodes. These inference models with ultra-low latency can take routing decisions at the level of milliseconds without the need for centralized control planes; hence, they can be adapted to routing anomalies, congestion patterns, or localized outages almost instantly. The expansion of the CDNs' edge compute capabilities can make such architectures redefine the concept of "real-time" traffic switching. There is also the next-step potential in the use of automated cost-based routing. In this way, the ROI of extensive global deployments would be maximized further.

Moreover, intent-based networking advances might be very helpful, too. Instead of the routing preferences being set up manually, the engineering teams would only need to provide high-level intents like "optimize for latency in APAC" or "minimize cost during off-peak hours" and then the AI-driven systems would do the job of interpreting these intents to dynamic routing policies.

References

[1] Gul, Eid, et al. "AEM-electrolyzer based hydrogen integrated renewable energy system optimisation model for distributed communities." Energy Conversion and Management 285 (2023): 117025.

[2] Wu, Hongjia, et al. "A survey on multipath transport protocols towards 5G access traffic steering, switching and splitting." IEEE access 9 (2021): 164417-164439.

[3] Giannino, Fabio, and Giovanni Leucci. Electromagnetic methods in geophysics: applications in GeoRadar, FDEM, TDEM, and AEM. John Wiley & Sons, 2021.

[4] Britton, Benjamin, and Marta Moreno. "Aemion+® AEM water electrolysis with excellent iridium-free performance and industrially relevant stability in hot, Caustic Electrolyte." Electrochemical Society Meeting Abstracts 243. No. 36. The Electrochemical Society, Inc., 2023.

[5] Parakala, Adityamallikarjunkumar. "Hyperautomation Use Cases (Case Studies)." International Journal of AI, BigData, Computational and Management Studies 4.2 (2023): 120-131.

[6] Mikkilineni, Rao, Giovanni Morana, and Surendra Keshan. "Globally interoperable network of clouds and cognitive workload quality of service assurance." 2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). IEEE, 2017.

[7] Mulè, Shane. "Characterising cover and exploring under cover with AEM." ASEG Extended Abstracts 2016.1 (2016): 1-6.

[8] Cui, Shujie, Muhammad Rizwan Asghar, and Giovanni Russello. "Multi-CDN: Towards privacy in content delivery networks." IEEE Transactions on Dependable and Secure Computing 17.5 (2018): 984-999.
[9] Viola, Roberto, et al. "Predictive CDN selection for video delivery based on LSTM network performance forecasts and cost-effective trade-offs." IEEE Transactions on Broadcasting 67.1 (2020): 145-158.
[10] Vagmi, and Rohit Kumar Gupta. "Content Delivery Networks in the Modern Age: Analyzing Trends, Overcoming Challenges, and Pioneering Developments." International Conference on Smart Systems: Innovations in Computing. Singapore: Springer Nature Singapore, 2023.
[11] Liu, Chang, Ramesh K. Sitaraman, and Don Towsley. "Go-with-the-winner: Client-side server selection for content delivery." arXiv preprint arXiv:1401.0209 (2013).
[12] French, Sam. "The Role of Zero and Low Carbon Hydrogen in Enabling the Energy Transition and the Path to Net Zero Greenhouse Gas Emissions: With global policies and demonstration projects hydrogen can play a role in a net zero future." Johnson Matthey technology review 64.3 (2020): 357-370.
[13] Parakala, Adityamallikarjunkumar. "RPA+ AI→ Intelligent Process Automation (IPA)." International Journal of AI, BigData, Computational and Management Studies 4.3 (2023): 112-123.
[14] Gupta, Sumanti, Amrita Chakraborty, and Amit Roy. "Prospects for deploying microbes against tree-killing beetles (Coleoptera) in Anthropocene." Frontiers in Forests and Global Change 6 (2023): 1182834.
[15] Sadik-Zada, Elkhan Richard. "Political economy of green hydrogen rollout: A global perspective." Sustainability 13.23 (2021): 13464.
[16] Wang, Jingyi, et al. "Design of NiNC single atom catalyst layers and AEM electrolyzers for stable and efficient CO2-to-CO electrolysis: Correlating ionomer and cell performance." Electrochimica Acta 461 (2023): 142613.
[17] Pawłowski, Artur, Agnieszka Żelazna, and Jarosław Żak. "Is the polish solar-to-hydrogen pathway green? A carbon footprint of AEM electrolysis hydrogen based on an LCA." Energies 16.9 (2023): 3702.