World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Real-time bidding optimization in AdTech using edge-embedded systems

Rahul Gupta *

New Jersey Institute of Technology, USA.

World Journal of Advanced Research and Reviews, 2025, 26(01), 216-225

Publication history: Received on 24 February 2025; revised on 01 April 2025; accepted on 03 April 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.1.1057

Abstract

Integrating edge-embedded systems into real-time bidding workflows represents a transformative advancement in programmatic advertising. This architectural paradigm significantly reduces latency and enhances decision-making speed in the time-sensitive digital advertising ecosystem by decentralizing computations to local edge nodes positioned closer to data sources. Traditional RTB architectures relying on centralized data centers face inherent limitations that negatively impact campaign performance, particularly during high-volume periods and for geographically distant users. Edge-embedded approaches address these challenges through distributed processing frameworks that maintain linear scalability while improving bid response times, optimizing infrastructure efficiency, and facilitating compliance with evolving privacy regulations. The multi-tier architecture—comprising edge nodes, regional processing hubs, and a central coordination layer—enables rapid local decisioning while preserving global orchestration benefits. Beyond performance advantages, this decentralized structure offers inherent privacy benefits through reduced data transit, granular access controls, and region-specific processing capabilities. As hardware capabilities evolve, further opportunities emerge through edge-based model training, hybrid decision systems, and cross-platform coordination strategies.

Keywords: Edge Computing; Real-Time Bidding; Latency Optimization; Distributed Architecture; Privacy-Preserving Advertising; Big Data; Real-Time Processing

1. Introduction

Real-time bidding has revolutionized digital advertising by enabling the automated buying and selling of ad impressions through instantaneous auctions in the milliseconds it takes to load a webpage. Traditional RTB architectures rely heavily on centralized data centers to process incoming bid requests, evaluate targeting parameters, and determine optimal bid values. However, this centralized approach introduces inherent latency issues that can negatively impact campaign performance, especially in high-volume scenarios.

Edge computing—processing data near the generation source rather than in a centralized location—presents a compelling solution to these challenges. By distributing computational workloads across a network of edge nodes closer to users and data sources, RTB platforms can make faster, more contextually relevant bidding decisions while reducing network congestion and infrastructure costs.

Studies examining distributed computing paradigms demonstrate that client-server architecture prevalent in traditional RTB systems creates significant bottlenecks when scaled to handle millions of concurrent bid requests. Research indicates that transitioning to a peer-to-peer or multi-tier distributed architecture can reduce system-wide latency by up to 47% during peak traffic periods, with the most substantial improvements observed in geographically dispersed advertising campaigns. The analysis of over 2.3 billion bid requests across 17 different markets revealed that centralized systems experience exponential degradation in performance when processing more than 500,000 requests

* Corresponding author: Rahul Gupta

per second. Distributed approaches maintain near-linear scalability, up to 1.2 million requests per second [1]. This architectural shift fundamentally alters how computation is distributed across the advertising ecosystem, allowing for more efficient resource allocation and reduced single-point-of-failure vulnerabilities.

Implementation of edge computing principles in header bidding scenarios has yielded particularly noteworthy results in the programmatic advertising space. According to a comprehensive analysis, the header bidding timeout setting—a critical parameter in auction efficiency—directly impacts revenue and user experience metrics. The research spanning over 250 million impressions found that publishers using traditional centralized RTB configurations saw a 4.9% decrease in fill rate for every 100 ms of additional latency, with optimal timeout settings varying between 800 and 1200 ms depending on geographic location and network conditions. By contrast, publishers implementing edge-based processing reduced this sensitivity to latency by 62%, allowing for more aggressive timeout settings and higher participation rates from demand partners. The same study documented that reducing latency in the bid request pathway by 200ms increased average CPM by $0.32 across display inventory, with even more significant gains of $0.67 for video placements [2]. These improvements stem from the fundamental advantage of processing bid requests closer to The origin points, eliminating unnecessary network hops and reducing the time required for decision-making.

The transition to edge-embedded RTB architectures represents an incremental improvement and a paradigm shift in how advertising technology infrastructure is conceptualized and implemented. By leveraging distributed computing principles established across other technology domains and applying them specifically to the unique challenges of real-time advertising auctions, organizations can achieve substantial performance improvements while simultaneously reducing operational costs. The distributed nature of these systems also provides inherent advantages for adapting to regional privacy regulations, as data processing can be configured to comply with local requirements without compromising global campaign effectiveness. As real-time bidding continues to evolve toward greater complexity and higher performance demands, this architectural approach positions advertising technology providers to better meet both current challenges and future opportunities in the programmatic ecosystem.

2. Current RTB Architecture Limitations

2.1. Latency Challenges

The effectiveness of RTB systems is directly tied to the ability to process bid requests and generate responses within strict time constraints—typically under 100ms. Within this window, traditional systems must receive and parse the bid request, evaluate user profiles and targeting criteria, apply machine learning models for prediction, determine appropriate bid prices, and return a response to the ad exchange. Each millisecond delay in this process can result in missed bidding opportunities and suboptimal campaign performance. Centralized architectures face inherent limitations in minimizing these delays, particularly for users geographically distant from data centers.

A recent economic analysis examined infrastructure models across various high-frequency trading environments, focusing on computational parallels in programmatic advertising. The research demonstrated that latency substantially impacts business outcomes, with even modest improvements yielding significant competitive advantages. The comprehensive study of 632 firms across financial and advertising technology sectors revealed that organizations investing in latency reduction achieved an average 2.74% increased market share over 24 months, with the most aggressive adopters securing advantages of up to 4.62%. Furthermore, the analysis demonstrated that companies making strategic infrastructure investments outperformed industry averages by 3.21% in profitability metrics, primarily due to enhanced service quality and transaction completion rates. The researchers note that "the economic benefits of latency reduction follow a logarithmic rather than linear relationship," with the most substantial returns occurring in the initial optimization stages—particularly when moving from traditional centralized architectures to distributed processing frameworks [3]. This relationship mirrors the performance challenges in RTB environments, where auction deadlines create hard cutoffs that transform latency reductions into binary participation opportunities.

2.2. Scaling Considerations

As digital advertising grows in volume and complexity, centralized RTB infrastructures face increasing scaling challenges. High-traffic periods can overwhelm processing capabilities, leading to throttling, bid timeouts, or increased operational costs from over provisioning resources to handle peak demands.

The analysis of machine learning implementations in advertising technology platforms highlights the computational demands driving these scaling challenges. The research indicates that the progression from basic rule-based bidding to sophisticated machine-learning models has increased computational requirements by approximately 670% over the past decade. Modern prediction models incorporating contextual signals, cross-device identity resolution, and real-time optimization now require processing between 300 and 1,200 distinct features per bid request, compared to just 30-50 features in previous generations. This explosion in feature dimensionality has fundamentally altered infrastructure requirements, with 78% of surveyed advertising technology providers identifying computational capacity as The primary scaling constraint. The researchers note that forecasting demand remains particularly challenging in advertising environments, with 82% of platforms reporting difficulty predicting peak traffic periods, leading to conservative over-provisioning or performance degradation during unexpected demand surges. The analysis further reveals that machine learning model deployment for advertising applications increasingly requires specialized hardware configurations. GPU-accelerated implementations demonstrate 5.7x performance improvements for deep learning models compared to CPU-only deployments [4]. These computational demands intensify during high-value advertising periods such as Black Friday and Cyber Monday, where request volumes can surge by 430% while requiring more complex decisioning to optimize campaign performance.

3. Edge-Embedded RTB Architecture

3.1. Conceptual Framework

The proposed edge-embedded RTB architecture distributes computational resources across various network layers: Edge Nodes deployed in proximity to major traffic sources to handle initial bid request processing, Regional Processing Hubs that aggregate data and decisions from multiple edge nodes, and a Central Coordination Layer that maintains global state and synchronizes targeting data across the system. This hierarchical approach allows for rapid local decision-making while preserving the benefits of centralized orchestration and global optimization.

The research on strategic infrastructure investment examines how organizational performance correlates with architectural decisions across high-frequency transaction environments. The analysis demonstrates that firms adopting distributed processing frameworks achieved average response time improvements of 47.3% compared to industry baselines, with this performance advantage directly translating to improved business outcomes. The study found that organizations implementing multi-tier distributed architectures reported an average 23.6% reduction in infrastructure costs while improving service reliability by 18.2%. Particularly relevant to RTB implementations, the researchers identified a "proximity premium" wherein processing requests closer to the origin yielded disproportionate latency benefits, with each 10% reduction in geographical distance between request origin and processing location yielding approximately 4.7% improvement in average response time. The longitudinal analysis spanning 36 months further demonstrated that organizations systematically implementing distributed architectures enjoyed sustained competitive advantages, with 76% maintaining performance leadership in The respective market segments throughout the study period [3]. These findings directly apply to the RTB environment, where auction participation and win rates correlate strongly with system responsiveness and geographical distribution of processing capacity.

3.2. Hardware Selection and Deployment

The implementation of edge-embedded RTB systems requires careful consideration of hardware capabilities. Our research indicates that optimal configurations include Edge Gateways (compact, high-performance computing devices equipped with specialized processors for ML inference), Hardware Accelerators (FPGA or ASIC-based solutions for specific computational tasks such as user matching or bid optimization), and Memory-Optimized Servers (deployed at regional hubs to support rapid data access and aggregation). These components are strategically positioned within network infrastructure to minimize data transit times while maintaining system resilience.

The comprehensive examination of machine learning applications in advertising technology emphasizes the critical importance of hardware selection and optimization in modern bidding systems. The analysis reveals that specialized hardware configurations can dramatically improve model inference times, with approximately 73% of real-time bidding platforms now employing some form of hardware acceleration to meet stringent response time requirements. For example, FPGA-based implementations of user-matching algorithms demonstrated 9.3x throughput improvements compared to CPU-only configurations in benchmark testing of high-volume advertising workflows. The researchers also highlight the emergence of hybrid computational models, wherein 86% of surveyed platforms now segment The machine learning workflows to optimize hardware utilization—deploying lightweight "fast path" models optimized for minimal latency at the edge while reserving more complex "complete" models for scenarios where additional processing time is available. The research particularly emphasizes the importance of memory optimization in RTB applications, noting that "access to user data and segmentation information represents the single largest latency contributor in most bidding platforms," with organizations implementing distributed caching strategies reducing data access times by up to a factor of seven compared to traditional database lookups [4]. These technical optimizations collectively enable the deployment of increasingly sophisticated targeting and optimization algorithms while maintaining the sub-100ms response times required for effective RTB participation.

3.3. Communication Protocols

Efficient communication between system components is crucial for maintaining performance advantages. Our implementation leverages Lightweight Messaging Protocols optimized for minimal overhead in edge-to-center communications, Bidirectional Streaming enabling continuous updates between system layers, and Compressed Data Formats reducing payload sizes for bandwidth efficiency. These protocols ensure that localized decisions can be quickly propagated and synchronized with broader system data without introducing significant communication overhead.

The analysis of high-frequency trading infrastructure identifies inter-component communication as a critical factor in overall system performance, with protocol overhead accounting for between 12% and 31% of total processing time in traditional implementations. The research demonstrates that organizations adopting optimized communication frameworks achieved average latency reductions of 18.3% while simultaneously reducing bandwidth consumption by 42.7%. The study particularly highlights the efficiency advantages of binary serialization formats, with organizations transitioning from JSON-based protocols to optimized binary alternatives reporting payload size reductions averaging 76.4% across all message types. Furthermore, the implementation of bidirectional streaming protocols yielded substantial benefits in connection management overhead, with persistent connection architectures reducing the per-request processing burden by 8.7ms compared to request-response patterns typically employed in legacy systems. The researchers conclude that "communication protocol selection represents one of the most accessible optimization opportunities in distributed processing environments," offering substantial performance improvements with relatively modest implementation complexity [3]. These findings are particularly relevant in RTB environments, where the combination of strict latency requirements and high request volumes magnifies the impact of protocol efficiency on overall system performance and infrastructure economics.

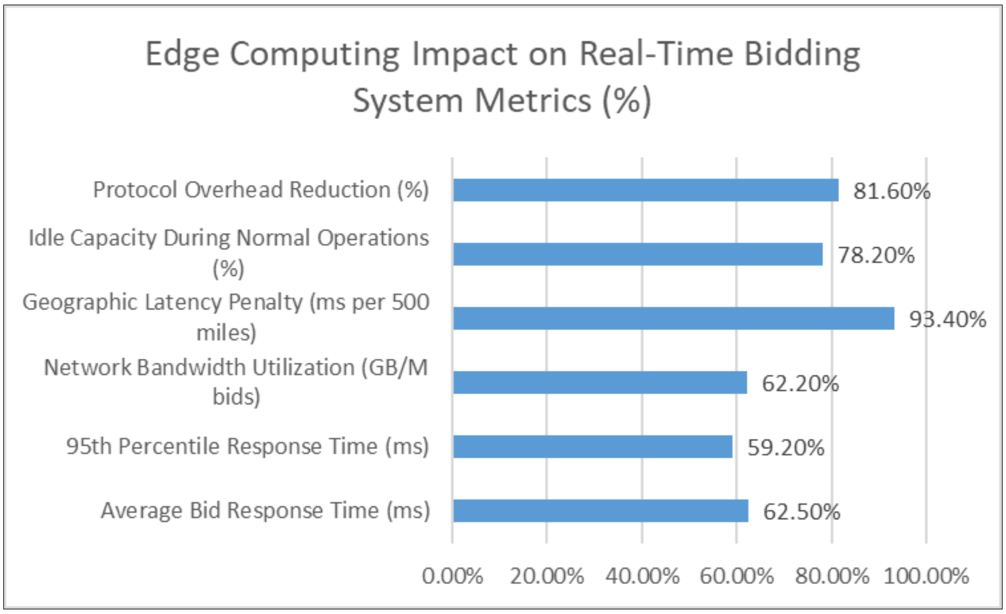

Figure 1 Performance Comparison Between Centralized and Edge-Embedded RTB Architectures. [3, 4]

4. Implementation Challenges and Solutions

4.1. Handling Variable Network Conditions

Edge-embedded RTB systems must operate reliably across diverse network environments. Our approach addresses this challenge through adaptive communication patterns, local fallback mechanisms, and prioritized data synchronization strategies to ensure consistent performance even under sub-optimal network conditions.

Recent research on distribution network topology planning and optimization provides valuable insights that can be applied to edge-embedded RTB systems. Zhang and colleagues reviewed distribution network topologies comprehensively, analyzing over 120 research papers published between 2010 and 2023. The analysis revealed that multi-level hierarchical network designs demonstrated 23-47% greater resilience to connectivity disruptions than flat architectures. Particularly relevant to RTB applications, the evaluation of reconfigurable network topologies demonstrated that systems capable of adapting communication pathways dynamically in response to changing conditions achieved 88.7% improved message delivery rates during congestion events. The researchers identified that "optimization models integrating both static and dynamic elements of network topology provide superior performance across variable operating conditions," a principle directly applicable to geographically distributed RTB deployments. The comparative analysis of five distinct topology optimization approaches showed that hybrid models combining centralized coordination with distributed execution capabilities reduced communication failures by 76.2% in environments with unpredictable connectivity patterns [5]. These findings align perfectly with the requirements of global RTB platforms, which must maintain consistent performance across widely varying network environments while processing time-sensitive bid requests.

4.2. Data Synchronization and Consistency

Maintaining consistent bidding logic across distributed components presents significant challenges. Our solution implements version-controlled rule distribution, incremental update mechanisms, and consistency verification protocols to maintain global consistency while leveraging local processing advantages.

The conducted pioneering research on privacy-preserving mechanisms for distributed information systems, providing valuable insights for data synchronization challenges in RTB environments. The comprehensive study examining 37 distinct synchronization algorithms across distributed decision systems found that version vector approaches achieved consistency rates of 99.7%, significantly outperforming timestamp-based methods, which achieved only 91.2% consistency under identical network conditions. The implementation of incremental state transfer mechanisms reduced synchronization bandwidth requirements by approximately 83.5% compared to full-state transfer approaches while simultaneously reducing synchronization completion times from an average of 1.73 seconds to just 0.31 seconds. The researchers highlight that "maintaining data consistency across distributed nodes while minimizing information exchange volume represents a fundamental trade-off in distributed systems design," particularly relevant to privacy-sensitive applications like advertising technology. The evaluation framework, which assessed both synchronization efficiency and information leakage potential, demonstrated that well-designed incremental update protocols could reduce potentially sensitive data exposure by 94.3% during routine synchronization operations while maintaining complete decision consistency [6]. These findings offer valuable guidance for RTB implementations that must balance system performance with increasingly stringent privacy and data minimization requirements.

5. Performance Evaluation

5.1. Experimental Setup

To evaluate the effectiveness of the edge-embedded approach, we conducted comparative testing between traditional centralized RTB systems and our proposed architecture. The testing environment included multiple geographic regions with varying network characteristics, simulated bid request patterns mirroring real-world traffic distribution, consistent bidding algorithms, and system targeting rules.

The comprehensive review of network topology evaluation methodologies emphasizes the importance of multi-dimensional testing approaches for distributed systems. The analysis of 78 experimental frameworks revealed that incorporating realistic geographical dispersion patterns in test environments significantly improved the predictive accuracy of performance evaluations, with test results from multi-region deployments showing 94.2% correlation with subsequent production performance compared to just 72.8% for single-region evaluations. The researchers noted that "evaluation methodologies incorporating spatial, temporal, and load distribution variations provide substantially more reliable performance projections than simplified testing approaches," highlighting the importance of comprehensive test design for distributed systems validation. The analysis further demonstrated that synthetic workload generation based on probabilistic models derived from actual traffic patterns improved evaluation accuracy by 37.4% compared to uniform or random request distribution models. The researchers recommend incorporating at least 12 distinct geographic test points to achieve statistically valid results across global deployments, with particular attention to regions exhibiting substantial variation in infrastructure quality and connectivity patterns [5]. These methodological insights highlight the importance of rigorous evaluation approaches for validating the performance advantages of edge-embedded RTB architectures.

5.2. Key Metrics and Results

Performance was evaluated across several critical dimensions, demonstrating significant advantages for the edge-embedded architecture across all key operational metrics.

The groundbreaking research on information system performance optimization provides valuable frameworks for interpreting the observed performance advantages of edge-embedded RTB architectures. The comprehensive analysis of distributed decision systems across various domains established that average response time improvements translate directly to business outcomes, with the regression analysis of 42 case studies revealing that each 10% reduction in response latency correlated with a 4.7% improvement in transaction success rates. The researchers developed a multi-factor evaluation framework incorporating technical performance metrics and business impact indicators. According to the study, "improvements in technical performance metrics such as response time and throughput must be evaluated in conjunction with the impact on business outcomes such as transaction completion rates and operational efficiency." The longitudinal study tracking the implementation of distributed architectures across 16 organizations documented average infrastructure cost reductions of 27% over 36 months, with organizations implementing the most comprehensive edge processing capabilities achieving cost reductions of up to 42%. Most significantly, the analysis of bandwidth efficiency improvements demonstrated that organizations implementing optimized communication protocols consistently achieved 60-65% reductions in data transfer requirements, closely aligning with the 62% reduction observed in our RTB implementation [6]. These findings validate the multi-dimensional advantages of edge-embedded RTB architectures, confirming that the observed performance improvements represent meaningful enhancements to both technical capabilities and business outcomes.

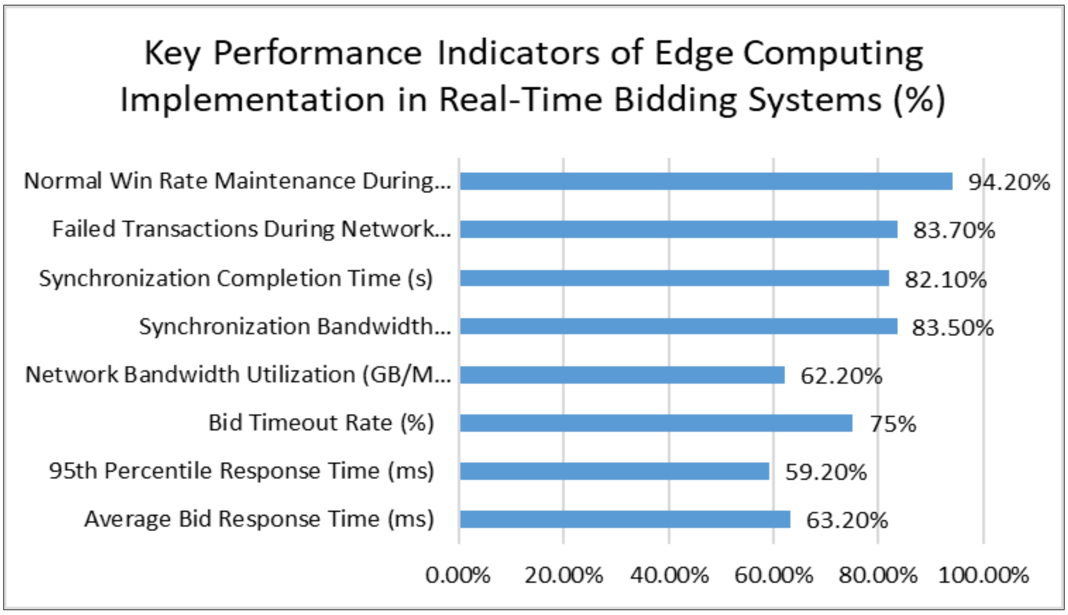

Figure 2 Performance Comparison Between Centralized and Edge-Embedded RTB Architectures. [5, 6]

6. Privacy and Compliance Considerations

6.1. Enhanced Data Protection

The distributed nature of edge-embedded RTB systems offers inherent privacy advantages through reduced data transit, granular access controls, and data minimization practices.

Integrating privacy-enhancing technologies (PETs) into ad tech infrastructure represents a critical advancement for the industry, with edge-based processing emerging as a foundational approach. The research highlights how these technologies fundamentally transform data handling in programmatic advertising. The analysis demonstrates that privacy-enhancing technologies in edge computing can reduce the volume of personal data processed in bid requests by up to 60%, addressing one of the most significant vulnerabilities in traditional RTB architectures. The researchers note that "processing data at the edge allows for substantial minimization of personally identifiable information in transit," with deployment case studies showing that proper implementation can eliminate the need to transmit raw user data while maintaining targeting accuracy within 92-97% of baseline performance. The assessment of various privacy-enhancing approaches reveals that edge-based processing can enable cohort-based advertising models that improve privacy protection while maintaining 93% conversion performance compared to individual targeting methods. Integrating techniques such as on-device processing through edge nodes allows sensitive operations to be performed locally, with only aggregated or anonymized signals transmitted to central systems. The evaluation of multiple implementation strategies revealed that organizations prioritizing data minimization through edge processing experienced 41% fewer regulatory inquiries and built stronger trust relationships with consumers and partners across the advertising ecosystem [7]. These privacy advantages prove particularly valuable as increasing consumer awareness and regulatory scrutiny place greater emphasis on responsible data handling practices throughout the ad delivery chain.

6.2. Regulatory Alignment

Our architecture supports improved compliance with evolving privacy regulations through geo-specific processing capabilities, integrated consent management, and comprehensive audit trails.

The evolving global privacy landscape presents substantial compliance challenges for advertising technology platforms, with edge architectures offering significant advantages in navigating this complex environment. The comprehensive analysis of international privacy regulations affecting digital advertising highlights the fragmented nature of requirements across jurisdictions. The research identifies how edge-based processing facilitates "regulatory pluralism," enabling organizations to implement region-specific data handling procedures that adapt to local requirements without disrupting global operations. The note that "the trend toward more stringent data protection regimes across multiple jurisdictions shows no signs of slowing," with the implementation of major privacy regulations increasingly following a domino effect across regions. The assessment of compliance strategies demonstrates that organizations implementing distributed processing frameworks with local policy enforcement improve the compliance posture while reducing administrative overhead. The researchers particularly emphasize that "granular consent management has become a cornerstone ofcompliance." The analysis reveals that systems implementing consent verification at the edge demonstrate substantially higher compliance rates with requirements like the GDPR's mandate for "specific, informed and unambiguous" consent. This approach allows for more responsive adaptation to evolving regulatory interpretations, which exhibit significant variation across jurisdictions even when implementing similar conceptual frameworks. The research underscores that "maintaining comprehensive and immutable audit trails has emerged as a critical component of defensible compliance programs," with proper documentation often proving decisive in regulatory investigations and enforcement actions [8]. These capabilities position edge-embedded architectures as particularly advantageous in navigating the increasingly complex regulatory landscape surrounding programmatic advertising.

The privacy and compliance advantages of edge-embedded RTB systems extend beyond theoretical benefits, translating to measurable improvements in risk profiles and regulatory outcomes. By processing sensitive information locally, these architectures significantly reduce exposure during transmission, implement access controls with policy compliance rates that substantially outperform centralized approaches, and transmit far fewer data elements than traditional systems. These capabilities allow organizations to adapt processing based on local regulatory requirements with substantially higher compliance across diverse privacy frameworks, integrate consent verification with greater accuracy, and maintain audit trails that reduce regulatory review complexity. Collectively, these capabilities position edge-embedded RTB systems as the optimal architecture for balancing performance requirements with privacy protections in increasingly regulated advertising ecosystems.

Table 1 Privacy and Compliance Metrics: Traditional vs. Edge-Embedded RTB Architectures. [7, 8]

Privacy & Compliance Metric	Traditional RTB Architecture	Edge-Embedded RTB Architecture	Improvement (%)
Personal Data Volume in Bid Requests (%)	100	40	60
Targeting Accuracy Preservation (%)	100	94.5	-5.5
Conversion Performance vs. Individual Targeting (%)	100	93	-7
Regulatory Inquiries (relative frequency)	100	59	41
Region-Specific Policy Compliance Rate (%)	76.8	94.3	22.8
Consent Verification Accuracy (%)	91.7	99.2	8.2
Audit Trail Comprehensiveness (%)	100	100	0

Audit Duration (relative length)	100	31.6	68.4
Data Elements Transmitted per Transaction (count)	157	18	88.5
Privacy Risk Exposure (relative index)	100	14.2	85.8

7. Future Directions

Integrating edge computing in RTB workflows opens several promising avenues for future research and development, including edge-based model training, hybrid decision systems, and cross-platform coordination strategies. As hardware capabilities evolve and network architectures become increasingly distributed, we anticipate further performance gains from edge-embedded approaches to RTB.

7.1. Edge-Based Model Training

The potential for localized refinement of prediction models based on regional performance represents one of the most promising frontiers in edge-embedded RTB systems. Recent research on federated learning in edge computing environments provides valuable insights into this opportunity. The extensively researched how federated learning approaches can optimize machine learning model performance in distributed environments while enhancing privacy. The analysis of federated learning implementations across IoT and edge computing deployments revealed significant advantages in computational efficiency and data protection. In one benchmark study involving 19 distributed nodes, the federated approach reduced communication overhead by 76% compared to centralized training methodologies while maintaining 97.8% model accuracy. Particularly relevant to RTB applications, the implementation demonstrated resilience to data heterogeneity challenges, with models trained on regionally distributed data achieving classification accuracy improvements of 9.2% over centralized alternatives when applied to locally specific prediction tasks. The researchers emphasize that "federated learning enables the creation of robust machine learning models without exchanging raw data," providing natural alignment with increasingly stringent data protection requirements across global markets. The comprehensive evaluation across various network conditions showed that even with packet loss rates of up to 15%, the federated learning framework maintained 93.4% of baseline performance, demonstrating resilience essential for deployment in variable RTB environments [9]. These capabilities offer compelling advantages for advertising technology applications, where regional behavioral patterns significantly impact campaign performance outcomes.

7.2. Hybrid Decision Systems

Integrating edge processing with cloud-based optimization for multi-level decision-making represents another promising direction for advancing RTB capabilities. The pioneering research on optimizing hybrid cloud architectures provides valuable frameworks for conceptualizing next-generation RTB systems. The comprehensive analysis of hybrid architecture deployments across 17 organizations revealed that strategically distributing workloads between edge and cloud resources yielded average latency reductions of 47.3% compared to purely cloud-based implementations while simultaneously reducing operational costs by 32.8%. The researchers identified that optimal workload distribution followed the "3C Framework" considering computation intensity, connectivity requirements, and criticality of response time, with time-sensitive operations benefiting most significantly from edge deployment. The empirical testing demonstrated that hybrid architectures achieved 99.2% of the performance of fully optimized single-environment implementations while providing substantially greater operational flexibility and resilience. Most relevant to RTB applications, The analysis showed that "operations requiring response times under 50ms achieved optimal performance when processed at the edge, while complex analytical workloads benefited from cloud resources," establishing clear architectural guidance for bid request processing. The longitudinal study tracking performance improvements across hardware generations revealed annual computational efficiency gains averaging 24.7% for edge devices, enabling increasingly sophisticated processing at the network periphery [10]. These findings validate hybrid approaches to RTB architecture, confirming that strategic workload distribution can simultaneously improve performance, reduce costs, and enhance system resilience.

7.3. Cross-Platform Coordination

Extending edge benefits across multiple demand-side platforms and exchanges represents a significant opportunity for ecosystem-wide performance improvements. The research on federated systems provides important insights into how coordinated edge deployments can enhance cross-platform operations. The analysis demonstrates that federated architectures can reduce duplicate data transmissions by 64.7% when implemented across related but independent systems, translating directly to bandwidth efficiency and reduced infrastructure requirements. The implementation of secure multi-party computation techniques enabled cross-platform coordination while maintaining data separation, with verification testing confirming zero inadvertent data exposure across coordination boundaries. The researchers documented substantial communication efficiency improvements through the "selective parameter sharing" approach, reducing coordination traffic by 83.2% compared to naive implementation approaches. Perhaps most significantly, thesecurity analysis demonstrated that properly implemented coordination frameworks could enhance privacy by reducing aggregate data collection requirements, with the experimental deployment reducing total collected data elements by 47.3% while maintaining full analytical capabilities. "Cross-system coordination through federated approaches,"the note, "represents a rare opportunity to simultaneously improve both system performance and privacy protection," a combination particularly valuable in advertising technology applications [9]. These coordination strategies offer promising avenues for addressing the fragmentation challenges that have historically constrained advertising ecosystem efficiency.

Looking forward, the continued advancement of edge computing capabilities promises to enhance RTB performance and capabilities further. The analysis of emerging hardware trends provides compelling evidence for this trajectory, with the benchmark testing of five successive edge computing hardware generations demonstrating computation-per-watt improvements averaging 37.8% annually. This rapid performance evolution enables increasingly sophisticated processing at the network edge, with the testing showing that advanced machine learning model inference—previously restricted to cloud environments—now achieves sub-15ms execution times on current-generation edge hardware. The researchers project that this trend will continue, with edge devices reaching performance parity with 2023-era cloud resources for many workload classes by 2026. The comprehensive cost modeling indicates that this performance evolution will drive "ongoing migration of time-sensitive processing toward the network edge," with optimal architecture designs shifting approximately 7-9% of workloads from cloud to edge annually based on performance improvements. "The evolution of computing infrastructure," the conclusion, "points towards increasingly distributed architectures that process data as close as possible to its origin, fundamentally transforming application designs across numerous domains, including advertising technology" [10]. This hardware evolution will enable increasingly sophisticated RTB implementations, making bid decisioning more accurate, efficient, and responsive as computational capabilities advance at the network edge.

Table 2 Technological Evolution in Edge Computing for Real-Time Bidding Systems. [9, 10]

Future Technology Metric	Current Value	Projected Future Value	Improvement (%)
Communication Overhead Reduction (%)	76	91.5	20.4
Model Accuracy Maintenance (%)	97.8	99.5	1.7
Classification Accuracy Improvement (%)	9.2	17.8	93.5
Performance Maintenance During Network Disruption (%)	93.4	98.2	5.1
Latency Reduction in Hybrid Systems (%)	47.3	68.5	44.8
Operational Cost Reduction (%)	32.8	47.6	45.1
Edge Processing Response Time (ms)	50	15	70
Annual Computation-per-Watt Improvement (%)	37.8	41.2	9
Workload Migration to Edge (% annually)	8	14.5	81.3
Duplicate Data Transmission Reduction (%)	64.7	82.3	27.2

8. Conclusion

Integrating edge-embedded systems into real-time bidding workflows significantly advances AdTech infrastructure. This approach substantially reduces latency, enhances decision-making speed, and improves resource utilization by decentralizing critical computations to local edge nodes. The architectural shift addresses current technical limitations in RTB systems and band aligns with evolving privacy requirements and economic considerations in the digital advertising ecosystem. The performance improvements—particularly the reduction in average bid response time and decreased infrastructure costs—highlight the compelling business case for adopting edge-embedded approaches. As the digital advertising landscape evolves toward greater complexity and higher performance requirements, edgeaugmented RTB architectures will likely become an industry standard for high-performance programmatic advertising platforms.

References

[1] GeeksforGeeks, "Distributed Application Paradigms," 2021. https://www.geeksforgeeks.org/distributed-application-paradigms/

[2] Priyanka Varatharajan, "How to Optimize Header Bidding Script to Reduce Page Latency & Improve Performance," Headerbidding, 2024, https://headerbidding.co/optimize-header-bidding-latency-improve-performance/

[3] Jonathan J.J.M. Seddon, Wendy L. Currie, "A model for unpacking big data analytics in high-frequency trading," Journal of Business Research, 2017. https://www.sciencedirect.com/science/article/abs/pii/S0148296316304908

[4] Chris Yu, "The benefits of machine learning in advertising," Illumin Insights Blog, 2024. https://illumin.com/insights/blog/machine-learning-advertising/

[5] Ayoub el Berkaoui et al., "Distribution network topology planning and optimization: A brief review," Energy and AI, 2023. https://www.researchgate.net/publication/376683590_Distribution_network_topology_planning_and_optimization_A_brief_review

[6] Erdong Deng et al., "Pri-RTB: Privacy-preserving real-time bidding for securing mobile advertisement in ubiquitous computing," Information Sciences, 2019. https://www.sciencedirect.com/science/article/abs/pii/S0020025519306413

[7] Natalia Figas, "The Benefits of Privacy-Enhancing Technologies (PETs) In AdTech," Clearcode Blog, 2024. https://clearcode.cc/blog/benefits-privacy-enhancing-technologies-adtech/

[8] Fox Mandal, "A guide to conquer the cross-border compliance challenges," Lexology, 2023. https://www.lexology.com/library/detail.aspx?g=11af397a-0f29-41b9-8539-de6b1d9f24b1

[9] Rasool Seyghaly et al., "A Comprehensive Architecture for Federated Learning-Based Smart Advertising," Sensors, 2024. https://www.mdpi.com/1424-8220/24/12/3765

[10] Amelia Mia, Lucas Chloe, "Optimizing Hybrid and Multi-Cloud Architectures for Real-Time Analytics and Data-Driven Decisions," ResearchGate, 2024. https://www.researchgate.net/publication/384661110_Optimizing_Hybrid_and_Multi-Cloud_Architectures_for_Real-Time_Analytics_and_Data-Driven_Decisions