International Science and Technology Journal
المجلة الدولية للعلوم والتقنية
Volume 39 العدد
Part 1 المجلد
المجلة الدولية للعلوم والتقنية
International Science and Technology Journal
ISTJ

Network Performance Evaluation Using Benchmarking, Monitoring, and Simulation
http://www.doi.org/10.62341/istj-vol39-1-ah18

Received	**2026/07/01**	**تم استلام الورقة العلمية في**
Accepted	**2026/07/30**	**تم قبول الورقة العلمية في**
Published	**2026/08/01**	**تم نشر الورقة العلمية في**

Network Performance Evaluation Using Benchmarking, Monitoring, and Simulation

Hasan Omar Ali
High Institute for Engineering Technologies - Al-Majori,
Benghazi, Libya
ORCID: 0009-0007-2682-6221
hassan.algbaly@gmail.com

Abstract

Network performance evaluation plays a critical role in ensuring the efficiency, reliability, and quality of modern communication systems. This paper presents a systematic review of three major network performance evaluation techniques: benchmarking, monitoring, and simulation. The study analyzes their principles, strengths, limitations, and practical applications. A comparative analysis is provided to highlight their differences in terms of accuracy, cost, flexibility, and real-time applicability. The qualitative ratings used in this comparison are linked directly to specific supporting studies from the reviewed literature, making the assessment traceable rather than purely subjective. A decision-oriented framework is further proposed to map common network scenarios to the most suitable evaluation technique. Furthermore, the paper discusses emerging trends such as Artificial Intelligence, Software Defined Networking, and Digital Twin technologies, supported by a dedicated comparison with prior surveys that clarifies the specific gap this review addresses. The findings show that selecting the appropriate evaluation method depends on the network environment, research objectives, and available resources.

Keywords: Network Performance Evaluation, Benchmarking, Monitoring, Simulation, QoS, Digital Twin.

International Science and Technology Journal
المجلة الدولية للعلوم والتقنية
Volume 39 العدد
Part 1 المجلد
المجلة الدولية للعلوم والتقنية
International Science and Technology Journal
ISTJ

Network Performance Evaluation Using Benchmarking, Monitoring, and Simulation
http://www.doi.org/10.62341/istj-vol39-1-ah18

تقييم أداء الشبكات باستخدام القياس المرجعي والمراقبة والمحاكاة

حسن عمر علي
المعهد العالي للتقنيات الهندسية – الماجوري، بنغازي، ليبيا
ORCID: 0009-0007-2682-6221
hassan.algbaly@gmail.com

الملخص

تلعب عملية تقييم أداء الشبكات دورا مهما في ضمان كفاءة وموثوقية وجودة أنظمة الاتصالات الحديثة .تقدم هذه الورقة مراجعة منهجية لثلاث تقنيات رئيسية في تقييم أداء الشبكات، وهي القياس المرجعي، والمراقبة، والمحاكاة .تناقش الدراسة مبادئ هذه التقنيات ومزاياها وحدودها وتطبيقاتها العملية، مع تقديم مقارنة تحليلية من حيث الدقة، والتكلفة، والمرونة، وإمكانية التحليل في الزمن الحقيقي .كما تقترح الورقة إطارا إرشاديا يساعد الباحثين والمهندسين على اختيار التقنية الأنسب وفقا لطبيعة الشبكة وأهداف البحث والموارد المتاحة .وتتناول الدراسة كذلك الاتجاهات الحديثة مثل الذكاء الاصطناعي، والشبكات المعرفة برمجيا، وتقنيات التوأم الرقمي .وتبين النتائج أن الاختيار الصحيح لطريقة التقييم يعتمد على بيئة الشبكة والغرض من التحليل ومستوى الدقة المطلوب.

الكلمات المفتاحية: تقييم أداء الشبكات، القياس المرجعي، المراقبة، المحاكاة، جودة الخدمة، التوأم الرقمي.

1. INTRODUCTION

Computer networks have become a fundamental part of modern digital infrastructures, supporting a wide range of services such as cloud computing, Internet of Things (IoT), multimedia applications, e-commerce, and enterprise systems (Kurose & Ross, 2021; Zhang et al., 2010). With the rapid growth of communication technologies and the increasing demand for high-speed, reliable, and scalable connectivity, maintaining efficient network performance has become a critical requirement for organizations, service providers, and researchers.

Network Performance Evaluation Using Benchmarking, Monitoring, and Simulation

http://www.doi.org/10.62341/istj-vol39-1-ah18

Network performance evaluation plays a vital role in assessing the effectiveness of network architectures, communication protocols, and management strategies. It helps researchers and network administrators identify bottlenecks, optimize resource utilization, improve Quality of Service (QoS), and enhance the overall user experience (Hofstede et al., 2014; Jain, 1991; Abbasi et al., 2021). Effective performance evaluation is essential to ensure network reliability, stability, and efficiency, especially in large-scale and dynamic communication environments.

Over the years, several techniques have been developed to evaluate network performance. Among the most widely adopted approaches are benchmarking, monitoring, and simulation (vom Lehn et al., 2008; Hofstede et al., 2014; Jain, 1991). Benchmarking focuses on measuring network performance under predefined and controlled conditions. Monitoring observes real-time network behavior in operational environments, allowing administrators to detect anomalies and performance degradation. Simulation provides a flexible and cost-effective environment for modeling network behavior and evaluating network scenarios before real deployment.

Although these techniques share the same general objective, each approach offers different advantages and limitations in terms of accuracy, cost, flexibility, and real-time applicability. Benchmarking provides highly accurate and repeatable measurements, monitoring delivers continuous visibility into live networks, and simulation enables researchers to test complex scenarios with minimal cost. Therefore, understanding the strengths and weaknesses of these techniques is essential for selecting the most appropriate performance evaluation method.

With the emergence of Software Defined Networking (SDN), Edge Computing, Artificial Intelligence (AI), and Digital Twin technologies, network environments have become increasingly complex and dynamic (Feamster et al., 2014; Mao et al., 2017). These advancements have created new challenges in network performance evaluation, requiring more intelligent, adaptive, and scalable evaluation techniques.

This paper presents a systematic review of network performance evaluation techniques, focusing on benchmarking,

http://www.doi.org/10.62341/istj-vol39-1-ah18

monitoring, and simulation. The study analyzes their principles, methodologies, strengths, limitations, and practical applications. Furthermore, a comparative analysis is provided to support researchers and practitioners in selecting suitable evaluation techniques for different networking environments and research objectives.

While several prior studies have addressed network performance evaluation, most focus on a single technique in isolation. Hofstede et al. (2014) and Lee et al. (2014) specifically surveyed network monitoring approaches, without addressing benchmarking or simulation. Jain (1991) provided a foundational treatment of performance analysis methods, but this work predates modern networking paradigms such as Software Defined Networking, cloud/edge computing, and AI-driven network management. Similarly, vom Lehn et al. (2008) focused exclusively on comparing simulation tools. To the best of our knowledge, no existing survey jointly examines benchmarking, monitoring, and simulation within a unified comparative framework while explicitly linking them to emerging paradigms such as SDN, AI, and Digital Twin technologies, this paper addresses this gap.

The main contributions of this paper are as follows:

- Providing a comprehensive overview of network performance evaluation concepts and metrics.
- Reviewing the fundamental principles of benchmarking, monitoring, and simulation techniques.
- Comparing the strengths and limitations of each evaluation approach.
- Identifying emerging trends and future research directions in network performance evaluation.
- Proposing a decision-oriented framework that maps evaluation techniques to specific network scenarios, translating the comparative analysis into practical, actionable guidance.

2. RESEARCH METHODOLOGY

This study follows a Systematic Literature Review (SLR) methodology to analyze and compare the major network performance evaluation techniques, including benchmarking, monitoring, and simulation.

http://www.doi.org/10.62341/istj-vol39-1-ah18

2.1. SEARCH STRATEGY

The literature search was conducted using several well-known academic databases, including IEEE Xplore, ACM Digital Library, SpringerLink, Elsevier ScienceDirect, and Google Scholar. The search focused on both foundational and recent studies published between 1991 and 2026 to ensure comprehensive coverage of classical and emerging network performance evaluation techniques.

2.2. INCLUSION CRITERIA

To ensure the quality and relevance of the selected studies, the following inclusion criteria were applied:

- Articles published between 1991 and 2026.
- Peer-reviewed journal and conference papers.
- Articles written in English.
- Studies directly related to network performance evaluation.
- Research focusing on benchmarking, monitoring, simulation, or performance metrics.

2.3. EXCLUSION CRITERIA

The following exclusion criteria were used to filter irrelevant studies:

- Articles published before 1991.
- Non-peer-reviewed articles, reports, or blogs.
- Duplicate studies across different databases.
- Studies not directly related to network performance evaluation.
- Articles focusing on unrelated fields outside computer networking.

2.4. STUDY SELECTION PROCESS

Table 1 summarizes the study selection process followed in the systematic search component of this review.

Table 1. Study selection process summary

Stage	No. of Studies
Records identified from databases	65
Duplicate records removed	10
Records screened by title/abstract	55
Records excluded as not relevant	28
Full-text studies assessed	27
Studies excluded after full-text review	7
Final studies included in the review	20

Network Performance Evaluation Using Benchmarking, Monitoring, and Simulation

http://www.doi.org/10.62341/istj-vol39-1-ah18

Figure 1 presents the same selection process as a PRISMA-style flow diagram, illustrating how the initial set of records identified through the systematic database search was progressively narrowed down to the 20 studies included through that process. The remaining 26 sources discussed in this review are not part of this PRISMA funnel: 4 recent works on AI-driven network digital twins (Li et al., 2025; Villa et al., 2024; Dhamala et al., 2024; Banisadr & Hesselbach, 2024) were added through targeted manual search, while 22 well-established surveys and foundational studies (Al-Fuqaha et al., 2015; Nguyen & Armitage, 2008; Pacheco et al., 2019; Abbasi et al., 2021; Zehra et al., 2023; Xia et al., 2015; Nunes et al., 2014; Hu et al., 2014; Ahmad et al., 2015; Scott-Hayward et al., 2016; Rafique et al., 2020; Mao et al., 2017; Iosup et al., 2011; Heidemann et al., 2001; Mihai et al., 2022; Wu et al., 2021; Khan et al., 2022; Bianchi, 2000; Al-Saadi et al., 2019; Adams, 2013; Briscoe et al., 2016; Guck et al., 2018) were incorporated to broaden topical coverage, as detailed in Section B.4.

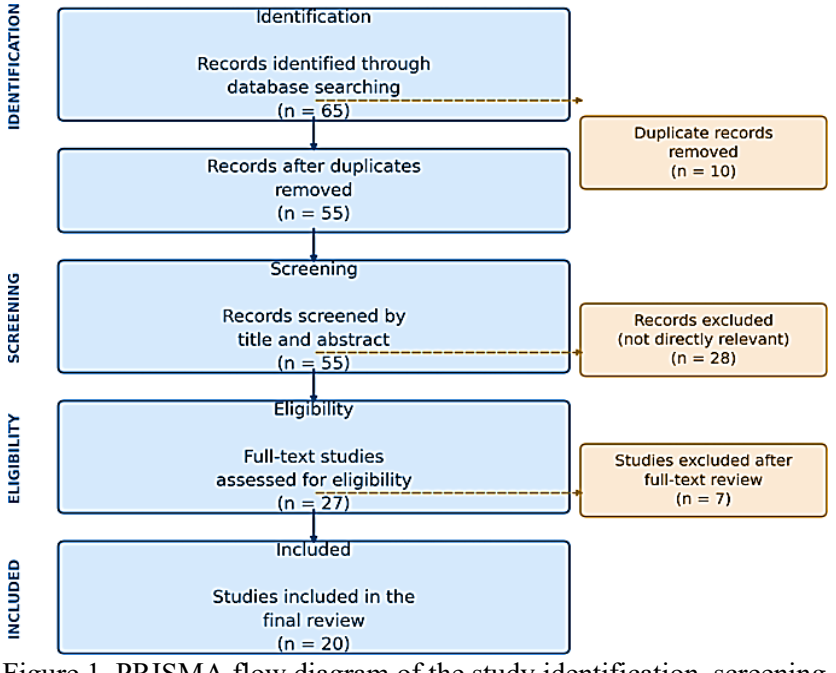

Figure 1. PRISMA flow diagram of the study identification, screening, eligibility, and inclusion process.

http://www.doi.org/10.62341/istj-vol39-1-ah18

This methodology provides a systematic and reliable foundation for analyzing current network performance evaluation techniques and identifying future research directions.

3 .NETWORK PERFORMANCE FUNDAMENTALS

Network performance refers to the effectiveness and efficiency of data communication within a network (Kurose & Ross, 2021; Xie et al., 2019).

It is a critical factor in determining the quality of network services and the ability of the network to support different applications and users. Network performance evaluation helps researchers and administrators measure how well a network operates under different conditions and workloads. By analyzing performance, organizations can identify bottlenecks, optimize resource utilization, and improve the overall Quality of Service (QoS).

Several Key Performance Indicators (KPIs) are commonly used to measure network performance, including bandwidth, throughput, latency, jitter, packet loss, and reliability (Hofstede et al., 2014; Xie et al., 2019; Abbasi et al., 2021).

Bandwidth represents the maximum amount of data that can be transmitted over a network connection within a given period of time. It is usually measured in bits per second (bps) and indicates the capacity of the communication channel. Throughput, on the other hand, refers to the actual amount of data successfully transmitted from source to destination during a specific time. While bandwidth shows the potential capacity, throughput reflects the real performance achieved, and is shaped in practice by lower-layer effects such as wireless medium contention, for which Bianchi's analytical model of the IEEE 802.11 distributed coordination function remains a foundational reference (Bianchi, 2000), as well as by transport-layer congestion control behavior (Al-Saadi et al., 2019).

Latency is the time required for data to travel from the sender to the receiver. Lower latency is essential for real-time applications such as video conferencing, online gaming, and voice communication, and a broad range of techniques for reducing it across different sources of delay has itself been the subject of

http://www.doi.org/10.62341/istj-vol39-1-ah18

dedicated survey work (Briscoe et al., 2016). Jitter refers to the variation in packet delay over time. High jitter can negatively affect the quality of real-time services by causing inconsistent data delivery. Packet loss occurs when data packets fail to reach their destination, and is also influenced by router-level buffer management policies, which active queue management schemes are specifically designed to address (Adams, 2013). It can reduce network performance and lead to communication errors, retransmissions, and service interruptions. Reliability measures the consistency and stability of the network over time, indicating its ability to maintain communication without failures. Table 2 summarizes the main network performance metrics and their significance (Jain, 1991).

TABLE 2. Key network performance metrics

KPI	Definition	Importance
Bandwidth	Max. data transfer capacity	Determines network capacity
Throughput	Actual data successfully sent	Reflects real performance
Latency	Time delay in transmission	Key for real-time use
Jitter	Variation in packet delay	Affects service stability
Packet Loss	Lost packets during transmission	Impacts comm. quality
Reliability	Network stability/consistency	Ensures continuous operation

Understanding these performance fundamentals is essential before applying benchmarking, monitoring, and simulation techniques. These metrics serve as the foundation for evaluating, comparing, and improving modern computer networks.

4 . BENCHMARKING TECHNIQUES

Benchmarking is one of the commonly used techniques for evaluating network performance (Jain, 1991; Li et al., 2010). It refers to the process of measuring the performance of a network, device, protocol, or system under predefined and controlled conditions. The main purpose of benchmarking is to provide objective performance results that can be used for comparison, validation, and improvement.

http://www.doi.org/10.62341/istj-vol39-1-ah18

In computer networks, benchmarking can be used to evaluate several performance aspects, such as throughput, latency, packet loss, jitter, bandwidth utilization, and response time. These metrics help researchers and network administrators understand how well a network performs under different traffic loads and operational scenarios. One of the main advantages of benchmarking is that it provides measurable and comparable results. For example, two routing protocols or two network devices can be tested under the same conditions to determine which one provides better performance. This makes benchmarking useful in both academic research and practical network design.

Several practical tools are commonly used to perform network benchmarking. iPerf, and its successor iPerf3, is one of the most widely used open-source tools for measuring achievable bandwidth and throughput between two endpoints over TCP and UDP. Netperf provides similar throughput measurement capabilities, with additional support for evaluating latency and request-response performance under controlled load. For benchmarking access-network performance from the end-user side, tools such as Ookla Speedtest are widely used to measure download and upload throughput and latency under real, rather than fully controlled, conditions. At a larger scale, structured benchmarking studies such as CloudCmp (Li et al., 2010) and the many-tasks scientific computing benchmarks of Iosup et al. (2011) apply these same principles systematically, running standardized benchmark suites to compare the performance of different public cloud providers across multiple metrics. These tools illustrate how the throughput, latency, and jitter metrics discussed earlier are operationalized in practice.

However, benchmarking also has some limitations. It often requires a controlled testing environment, specific tools, and well-defined test scenarios. In addition, benchmarking results may not always fully represent real-world network behavior because actual networks are affected by many dynamic factors, such as user activity, traffic variation, hardware limitations, and unexpected failures. Despite these limitations, benchmarking remains an important method in network performance evaluation (Jain, 1991; Li et al., 2010). When properly designed, it can provide valuable

http://www.doi.org/10.62341/istj-vol39-1-ah18

insights into the strengths and weaknesses of network components and help in making informed decisions regarding network optimization, protocol selection, and infrastructure planning.

5 .NETWORK MONITORING TECHNIQUES

Network monitoring is a key technique used to evaluate operational computer networks in real-time (Hofstede et al., 2014; Lee et al., 2014; Claffy et al., 1995; Abbasi et al., 2021). Unlike benchmarking, which is usually conducted under controlled testing conditions, network monitoring focuses on observing real-time network behavior in live environments. It enables network administrators to continuously collect performance data, detect faults, analyze traffic patterns, and respond to service degradation before it affects users (Barford & Plonka, 2001; Dovrolis et al., 2001).

Network monitoring can be classified into two main approaches: active monitoring and passive monitoring. Active monitoring involves generating test traffic or probe packets to measure specific performance metrics such as latency, packet loss, response time, and availability. This approach is useful for testing network paths and service responsiveness. However, it may introduce additional traffic overhead. Passive monitoring, on the other hand, observes actual network traffic without injecting additional packets. It is useful for analyzing real user traffic, detecting anomalies, and understanding application behavior, but it may require advanced tools and careful data analysis. Machine learning has increasingly been applied to passive monitoring data to automate traffic classification, building on early foundational work (Nguyen & Armitage, 2008) through more recent systematic surveys of machine-learning-based classification approaches (Pacheco et al., 2019) and deep-learning-based traffic monitoring and analysis frameworks (Abbasi et al., 2021).

Fault detection is one of the most important functions of network monitoring. By continuously observing network devices, links, and services, monitoring systems can identify failures, congestion, abnormal traffic, and performance degradation (Zehra et al., 2023). This helps administrators take corrective actions quickly and maintain service continuity. Traffic analysis is another important

http://www.doi.org/10.62341/istj-vol39-1-ah18

aspect, as it allows organizations to understand bandwidth usage, detect bottlenecks, identify unusual traffic patterns, and improve network planning.

Several tools are commonly used for network monitoring and traffic analysis. Wireshark is widely used for packet-level traffic inspection and protocol analysis. Nagios provides infrastructure monitoring, alerting, and service availability checking. Zabbix is another powerful monitoring platform that supports network device monitoring, performance metrics collection, visualization, and alerting. These tools help administrators gain practical insight into network behavior and support effective performance management (Lee et al., 2014).

Despite its advantages, network monitoring also has some limitations. It may require continuous resource usage, storage for collected data, and proper configuration of monitoring tools. In addition, analyzing large volumes of monitoring data can be complex, especially in large-scale networks. Nevertheless, network monitoring remains essential for ensuring reliability, detecting faults, maintaining Quality of Service (QoS), and improving the overall performance of modern computer networks.

6 . NETWORK SIMULATION TECHNIQUES

Network simulation is one of the most widely used approaches for evaluating network performance in research and development (vom Lehn et al., 2008; Varga & Hornig, 2008; Henderson et al., 2008).

It provides a virtual environment where network behavior, protocols, and traffic patterns can be modeled and analyzed without the need for physical infrastructure.

Simulation techniques allow researchers to study different network scenarios, test protocol performance, and analyze the impact of various parameters such as bandwidth, delay, packet loss, and node mobility. This approach is particularly useful in large-scale or complex network environments where real-world implementation may be costly, time-consuming, or impractical.

Several network simulation tools such as NS-3, OMNeT++, and OPNET are commonly used in performance evaluation (Varga & Hornig, 2008; Henderson et al., 2008). These tools enable

http://www.doi.org/10.62341/istj-vol39-1-ah18

researchers to create customizable network models and simulate real communication behaviors under different conditions, although, as discussed early on in the simulation literature, the credibility of such results depends heavily on careful validation of the underlying simulation models (Heidemann et al., 2001).

More recently, simulation research has begun to move beyond traditional discrete-event simulators toward AI-assisted and digital-twin-based emulation platforms. For example, Villa et al. (2024) demonstrated how the Colosseum wireless testbed can function as a digital twin that bridges real-world experimentation with large-scale network emulation, while Dhamala et al. (2024) applied graph neural networks to improve the accuracy of simulation-based performance prediction in the RouteNet model. These developments indicate a gradual convergence between classical network simulation and the data-driven techniques discussed further in Section H.

One of the main advantages of network simulation is its flexibility. Researchers can modify network topologies, traffic loads, and protocol configurations easily. In addition, simulation reduces the cost of experimentation and allows repeated testing under identical conditions.

However, simulation also has limitations. The results depend heavily on the accuracy of the simulation model and assumptions. In some cases, simulated environments may not fully reflect real-world network behavior due to unpredictable factors such as hardware limitations, user behavior, and environmental conditions (vom Lehn et al., 2008).

Despite these challenges, network simulation remains a powerful and essential technique in network performance evaluation. It is widely used for protocol design, performance comparison, scalability testing, and future network planning.

7 .COMPARATIVE ANALYSIS

A comparative analysis of network performance evaluation techniques is essential to understand the strengths, limitations, and applicability of each method (Hofstede et al., 2014; Jain, 1991; Henderson et al., 2008).

http://www.doi.org/10.62341/istj-vol39-1-ah18

Benchmarking, Monitoring, and Simulation are widely used approaches, but each serves different purposes depending on the network environment and research objectives. Benchmarking provides highly accurate and measurable results under controlled conditions. It is useful for comparing devices, protocols, and network configurations. However, it requires a dedicated testing environment and may not fully represent real-world network dynamics.

Monitoring, on the other hand, focuses on observing live network behavior in real-time. It allows network administrators to detect faults, measure performance continuously, and respond to anomalies quickly. Despite its practical advantages, monitoring may introduce overhead and can be limited by the tools used for data collection.

Simulation offers a flexible and cost-effective way to model and evaluate networks before deployment. It allows researchers to test multiple scenarios without physical hardware. This makes simulation ideal for academic studies and protocol development. However, its accuracy depends on the quality of the simulation model and assumptions.

The qualitative ratings shown in Table 3 were derived from the reviewed literature and based on five evaluation criteria: accuracy, cost, flexibility, real-time analysis, and implementation complexity. The terms High, Medium, and Low are used as relative indicators rather than absolute measurements. To make this assessment traceable rather than purely subjective, the last row of Table 3 lists the specific studies from the reviewed literature that support the rating assigned to each technique. Table 3 summarizes the key differences between these three techniques.

http://www.doi.org/10.62341/istj-vol39-1-ah18

Table 3. Comparative Analysis of Network Performance Evaluation Techniques

Criterion	Benchmarking	Monitoring	Simulation
Accuracy	High	High	Med.
Cost	High	Med.	Low
Flexibility	Med.	Low	High
Real-Time Applicability	No	Yes	No
Implementation Complexity	Med.	High	Med.
Supporting References	(Li et al., 2010; Claffy et al., 1995; Dovrolis et al., 2001)	(Hofstede et al., 2014; Lee et al., 2014; Barford & Plonka, 2001; Abbasi et al., 2021)	(vom Lehn et al., 2008; Varga & Hornig, 2008; Henderson et al., 2008; Villa et al., 2024)

Ratings (High/Med./Low) are relative qualitative indicators derived from the cited literature, not absolute measurements.

To translate this comparison into practical guidance, Table 4 maps common network scenarios to the most suitable evaluation technique.

Table 4. Recommended evaluation technique by use case

Use Case	Technique	Justification
Controlled lab testing / protocol validation	Benchmarking	Highest accuracy, repeatable results
Live production network management	Monitoring	Real-time visibility, fault detection
Pre-deployment topology / capacity planning	Simulation	Low cost, scalable scenario testing
Large-scale dynamic SDN/IoT environments	Hybrid (Monitoring + Simulation) / Digital Twin	Combines live data with predictive modeling

This decision framework — in which the Digital Twin approach, discussed in detail in Section H, is treated as a hybrid option — allows researchers and network engineers to move directly from the qualitative comparison in Table 3 to a concrete choice of method based on their specific operational context, rather than relying on general intuition.

http://www.doi.org/10.62341/istj-vol39-1-ah18

Based on this comparison, benchmarking is more suitable for controlled experimental validation, monitoring is ideal for operational network management, and simulation is the preferred choice for research, testing, and future network planning. Therefore, selecting the appropriate technique depends on the specific goals, available resources, and network conditions.

8 . FUTURE TRENDS

The rapid evolution of computer networks and communication technologies has introduced new challenges in network performance evaluation. Traditional techniques such as benchmarking, monitoring, and simulation remain important; however, emerging technologies are transforming the way network performance is analyzed and optimized (Boutaba et al., 2018; Xie et al., 2019).

One of the most significant future trends is the integration of Artificial Intelligence (AI) and Machine Learning (ML) into network performance evaluation (Mao et al., 2018; Tang et al., 2016). AI-based systems can analyze large volumes of network data, detect anomalies, predict failures, and optimize traffic management automatically. This approach improves the accuracy and efficiency of network monitoring and decision-making.

Another important trend is the adoption of Software Defined Networking (SDN) (Feamster et al., 2014; McKeown et al., 2008; Xia et al., 2015; Nunes et al., 2014; Hu et al., 2014). SDN provides centralized control over network operations, making it easier to monitor performance, manage resources, and implement dynamic optimization strategies. Performance evaluation in SDN environments has become an active research area due to its flexibility and programmability, including dedicated comparative evaluations of QoS-aware routing algorithms for centrally controlled SDN networks (Guck et al., 2018), and the security implications of this centralized control have themselves been the subject of dedicated surveys (Ahmad et al., 2015; Scott-Hayward et al., 2016).

Cloud computing and edge computing also represent important areas for future performance evaluation (Lee et al., 2014; Barford & Plonka, 2001; Rafique et al., 2020; Mao et al., 2017). As network

http://www.doi.org/10.62341/istj-vol39-1-ah18

services become more distributed, evaluating latency, resource allocation, and service reliability across cloud and edge infrastructures becomes increasingly critical. In addition, Digital Twin technology is emerging as a promising approach for network performance evaluation (Hui et al., 2023; Mihai et al., 2022; Wu et al., 2021; Khan et al., 2022). A digital twin creates a virtual representation of a physical network, allowing real-time analysis, simulation, and predictive performance modeling. This technology has the potential to combine the advantages of monitoring and simulation in a single framework.

Recent work has begun to operationalize this concept. Li et al. (2025) reviewed how generative AI techniques are being integrated into network digital twins to improve their architecture and predictive capabilities, while Banisadr and Hesselbach (2024) proposed a novel architecture for deploying virtual network twins in practice. Together with the emulation-based digital twin approach of Villa et al. (2024) discussed in Section F, these studies suggest that digital twin research is rapidly transitioning from conceptual frameworks toward practical, deployable systems.

Finally, the growth of Internet of Things (IoT) networks presents new challenges related to scalability, energy efficiency, and security (Atzori et al., 2010; Al-Fuqaha et al., 2015). Future performance evaluation methods must adapt to the large-scale and dynamic nature of IoT environments.

Overall, future network performance evaluation will increasingly rely on intelligent, automated, and adaptive systems to handle the complexity of next-generation networks (Mao et al., 2018; Tang et al., 2016).

9. LIMITATIONS AND COMPARISON WITH PRIOR REVIEWS

While this review aims to provide a structured and comprehensive synthesis of network performance evaluation techniques, several limitations should be acknowledged. First, the systematic search was limited to five databases; relevant work indexed exclusively in other venues may not have been captured. Second, the review was restricted to English-language publications. Third, although the qualitative ratings in Table 3 are now linked to specific supporting

http://www.doi.org/10.62341/istj-vol39-1-ah18

references, they remain relative, literature-grounded assessments rather than outputs of a standardized quantitative meta-analysis. Fourth, this review focuses specifically on benchmarking, monitoring, and simulation; adjacent evaluation paradigms were discussed only where they intersect with these three techniques. Finally, the four additional recent works on AI-driven and Digital Twin-based evaluation (Li et al., 2025; Villa et al., 2024; Dhamala et al., 2024; Banisadr & Hesselbach, 2024) were identified through targeted manual search rather than the formal systematic protocol, and were included on the basis of direct topical relevance rather than a full screening procedure. In addition to the research gap outlined in the Introduction, Table 5 explicitly compares the scope of this review with the most closely related prior surveys identified during the literature search, clarifying the specific gap this paper addresses.

Table 5. Comparison with prior surveys

Prior Survey	Scope	Key Limitation
vom Lehn et al. (2008)	Comparison of simulation tools only	No benchmarking/monitoring, no AI/Digital Twin
Hofstede et al. (2014)	Tutorial on flow monitoring (NetFlow/IPFIX)	Focuses on flow-based monitoring only; no benchmarking/simulation
Jain (1991)	Foundational performance analysis methodology	Predates SDN, cloud/edge, AI-driven management
Lee et al. (2014)	Monitoring: present and future	Monitoring only; no unified framework

As shown in Table 5, prior surveys have generally addressed benchmarking, monitoring, and simulation in isolation, or have not incorporated recent AI- and Digital Twin-driven developments. The unified comparative framework (Table 3), decision-oriented guidance (Table 4), and traceable rating methodology developed in this review are intended to directly address these gaps.

10. CONCLUSION

Network performance evaluation is essential for ensuring the reliability, efficiency, and quality of modern communication systems. This paper presented a systematic review of three major

http://www.doi.org/10.62341/istj-vol39-1-ah18

evaluation techniques — benchmarking, monitoring, and simulation (Kurose & Ross, 2021; Hofstede et al., 2014; Jain, 1991) — analyzing their principles, strengths, and limitations. The comparative analysis showed that benchmarking offers the highest accuracy under controlled conditions, monitoring provides real-time operational visibility, and simulation offers flexible, low-cost scenario testing. Selecting the appropriate technique depends on the research objectives, available resources, and network conditions, while emerging technologies such as Artificial Intelligence, Software Defined Networking, and Digital Twin systems are expected to reshape future evaluation approaches. Future research should focus on developing intelligent, adaptive frameworks capable of handling the growing complexity of next-generation networks.

ACKNOWLEDGMENT

The author acknowledges the use of ChatGPT (OpenAI) as an AI-assisted language support tool during the preparation of this manuscript, particularly for language refinement, structural suggestions, and writing improvement. All technical content, analysis, and final decisions were reviewed and validated by the author.

REFERENCES

Abbasi, M., Shahraki, A., & Taherkordi, A. (2021). “Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey,” Computer Communications, vol. 170, pp. 19–41.

Adams, R. (2013). “Active Queue Management: A Survey,” IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1425–1476.

Ahmad, I., Namal, S., Ylianttila, M., & Gurtov, A. (2015). “Security in Software Defined Networks: A Survey,” IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2317–2346.

Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). “Internet of Things: A Survey on

http://www.doi.org/10.62341/istj-vol39-1-ah18

Enabling Technologies, Protocols, and Applications," IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2347–2376.

Al-Saadi, R., Armitage, G., But, J., & Branch, P. (2019). "A Survey of Delay-Based and Hybrid TCP Congestion Control Algorithms," IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3609–3638.

Atzori, L., Iera, A., & Morabito, G. (2010). "The Internet of Things: A Survey," Computer Networks, vol. 54, no. 15, pp. 2787–2805.

Banisadr, A. H., & Hesselbach, X. (2024). "A Novel Architecture for Virtual Network Twin Deployment," Electronics, vol. 13, no. 24, p. 5045.

Barford, P., & Plonka, D. (2001). "Characteristics of Network Traffic Flow Anomalies," in Proc. 1st ACM SIGCOMM Internet Measurement Workshop (IMW), pp. 69–73.

Bianchi, G. (2000). "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," IEEE Journal on Selected Areas in Communications, vol. 18, no. 3, pp. 535–547.

Boutaba, R., Salahuddin, M. A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., & Caicedo, O. M. (2018). "A Comprehensive Survey on Machine Learning for Networking," Journal of Internet Services and Applications, vol. 9, no. 1, pp. 1–99.

Briscoe, B., Brunstrom, A., Petlund, A., et al. (2016). "Reducing Internet Latency: A Survey of Techniques and Their Merits," IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 2149–2196.

Claffy, K., Braun, H.-W., & Polyzos, G. (1995). "A Parameterizable Methodology for Internet Traffic Flow Profiling," IEEE Journal on Selected Areas in Communications, vol. 13, no. 8, pp. 1481–1494.

http://www.doi.org/10.62341/istj-vol39-1-ah18

Dhamala, B. K., Dawadi, B. R., Manzoni, P., & Acharya, B. K. (2024). "Performance Evaluation of Graph Neural Network-Based RouteNet Model," Future Internet, vol. 16, no. 4, p. 116.

Dovrolis, C., Ramanathan, P., & Moore, D. (2001). "What Do Packet Dispersion Techniques Measure?" Proc. IEEE INFOCOM, pp. 905–914.

Feamster, N., Rexford, J., & Zegura, E. (2014). "The Road to SDN: An Intellectual History of Programmable Networks," ACM SIGCOMM Computer Communication Review, vol. 44, no. 2, pp. 87–98.

Guck, J. W., Van Bemten, A., Reisslein, M., & Kellerer, W. (2018). "Unicast QoS Routing Algorithms for SDN," IEEE Communications Surveys & Tutorials, vol. 20, no. 1, pp. 388–415.

Heidemann, J., Mills, K., & Kumar, S. (2001). "Expanding Confidence in Network Simulations," IEEE Network, vol. 15, no. 5, pp. 58–63.

Henderson, T. R., Lacage, M., Riley, G. F., Dowell, C., & Kopena, J. (2008). "Network Simulations with the NS-3 Simulator," SIGCOMM Demonstration.

Hofstede, R., Čeleda, P., Trammell, B., Drago, I., Sadre, R., Sperotto, A., & Pras, A. (2014). "Flow Monitoring Explained: From Packet Capture to Data Analysis With NetFlow and IPFIX," IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 2037–2064.

Hu, F., Hao, Q., & Bao, K. (2014). "A Survey on Software-Defined Network and OpenFlow," IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 2181–2206.

Hui, L., Wang, M., Zhang, L., Lu, L., & Cui, Y. (2023). "Digital Twin for Networking: A Data-Driven Performance Modeling Perspective," IEEE Network, vol. 37, no. 3, pp. 202–209.

http://www.doi.org/10.62341/istj-vol39-1-ah18

Iosup, A., Ostermann, S., Yigitbasi, N., Prodan, R., Fahringer, T., & Epema, D. (2011). "Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing," IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 6, pp. 931–945.

Jain, R. (1991). The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling, New York, NY, USA: Wiley.

Khan, L. U., Han, Z., Saad, W., Hossain, E., Guizani, M., & Hong, C. S. (2022). "Digital Twin of Wireless Systems," IEEE Communications Surveys & Tutorials, vol. 24, no. 4, pp. 2230–2254.

Kurose, J. F., & Ross, K. W. (2021). Computer Networking: A Top-Down Approach, 8th ed., Pearson.

Lee, S., Levanti, K., & Kim, H. S. (2014). "Network Monitoring: Present and Future," Computer Networks, vol. 65, pp. 84–98.

Li, A., Yang, X., Kandula, S., & Zhang, M. (2010). "CloudCmp: Comparing Public Cloud Providers," Proc. ACM SIGCOMM Internet Measurement Conf. (IMC), pp. 1–14.

Li, T., Long, Q., Chai, H., Zhang, S., Jiang, F., Liu, H., Huang, W., Jin, D., & Li, Y. (2025). "Generative AI Empowered Network Digital Twins," ACM Computing Surveys, vol. 57, no. 6, pp. 1–43.

Mao, Q., Hu, F., & Hao, Q. (2018). "Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey," IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 2595–2621.

Mao, Y., You, C., Zhang, J., Huang, K., & Letaief, K. B. (2017). "A Survey on Mobile Edge Computing," IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2322–2358.

http://www.doi.org/10.62341/istj-vol39-1-ah18

McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Shenker, S., & Turner, J. (2008). "OpenFlow: Enabling Innovation in Campus Networks," ACM SIGCOMM Computer Communication Review, vol. 38, no. 2, pp. 69–74.

Mihai, S., Yaqoob, M., Hung, D. V., et al. (2022). "Digital Twins: A Survey on Enabling Technologies, Challenges, Trends and Future Prospects," IEEE Communications Surveys & Tutorials, vol. 24, no. 4, pp. 2255–2291.

Nguyen, T. T. T., & Armitage, G. (2008). "A Survey of Techniques for Internet Traffic Classification Using Machine Learning," IEEE Communications Surveys & Tutorials, vol. 10, no. 4, pp. 56–76.

Nunes, B., Mendonca, M., Nguyen, X., Obraczka, K., & Turletti, T. (2014). "A Survey of Software-Defined Networking," IEEE Communications Surveys & Tutorials, vol. 16, no. 3, pp. 1617–1634.

Pacheco, F., Exposito, E., Gineste, M., Baudoin, C., & Aguilar, J. (2019). "Towards the Deployment of Machine Learning Solutions in Network Traffic Classification," IEEE Communications Surveys & Tutorials, vol. 21, no. 2, pp. 1988–2014.

Rafique, W., Qi, L., Yaqoob, I., Imran, M., Rasool, R. U., & Dou, W. (2020). "Complementing IoT Services Through SDN and Edge Computing," IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1761–1804.

Scott-Hayward, S., Natarajan, S., & Sezer, S. (2016). "A Survey of Security in Software Defined Networks," IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 623–654.

Tang, T. A., Mhamdi, L., McLernon, D., Zaidi, S. A. R., & Ghogho, M. (2016). "Deep Learning Approach for Network Intrusion Detection in SDN," in Proc. Int. Conf. Wireless Networks

and Mobile Communications (WINCOM), 2016, pp. 258–263.

Varga, A., & Hornig, R. (2008). "An Overview of the OMNeT++ Simulation Environment," Proc. 1st Int. Conf. Simulation Tools and Techniques, pp. 1–10.

Villa, D., Tehrani-Moayyed, M., Robinson, C. P., Bonati, L., Johari, P., Polese, M., & Melodia, T. (2024). "Colosseum as a Digital Twin," IEEE Transactions on Mobile Computing, vol. 23, no. 10, pp. 9150–9166.

vom Lehn, H., Weingärtner, E., & Wehrle, K. (2008). "Comparing Recent Network Simulators: A Performance Evaluation Study," Aachener Informatik Berichte (AIB) 2008-16, RWTH Aachen University.

Wu, Y., Zhang, K., & Zhang, Y. (2021). "Digital Twin Networks: A Survey," IEEE Internet of Things Journal, vol. 8, no. 18, pp. 13789–13804.

Xia, W., Wen, Y., Foh, C. H., Niyato, D., & Xie, H. (2015). "A Survey on Software-Defined Networking," IEEE Communications Surveys & Tutorials, vol. 17, no. 1, pp. 27–51.

Xie, J., Yu, F. R., Huang, T., Xie, R., Liu, J., Wang, C., & Liu, Y. (2019). "A Survey of Machine Learning Techniques Applied to SDN," IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 393–430.

Zehra, S., Faseeha, U., Syed, H. J., Samad, F., Ibrahim, A. O., Abulfaraj, A. W., & Nagmeldin, W. (2023). "Machine Learning-Based Anomaly Detection in NFV," Sensors, vol. 23, no. 11, p. 5340.

Zhang, Q., Cheng, L., & Boutaba, R. (2010). "Cloud Computing: State-of-the-Art and Research Challenges," Journal of Internet Services and Applications, vol. 1, no. 1, pp. 7–18.