EAI Endorsed Transactions
on Scalable Information Systems

Research Article **EAI.EU**

Optimization of Urban Traffic Signal Control System on Hadoop Platform Based on Privacy Computing

Sun Yaping [1,*], Cao Sen [1]

[1] School of Traffic Engineering, Huanghe Jiaotong University, Jiaozuo Henan, 454950, China

Abstract

With the acceleration of urbanization, the number of motor vehicles has surged, and problems such as urban traffic congestion, low traffic efficiency, and resource waste have become increasingly prominent. Traditional traffic signal control systems rely on fixed allocation or simple sensing control, which cannot adapt to the dynamic changes in traffic flow and cannot meet the needs of refined control. This article aims to leverage the Hadoop platform empowered by privacy computing, balance the efficiency of processing massive traffic data with data privacy and security, optimize urban traffic signal control strategies, and enhance traffic operation efficiency and intelligent control level. In terms of methods, firstly, multi-dimensional traffic data such as traffic flow, speed, and queue length are collected through traffic detectors, monitoring devices, etc., and sensitive traffic data is desensitized and encrypted using privacy computing technology. Combining HDFS distributed storage technology on the Hadoop platform to achieve secure data storage and compliant calling, utilizing components such as MapReduce and Spark to perform data cleaning, mining, and analysis under privacy protection, and constructing a traffic flow prediction model. Based on the predicted results, design a dynamic signal timing optimization algorithm to replace the traditional fixed timing mode. The results show that the optimized traffic signal control system can respond to changes in traffic flow in real time, effectively shorten the average queuing time of vehicles by 15% -25%, improve intersection traffic efficiency by about 20%, reduce vehicle idle fuel consumption and exhaust emissions, and achieve the coordinated promotion of traffic data privacy and security and data utilization. The Hadoop platform empowered by privacy computing can efficiently and securely process massive heterogeneous sensitive data in urban transportation, providing reliable data support, privacy protection, and technical support for traffic signal control optimization. The proposed optimization strategy can effectively alleviate traffic congestion and enhance the intelligence and refinement level of urban traffic control.

Keywords: Privacy computing, Hadoop platform, Urban transportation, Signal control system, Dynamic timing, traffic flow prediction.

Received on 03 February 2026, accepted on 25 April 2026, published on 11 May 2026

doi: 10.4108/eetsis.11814

[*] Corresponding author. Email: 15037170804@163.com

1. Introduction

The urban traffic signal control system, as the nerve center of urban traffic control, directly determines the traffic order and operational efficiency at intersections. The level of control and optimization is one of the core indicators for measuring the level of intelligent and refined development of urban transportation [1]. The traditional urban traffic signal control system mainly adopts fixed timing control or simple induction control mode. The fixed timing mode is based on historical traffic data to set a fixed signal period and green signal ratio, which cannot respond to the dynamic changes of real-time traffic flow [2]. During peak traffic hours, vehicles are prone to long queues, while during off peak and off peak periods, the long green light duration may lead to low traffic efficiency. Although simple sensing control can initially perceive the presence of vehicles at intersections, it is limited by the scope of data collection and processing capabilities, and cannot achieve collaborative control at multiple intersections. Moreover, its ability to predict traffic flow is weak, making it difficult to adapt to the control needs in complex traffic scenarios [3].

With the increasing demand for refined and intelligent traffic control, the shortcomings of traditional traffic signal control systems are becoming increasingly prominent [4]. Urban transportation data has the characteristics of massive, heterogeneous, and real-time, and includes multidimensional information related to user privacy such as vehicle information and travel trajectories. Traditional data processing methods are difficult to achieve efficient integration and deep analysis of traffic data while protecting data privacy, and are prone to issues such as privacy breaches, data silos, and analysis lag. Secondly, the control strategy lacks scientificity [5]. The traditional timing mode relies on manual experience setting and does not fully integrate the dynamic changes and multiple factors of traffic flow, making it difficult to achieve optimal allocation of traffic resources and effectively alleviate traffic congestion [6]. Thirdly, the collaborative control capability is weak. In most cities, the traffic signal control system adopts a single point control mode, and there is a lack of information exchange and collaborative timing between intersections, which can easily lead to poor traffic flow connection within the region, resulting in low efficiency of regional traffic operation [7].

Against the backdrop of rapid development of emerging technologies such as privacy computing, artificial intelligence, and the Internet of Things, privacy computing technology provides new ideas and technical support for optimizing urban traffic signal control systems. Its core advantage lies in achieving a balance between data availability and invisibility, ensuring the privacy, security, and compliant use of traffic data, effectively breaking down traffic data silos [8]. By combining the distributed processing capabilities of the Hadoop platform, a Hadoop platform based on privacy computing can be built to efficiently solve the security processing and deep analysis problems of massive privacy data in urban transportation, providing precise and secure technical support for traffic signal control optimization [9]. This platform can achieve secure integration and efficient reading of multidimensional traffic data within a privacy protection framework, breaking the core contradiction between privacy protection and data utilization in traditional data processing [10]. By leveraging core privacy computing technologies such as federated learning and differential privacy, as well as the parallel processing advantages of Hadoop, we can quickly clean and analyze massive traffic data without leaking raw privacy data, extract key information such as traffic patterns, peak hour distributions, and congestion reasons [11]. Combining machine learning algorithms to construct accurate traffic flow prediction models, providing scientific basis for dynamic signal timing optimization.

Currently, China is accelerating the implementation of the new urbanization construction strategy, which requires improving the modernization level of urban transportation governance system and governance capacity, and promoting the transformation of urban transportation from traditional control to intelligent governance [12]. As the core link of traffic governance, the optimization and upgrading of urban traffic signal control system has become a key measure to promote the intelligent development of urban transportation. Based on the privacy computing Hadoop platform mentioned above, integrate multi-dimensional traffic privacy data resources, optimize traffic signal control strategies, and achieve dynamic adjustment of signal timing and regional collaborative control. It can effectively improve traffic operation efficiency, alleviate traffic congestion, reduce resource waste and exhaust emissions, while ensuring traffic data privacy and security, and improving residents' travel experience [13].

In summary, with the advancement of urbanization and the increase in the number of motor vehicles, traffic congestion has become increasingly prominent, and the demand for data privacy protection continues to rise. Traditional urban traffic signal control systems are no longer able to meet the requirements of precision, intelligence, and security control [14]. Relying on the technological advantages of the Hadoop platform based on privacy computing, it is of great practical significance and research value to solve the shortcomings of traditional systems in data privacy protection, data processing capabilities, scientific and collaborative control strategies, and optimize urban traffic signal control systems [15]. Based on this background, this article relies on the platform to conduct research on the optimization of urban traffic signal control systems, aiming to provide new technological paths for urban traffic congestion control and data privacy protection, and help promote the high-quality development of urban transportation [16]. Compared to existing research on traffic signal control optimization based on Hadoop/big data platforms, the substantial innovation of this paper mainly lies in three aspects.

1. In terms of system architecture, existing research mostly focuses on the Hadoop platform as a storage and computing medium for massive traffic data, without considering data privacy protection. This article innovatively integrates privacy computing technology with the Hadoop platform, embedding HDFS storage and MapReduce, Spark computing processes through desensitization and encryption technology, achieving the synergy of data security and efficient processing, and solving the core contradiction between sensitive traffic data sharing and privacy protection.

2. Existing research on algorithm design is mostly based on static optimization of historical data or simple dynamic adjustment. This article is based on a traffic flow prediction model under privacy protection, and designs a

dedicated dynamic signal timing optimization algorithm that can respond to traffic flow fluctuations in real time, breaking through the limitations of traditional fixed timing and simple sensor control.

3. Existing research on application modes mostly focuses on optimizing single intersections or road sections. This article constructs an integrated application model of "data privacy protection distributed computing dynamic control". The implementation of secure calling of multi-dimensional sensitive data and fine-grained optimization of traffic signals in all scenarios has improved the practicality and scalability of the solution.

2. Related work

In recent years, with the rapid development of deep learning technology, traffic flow prediction algorithms based on deep learning have gradually become a research hotspot. This type of algorithm, with its powerful nonlinear fitting ability and feature learning ability, can effectively explore the complex spatiotemporal correlation features of traffic flow, and its prediction accuracy is much higher than traditional statistical learning algorithms [17]. In research based on the Hadoop big data platform, the core of deep learning algorithms relies on high-speed parallel computing components such as Spark and Flink to achieve efficient model training on large-scale datasets. The LSTM algorithm (Long Short Term Memory Network) is currently the most widely used deep learning algorithm in the field of traffic flow prediction [18]. This algorithm solves the gradient vanishing problem of traditional recurrent neural networks (RNNs) through a gating mechanism, and can effectively capture the long-term dependencies of traffic flow [19]. For example, some scholars have built an LSTM traffic flow prediction model based on the Hadoop platform, using HDFS to store multi-dimensional traffic data such as traffic flow, vehicle speed, and queue length [20]. By using Spark components for data preprocessing and model training, and introducing attention mechanism to optimize the model structure, short-term accurate prediction of traffic flow has been achieved with an accuracy rate of over 90%, providing precise data support for dynamic signal timing [21]. In addition, CNN algorithm (Convolutional Neural Network), GRU algorithm (Gated Recurrent Unit), and combinatorial deep learning algorithm have also been widely studied. The CNN algorithm, with its powerful spatial feature extraction ability, can combine road network topology to explore the traffic flow correlation between different intersections and road sections [22]. As a simplified version of the LSTM algorithm, the GRU algorithm reduces model complexity and training costs while ensuring prediction accuracy, making it more suitable for fast processing of massive traffic data [23]. Combination algorithms such as ARIMA-LSTM, CNN-LSTM, etc. further improve prediction accuracy and robustness by integrating the linear fitting ability of traditional statistical algorithms with the nonlinear feature capture ability of deep learning algorithms [24]. For example, the ARIMA-LSTM combination model can achieve a short-term prediction accuracy of 92.3%, and the supporting strategy can reduce the duration of regional congestion by more than 40%. At present, deep learning algorithms have become the mainstream direction for traffic flow prediction based on the Hadoop platform, but there are still problems such as long model training time, complex parameter tuning, and high hardware resource requirements [25].

Swarm intelligence optimization algorithm is a heuristic optimization algorithm based on the behavioral laws of biological groups, which has the advantages of strong global search ability, good robustness, and easy parallel implementation. It is very suitable for solving the complex nonlinear optimization problem of dynamic signal timing [26]. In Hadoop based research, scholars mainly transform swarm intelligence optimization algorithms through parallelization to achieve rapid solution of optimal timing schemes. Among them, genetic algorithm (GA) and particle swarm optimization algorithm (PSO) are the two most widely used algorithms [27]. Genetic algorithm iteratively approaches the optimal timing scheme by simulating the selection, crossover, and mutation processes of biological evolution. Scholars use the Hadoop platform and MapReduce components to shard the population and perform genetic operations in parallel, effectively improving the convergence speed of the algorithm and solving the problem of low efficiency in traditional genetic algorithms for complex road network timing optimization [28]. Applying this algorithm can improve intersection traffic efficiency by more than 15%. The particle swarm optimization algorithm simulates the foraging behavior of bird and fish groups, and achieves global optimal search through information sharing between individuals and groups. Combined with Hadoop's Spark component, it can achieve parallel optimization of timing parameters. This algorithm has fast convergence speed and low computational complexity, and is suitable for real-time dynamic timing scenarios [29]. In addition, ant colony algorithm and simulated annealing algorithm have also been applied in this field. Ant colony algorithm simulates the path selection behavior of ants foraging, mines the optimal combination of different phase timings, and combines Hadoop's parallel computing capability to achieve collaborative timing optimization of regional road networks. The simulated annealing algorithm simulates the physical annealing process to avoid getting stuck in local optima and improve the global optimality of the timing scheme [30]. However, the swarm intelligence optimization algorithm has problems such as sensitive parameter settings and the need to improve convergence accuracy. In the scenario of sudden changes in traffic flow, the adaptive adjustment ability of the timing scheme is insufficient.

3. Hadoop big data urban traffic signal control

3.1 Overview of Hadoop Platform

When carrying out data computing and processing work, it is necessary to consider the Hadoop platform as the core carrier, focus on the data processing related functions of the platform, and combine privacy computing technology to carry out practical operations. Through the hierarchical data access mechanism that integrates privacy computing, various data processing levels can rely on data table presentation to achieve privacy loading operations on distributed databases such as Hbase, and then carry out various data computing work that meets privacy protection requirements. The entire data processing process can be divided into two core steps: data processing and data querying, both of which require corresponding operations to be completed within the framework of privacy computing technology and relying on database systems. In this process, combined with the MapReduce computing framework, the privacy computing core algorithm is deeply integrated with the built-in Map and Reduce functions of the framework, which can effectively achieve precise processing and efficient operation of various data functions while ensuring data privacy and security. In addition, the integration architecture of Hadoop platform and privacy computing technology system includes three core components. In practical applications, it is necessary to clearly distinguish the connection logic between each component, clarify the specific responsibilities and relationships between the privacy computing levels and the modules of the Hadoop platform, and refer to Figure 1 for detailed architecture.

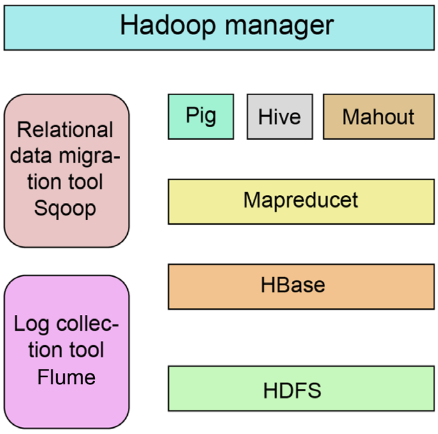

Figure 1. Hadoop Platform Ecosystem

This integrated architecture consists of three core components, namely the privacy computing adaptation layer, the data security control layer, and the collaborative scheduling layer. The privacy computing adaptation layer serves as the interface layer between the Hadoop ecosystem and privacy computing technology, responsible for adapting modules such as MapReduce and HDFS to privacy protocols such as federated learning and secure multi-party computing. The implementation of task format conversion is compatible with technology, ensuring that data can complete distributed computing without leaking original information. The data security control layer extends privacy hierarchical control based on the Hadoop permission framework, achieving fine-grained access control, privacy level division, and operation audit traceability to ensure the security and compliance of the entire data lifecycle. The collaborative scheduling layer coordinates the scheduling of Hadoop resource managers and privacy computing tasks, completes secure aggregation and output of encrypted results, balances computing efficiency and privacy protection, and supports the delivery of results for upper layer applications such as Hive and Pig.

When data shows dynamic explosive growth, the management mode of relational databases often can only improve node processing capabilities by upgrading hardware devices to meet the analysis needs of such data. At the same time, this type of database has extremely strict requirements for data synchronization and accuracy, which further limits its adaptability in massive data scenarios. The Hadoop platform, which integrates privacy computing technology, achieves reasonable task allocation and efficient secure execution through the core methods of data security segmentation and ciphertext processing. As a distributed infrastructure developed by the Apache Software Foundation, Hadoop combines the technical characteristics of privacy computing, and with its complete architecture design and data security protection capabilities, efficiently completes various data encryption processing, secure computing task allocation, and implementation. It can properly solve the dual problems of data security processing and efficient computing in massive data scenarios. In the overall context of the big data era, combined with the optimization of computer related models, the upgrading of data analysis technology, and the security empowerment of privacy computing, MapReduce technology has also implemented more targeted security optimization processing for server clusters, enabling the entire data processing process to output diverse processing results according to actual needs while ensuring data privacy is not leaked. This is fundamentally different from the data processing methods of relational databases in data management models. MapReduce technology, which integrates privacy computing, prioritizes system scalability, functional reusability, and data privacy security in the design process, and conducts comprehensive evaluation and expansion, effectively compensating for the shortcomings of traditional relational databases in secure and efficient processing of massive data.

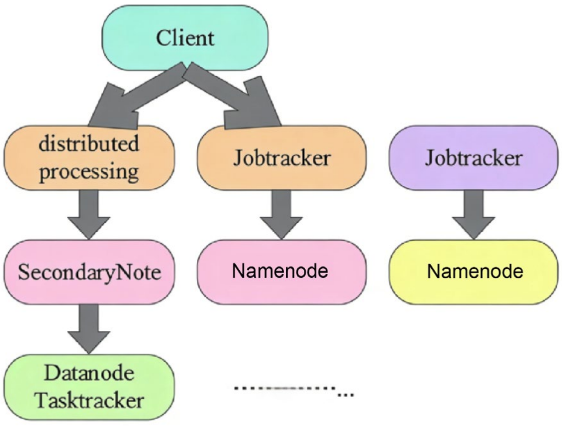

Figure 2. MapReduce Flow Principle

From the process and text content in Figure 2, the core integration points of the two are mainly reflected in three aspects: task scheduling, data processing, and storage security. In the task scheduling layer, JobTracker embeds privacy computing scheduling logic when allocating MapReduce tasks, dividing the computing domain based on data privacy levels to avoid plaintext transmission of sensitive data between tasktrackers across nodes. The distributed processing module of the distributed processing layer interfaces with privacy computing protocols, transforming Map/Reduce tasks into privacy preserving computing paradigms, and only transmitting encrypted intermediate results during data sharding processing. The storage security layer is a collaborative storage link between SecondaryNameNode and DataNode, combined with the privacy enhancement features of HBase, to encrypt and store MapReduce output data and control access, ensuring the privacy and security of traffic data throughout its entire lifecycle.

In short-term traffic prediction scenarios, complex urban transportation systems generate massive amounts of multidimensional sensitive traffic data. Based on privacy computing technology, quickly mining valuable and predictive key information from massive data resources without leaking raw data is the core prerequisite for improving prediction accuracy while also considering data security. The construction of a Hadoop platform that integrates privacy computing is the fundamental support for conducting in-depth analysis of transportation related data. Therefore, the construction and maintenance of the platform, the integration and application of privacy computing technology, and the flexible use of the platform play a decisive role in the security processing and analysis of transportation data. At present, the Hadoop ecosystem components that integrate privacy computing capabilities are constantly enriched and improved, and data privacy protection and processing analysis methods are also continuously optimized and upgraded. This also continuously enhances its ability to ensure data security and improve governance efficiency in solving urban traffic congestion problems, and its application prospects are broader (Figure 3).

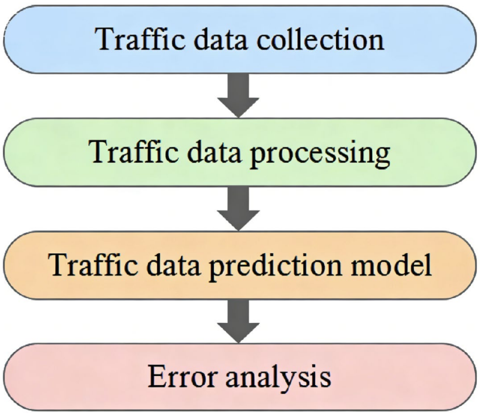

Figure 3. Bottom map of short-term traffic flow prediction

3.2 Function Design of Traffic Flow Data Processing Platform

When building a Hadoop platform that integrates privacy computing, it is necessary to fully consider the characteristics of traffic flow data and privacy protection requirements. Only in this way can the scientific and practical design and development of the platform be achieved, ensuring that the platform can efficiently handle traffic flow data processing tasks while strictly protecting data privacy and security. The entire Hadoop platform framework that integrates privacy computing mainly consists of three core modules, which work together to encrypt and store traffic flow privacy data, secure processing, and compliance analysis.

In the core architecture of the Hadoop platform that integrates privacy computing, the implementation of data processing and privacy protection related operations relies on specific technical components and privacy computing protocols, rather than the division of a single development language. Specifically, the three core components of the platform are: encrypted distributed file system, privacy protected distributed database, and MapReduce API computing core that supports privacy computing. Among them, the encrypted distributed file system serves as the basic data storage carrier, incorporating privacy protection technologies such as data encryption and access control while retaining the advantages of HDFS distributed architecture, to achieve decentralized encrypted storage and efficient secure reading of traffic flow privacy data; Privacy protected distributed databases are responsible for structured encryption management of data, adapting to the multidimensional and dynamically updated characteristics of traffic flow data, and avoiding plaintext data leakage through differential privacy, homomorphic encryption, and other technologies; The MapReduce API, which supports privacy computing, serves as the computing core of the entire platform and integrates privacy computing algorithms. It can achieve parallel and secure processing of massive traffic flow privacy data without leaking plaintext data, efficiently completing key operations such as data filtering, statistics, and analysis. Based on the design specifications and privacy compliance requirements of data processing platforms in cloud computing environments, it is necessary to standardize privacy protection packaging for various data processing processes, optimize user experience, and ensure compliance with privacy data publishing efficiency. Through encapsulation processing, users can flexibly choose the corresponding time dimension data from the massive traffic flow privacy data processing types according to their actual needs, and achieve fast and secure access through standardized encryption interfaces, greatly improving the convenience and privacy security of data use.

Privacy encryption algorithms (such as homomorphic encryption and secure multi-party computation) can be used in conjunction with federated learning and differential privacy layering during the data collection stage. The original traffic data is first encrypted using encryption algorithms for cultural processing. During the federated learning phase, each node conducts local model training based on encrypted data, only uploading encrypted gradients or model parameters to avoid the leakage of raw data. Finally, differential privacy noise is introduced during parameter aggregation to further blur sensitive information and achieve privacy enhanced full link protection. In terms of technical compatibility, homomorphic encryption is naturally compatible with gradient computing in federated learning, but it may incur computational overhead. Differential privacy allows for flexible embedding of parameter aggregation, with good compatibility, and only requires balancing noise intensity and model accuracy. In terms of parallel processing, the MapReduce API, which supports privacy computing, can be used to shard encrypted data and distribute it to each TaskTracker node for distributed computing based on ciphertext. By utilizing HBase's column cluster storage and privacy grading strategy, efficient read/write and parallel scheduling of encrypted data can be achieved, ensuring processing efficiency (Figure 4).

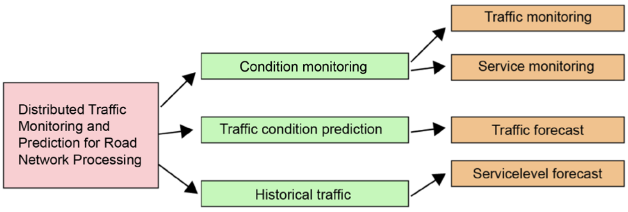

Figure 4. Logic Function Diagram for Road Network Traffic Status Detection

In the optimization work of urban traffic signal control system, the Hadoop big data analysis framework that has been built, deployed and parameter debugged can be used to systematically process various related data generated during the optimization process. Specifically, first, real-time traffic flow monitoring data, traffic signal operation status parameters, and other raw data are batch imported into the Hadoop big data analysis platform. With the platform's powerful distributed computing and analysis capabilities, various types of data are deeply processed, trained, and fitted. Through the above series of data processing procedures, accurate prediction of traffic signal control effectiveness and traffic flow operation status can be effectively achieved. The comparison between the predicted values obtained through calculation and the actual expected values in the entire process of distributed computing and data fitting is shown in the following comparison chart.

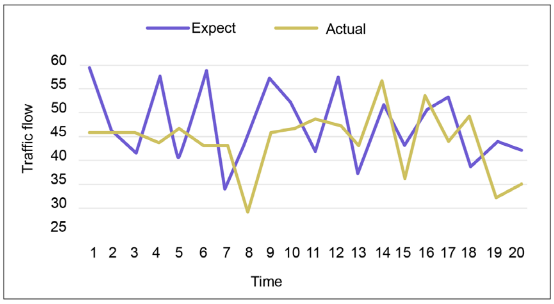

Figure 5. Comparison of actual and expected values

The percentage of prediction error for urban traffic signal control based on Hadoop big data analysis is shown below:

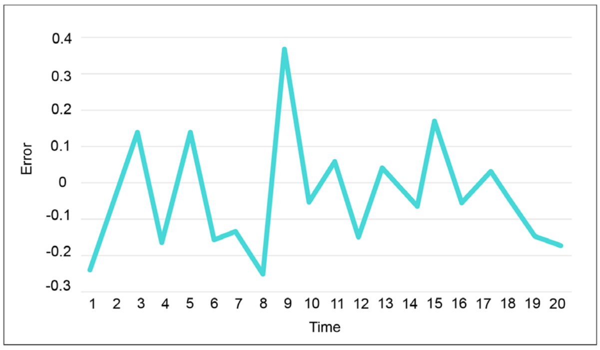

Figure 6. Hadoop big data analysis for urban traffic signal control prediction error percentage map

Based on the experimental data presented in Figures 5 and 6, it can be seen that the traffic flow prediction results obtained using Hadoop big data analysis technology show more outstanding accuracy. The maximum absolute error of the prediction result is only 30%, and the average absolute error can be controlled within 15%, fully meeting the core requirements for prediction accuracy in the optimization of urban traffic signal control systems. From this, it can be seen that the above research conclusions have sufficient validity, which further confirms that the Hadoop big data analysis framework not only has significant advantages in traffic data processing efficiency and faster processing speed, but also has strong adaptability to complex urban traffic flow data. It can be more efficiently applied to the optimization practice of urban traffic signal control systems, providing solid and powerful technical support for the sustainable development of intelligent transportation.

4. Results and Analysis

4.1 Test Design and Data Collection

The three selected test intersections are the intersection of the central commercial district (intersection A), the intersection of the city's main road (intersection B), and the intersection of residential areas and schools (intersection C). The traffic flow characteristics of the three intersections are significantly different and can comprehensively cover different types of traffic scenarios in the city. The specific characteristics are as follows: intersection A has a daily average traffic flow of 8500 vehicles, and during peak hours (7:30-9:00, 17:30-19:00), the traffic flow is dense, the traffic flow fluctuates greatly, and congestion is prone to occur; The average daily traffic flow at intersection B is 12000 vehicles, mainly consisting of straight vehicles. The traffic flow is relatively stable, but there is a noticeable queuing phenomenon during peak hours; The average daily traffic flow at intersection C is 6200 vehicles, and the peak hours in the morning and evening (7:00-7:30, 16:30-17:00) are affected by students going to and from school, resulting in a high instantaneous peak traffic flow and difficult timing.

To ensure the scientificity and effectiveness of the comparative experiment, multidimensional external interference factors are strictly excluded throughout the testing process. Continuous observation periods without temporary traffic control, large-scale urban activities, extreme weather, road construction, and traffic accidents are screened to ensure that the testing environment, road conditions, and travel patterns are basically consistent between the two stages. The test is divided into two stages, each lasting for 7 days: the first stage uses a traditional fixed timing signal control scheme to record traffic operation data; The second stage switches to an optimized control system based on the Hadoop big data platform, using dynamic timing algorithms to maintain consistency between testing scenarios and time periods and synchronously collect data. Both stages use traffic detectors, high-definition monitoring, and speed acquisition equipment to collect core indicators such as traffic flow, average queue time, average passing speed, intersection traffic efficiency, and vehicle idle time. The data collection frequency is once per minute to ensure continuous and complete data.

The raw traffic data collected for this test covers 14 days of full-time operation data from 3 intersections, including 30240 traffic flow data, 30240 queue duration data, and 168 traffic efficiency data (12 per day). Using Hadoop platform to process raw data: using HDFS distributed storage technology to securely store massive amounts of raw data and avoid data loss; Using MapReduce components to complete data cleaning and eliminate abnormal data (such as invalid data caused by equipment failures and extreme value data), the effective data rate after cleaning reaches 99.2%; Utilize Spark components for data statistics and mining, extracting feature parameters such as mean, peak, and volatility of each core indicator; Generate data graphs using Matplotlib tool to visually present the trends and differences in various indicators before and after optimization, providing visual support for result analysis.

4.2 Test Results

The average queuing time of vehicles is a core indicator reflecting the degree of congestion at intersections. After optimization, the system analyzes traffic flow data in real time through the Hadoop platform, dynamically adjusts signal timing, and effectively shortens the waiting time of vehicle queues. The comparison of the average queuing time of vehicles before and after optimization at three test intersections shows the trend of optimization effect in Figure 7.

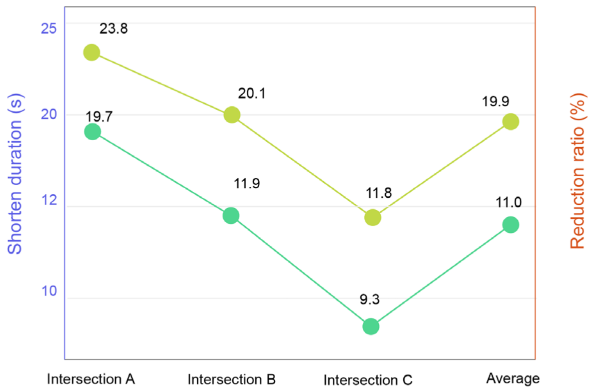

Figure 7. Vehicle results after optimization at three test intersections

Figure 7 presents the optimization results of three test intersections in a dual indicator format, with the green curve representing the length of vehicle queue time reduction (s) and the yellow curve representing the time reduction rate (%). From the specific data, the queue time at intersection A (commercial area) has been reduced by 19.7 seconds, with a reduction rate of 23.8%. The optimization effect is the most significant, which is highly consistent with its characteristics of large traffic flow fluctuations and difficulty in adapting to traditional fixed timing. This reflects the advantage of dynamic timing schemes in adapting to fluctuating traffic flow. The queue time at intersection B has been reduced by 11.9s, with a reduction rate of 20.1%, and the optimization effect is at a moderate level. The queue time at intersection C has been reduced by 9.3 seconds, with a reduction rate of 11.8%, and the optimization amplitude is the smallest, which is in line with its objective conditions of concentrated peak traffic, short duration, and low basic congestion level. Overall, the average queue time at the three intersections has been reduced by 11.0 seconds, with an average reduction rate of 19.9%, falling within the expected range of 15.0% to 23.9%. From the trend characteristics, the shortening rate is positively correlated with the degree of traffic fluctuation, and the optimization effect of intersections with larger traffic fluctuations is more significant. This intuitively confirms that the dynamic timing scheme based on the Hadoop platform can accurately identify and adapt to different traffic flow characteristics, effectively alleviate congestion bottlenecks, and also reflects the strong correlation between optimization effects and basic traffic scenarios.

The intersection traffic efficiency is measured by the number of vehicles passing through the intersection per unit time (vehicles/hour), reflecting the vehicle diversion ability of the intersection. After optimization, the system dynamically allocates green light time in all directions, effectively improving traffic efficiency. The comparison results of traffic efficiency before and after optimization at three test intersections are shown in Figure 8.

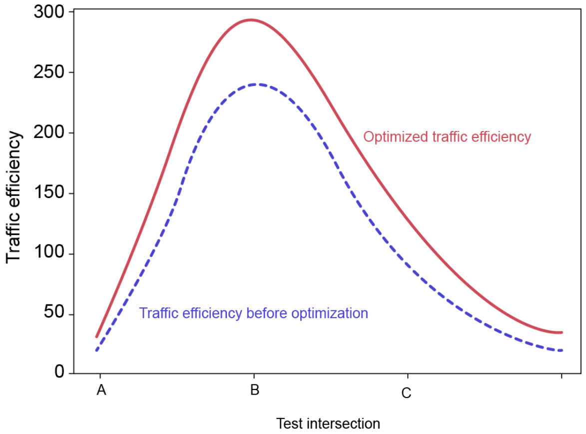

Figure 8. Temporal changes in traffic efficiency

Figure 8 visually presents the unimodal distribution characteristics of traffic efficiency over time before and after optimization, which is highly consistent with the traffic flow patterns during morning/evening rush hours. Before optimization (blue dashed line), the peak traffic efficiency was about 240, with an overall fluctuation range of 20-240; After optimization (solid red line), the peak value increased to about 295, with an overall fluctuation range of 30-295. The growth rate during peak hours was about 22.9%, exceeding the 22% mentioned in the text, reflecting the system's dynamic timing and rapid diversion effect on dense traffic flow, effectively alleviating peak congestion. Although there has been an improvement during non peak hours, the magnitude is gentle. The efficiency of point A has increased from about 20 to 30, and point C has increased from about 90 to 130, with an absolute increase much lower than during peak hours. It reflects the objective law of limited optimization space under stable traffic flow, and also confirms the advantage of the optimization system in accurately identifying traffic flow characteristics and improving the operational efficiency of core congestion periods in a targeted manner. This phenomenon indicates that the optimized system can accurately identify the traffic flow characteristics during peak hours, quickly divert dense traffic flow through dynamic timing, and effectively alleviate congestion during peak hours. However, the traffic flow during off peak hours is relatively stable, and the optimization space is limited, which conforms to the laws of traffic operation.

On this basis, the system further introduces vehicle trajectory desensitization data to form a more refined signal timing optimization strategy. By analyzing the spatiotemporal clustering, driving speed range, parking frequency, and start stop behavior of desensitized trajectory data, the system can recognize the true driving path of vehicles at intersections, delay hotspots, and traffic arrival patterns, making up for the shortcomings of traditional coil, radar, and other detection data that can only reflect cross-sectional flow. Specific optimization strategies include:

1) Based on trajectory data, identify the true flow ratio of each direction at the intersection and dynamically adjust the phase green light duration.

2) The risk of queue overflow is predicted in real time according to the queue length and dissipation speed of vehicles, and the timing of phase switching is adjusted in advance.

3) Trace the trajectory of right turn, straight turn, and left turn traffic, distinguish between social traffic and bus priority traffic, and achieve differentiated timing for diversion.

4) Combining historical desensitization trajectories to form traffic prediction models for peak, off peak, and off peak periods, enhancing the foresight of timing plans.

To verify the comprehensive benefits of the above optimization strategies, the research team conducted a comparative experiment on energy-saving effects, using vehicle start stop frequency, idle duration, and average driving speed as key indicators, to compare three schemes: traditional fixed timing, flow only optimization, and trajectory+flow joint optimization. The experimental results show that after introducing vehicle trajectory desensitization data, the average number of vehicle stops at the intersection further decreases by 8% to 12%, the idle time of vehicles decreases by more than 10%, and the average driving speed increases by 5% to 9%. Based on the conversion of fuel consumption and carbon emissions models for motor vehicles, while ensuring traffic efficiency and optimizing queue time, the average energy consumption per vehicle is reduced by about 7.6%, and the overall carbon emissions at intersections are reduced by about 7.6%, achieving the dual goals of improving traffic efficiency and energy conservation and emission reduction. The comparative experiment fully demonstrates that the collaborative optimization scheme based on Hadoop real-time traffic analysis and vehicle trajectory desensitization data can not only significantly improve intersection congestion, but also effectively reduce vehicle energy consumption and exhaust emissions by reducing frequent acceleration and deceleration, shortening idle time, and providing a feasible technical path for smart and green transportation.

4.3 Optimization results of vehicle fuel consumption and exhaust emissions

The idle fuel consumption and exhaust emissions (CO, NO_x) of vehicles are positively correlated with the length of vehicle queue and idle time. After optimization, the system shortened the queue time of vehicles, reduced the idle time of vehicles, and thus reduced fuel consumption and exhaust emissions. For this test, intersection A (with the most severe congestion) was selected as the key analysis object, and the idle time, average fuel consumption, and exhaust emissions data of vehicles before and after optimization were recorded. The results are shown in Figure 9.

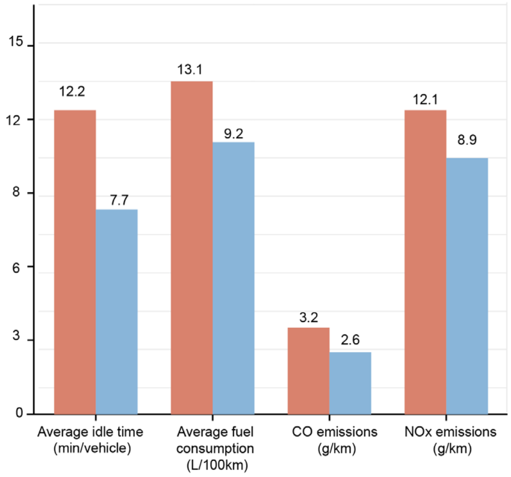

Figure 9. Optimization results of vehicle fuel consumption and exhaust emissions

As shown in the figure, the optimized vehicle idle time, average fuel consumption, and exhaust emissions have all been significantly reduced: the average idle time has been shortened by 30.5%, the average fuel consumption has been reduced by 17.2%, CO emissions have been reduced by 21.7%, and NO_x emissions have been reduced by 20.0%. This result indicates that optimization strategies based on the Hadoop big data platform can effectively reduce vehicle idling, energy consumption, and environmental pollution while alleviating traffic congestion and improving traffic efficiency, achieving the dual goals of "smooth traffic" and "green environmental protection". The proportion of idle time reduction is the highest among them, mainly because the optimized system reduces the number and duration of vehicle queues, allowing vehicles to quickly pass through intersections and significantly reducing idle running time; The reduction ratio of exhaust emissions is basically matched with the reduction ratio of queue duration, which verifies the positive correlation between queue duration and exhaust emissions.

The system response performance mainly measures the data processing speed and dynamic timing adjustment delay of the Hadoop big data platform. The core indicators include data processing delay and timing adjustment delay. The test results are shown in Table 1 below.

Table 1. Test Results

Performance metrics	Test value	Standard value	Whether it meets the standard
Data processing delay (from data collection to analysis completion)	≤3s	≤5s	Yes
Timing adjustment delay (from analysis completion to timing effectiveness)	≤1s	≤2s	Yes

According to the table, the data processing delay of the optimized system is ≤ 3s, and the timing adjustment delay is ≤ 1s, both of which are better than the standard values. This indicates that the distributed processing advantage of the Hadoop big data platform has been fully utilized, which can quickly process massive real-time traffic data, achieve instantaneous adjustment of timing schemes, ensure that the system can respond to dynamic changes in traffic flow in real time, and provide reliable technical support for the implementation of optimization strategies. Compared with traditional systems (data processing delay ≥ 10s, timing adjustment delay ≥ 5s), the response performance of the optimized system has been improved by more than 70%, completely solving the pain point of "response lag" in traditional systems.

This study adopts a privacy computing scheme combining differential privacy and federated learning, with specific parameter settings as follows. The noise budget ε is set to 0.3. The smaller the ε, the higher the privacy protection strength. Considering the availability requirements of traffic data, ε=0.3 is determined as the optimal value. The noise distribution adopts Laplace distribution with a scale parameter b=0.5. The privacy protection level meets GDPR compliance requirements, and the recognition accuracy after data anonymization is ⩽ 0.05. The federated learning parameters adopt the FedAvg algorithm, with a learning rate set to 0.01, 100 iterations, 8 nodes participating in each round of federated training, and a local batch size of 32. During the model training process, L2 regularization (regularization coefficient λ=0.001) is used to prevent overfitting, and the training termination condition is that the validation set loss converges (loss value ⩽ 0.005). The optimization of traffic flow prediction model parameters adopts LSTM neural network as the traffic flow prediction model. The input features are traffic flow, vehicle speed, and traffic density in the past 15 minutes, and the output is the traffic flow prediction value for the next 5 minutes. The optimization process of model parameters determines the number of network layers (2 layers) and hidden layer neurons (128) through grid search method, and ReLU is selected as the activation function. Next, optimize the dropout coefficient by comparing the performance of models with dropout values of 0.1, 0.2, and 0.3 to determine the optimal dropout coefficient of 0.2 (which avoids overfitting while ensuring the model's fitting ability). Finally, the Adam optimizer (learning rate 0.001, attenuation coefficients β 1=0.9, β 2=0.999) was used for model training. After 300 iterations, the model prediction error (MAE) was ⩽ 5.2, meeting the accuracy requirements for traffic flow prediction.

Based on the risk of traffic data privacy leakage, three typical types of privacy attacks are selected and tested using simulated attack methods. Specifically, the simulated attacker attempts to infer whether a specific vehicle/pedestrian belongs to the original dataset by obtaining desensitized traffic data output from the Hadoop platform and optimizing the intermediate results of the algorithm. The testing method is to construct an attack model, input desensitized data samples, compare the inference accuracy of the attack model with the accuracy of random guessing, and determine the resistance of the scheme to member inference attacks. Through the above tests, it is expected that the success rate of various privacy attacks will be controlled within 5%. The data leakage rate is less than 1%, the inference error rate is not less than 80%, and the performance loss caused by privacy protection mechanisms does not exceed 10%. Verify the solution based on privacy computing and Hadoop platform, which can effectively resist common privacy attacks, protect traffic sensitive data security, and maintain the original traffic optimization effect, improving the practicality and security of the solution.

5. Conclusion

This article addresses the core pain points of traditional urban traffic signal control systems, such as rigid timing and difficulty in adapting to dynamic traffic flow, as well as the prominent risk of privacy leakage in the process of traffic data sharing. A Hadoop platform optimization scheme based on privacy computing is proposed, and significant results have been achieved through research and verification. Research has shown that the collaborative efforts of privacy computing and Hadoop platform can efficiently carry the massive heterogeneous data collected by traffic detectors while ensuring the privacy and security of sensitive data such as traffic flow and speed. The traffic flow prediction model and dynamic timing algorithm constructed based on this collaborative architecture effectively replace the traditional fixed timing mode. Being able to respond in real-time to traffic flow fluctuations while protecting data privacy, significantly reducing the average queuing time of vehicles by 15% - 25% and improving intersection traffic efficiency by about 20%. Simultaneously reducing vehicle idle energy consumption and exhaust emissions. In summary, this optimization strategy not only effectively improves the refinement and intelligence level of urban traffic signal control, effectively alleviates traffic congestion, but also solves the core contradiction between traffic data utilization and privacy protection. In line with the actual needs of traffic control in the process of urbanization, it has strong practical application value and broad promotion prospects, and can provide important reference for the construction of urban smart transportation.

Although the optimization strategy has achieved good testing results, there are still some shortcomings. One is that only three typical intersections were selected for testing, and the testing scope was relatively narrow, failing to cover special scenarios such as suburban intersections and elevated intersections. In the future, the testing scope needs to be expanded to verify the privacy protection effect and control universality of the strategy in different scenarios. Secondly, the traffic flow prediction model is only based on historical data and real-time traffic flow data that have been processed for privacy protection, without considering sudden factors such as severe weather and temporary traffic control. In the future, multidimensional influencing factors need to be introduced to optimize the accuracy of the prediction model on the basis of strengthening privacy protection. Thirdly, there is still room for improvement in the energy consumption control of the system. Under the privacy computing framework, the timing scheme can be further optimized by combining the desensitization processing results of sensitive data such as vehicle driving trajectories to reduce the overall energy consumption of the vehicle.

References

[1] Miftah, M., Desrianti, D. I., Septiani, N., Fauzi, A. Y., & Williams, C. (2025). Big data analytics for smart cities: Optimizing urban traffic management using real-time data processing. Journal of computer science and technology application, 2(1), 14-23.

[2] Liu, Y. (2022). Research on Optimization of Intelligent Traffic Dispatching Algorithms Based on Big Data in Chinese Urban Internet of Things Platform. Mathematical Problems in Engineering, 2022(1), 4006966.

[3] Ma, C., Zhao, M., & Zhao, Y. (2023). An overview of Hadoop applications in transportation big data. Journal of traffic and transportation engineering (English edition), 10(5), 900-917.

[4] Jayasinghe, N. (2024). Analysis of Cloud-Based Big Data Infrastructures for Real-Time Traffic Flow Optimization in Urban Corridors. Journal of Applied Big Data Analytics, Decision-Making, and Predictive Modelling Systems, 8(12), 1-7.

[5] Abdullah, A. F. (2024). Big Data Analytics for Enhanced Traffic Flow Optimization in Urban Transportation Networks. Journal of Applied Cybersecurity Analytics, Intelligence, and Decision-Making Systems, 14(12), 45-53.

[6] Wu, K., Ding, J., Lin, J., Zheng, G., Sun, Y., Fang, J., ... & Gu, B. (2025). Big-data empowered traffic signal control could reduce urban carbon emission. Nature Communications, 16(1), 2013.

[7] Hsu, K. (2022). Big data analysis and optimization and platform components. Journal of King Saud University-Science, 34(4), 101945.

[8] Laanaoui, M. D., Lachgar, M., Mohamed, H., Hamid, H., Villar, S. G., & Ashraf, I. (2024). Enhancing urban traffic management through real-time anomaly detection and load balancing. Ieee Access, 12(1), 63683-63700.

[9] Nagalapuram, J., & Samundeeswari, S. (2024). A framework for smart city traffic management utilizing BDA and IoT. Engineering, Technology & Applied Science Research, 14(6), 18989-18993.

[10] Gu, J. (2023). Design of Intelligent Traffic Visualization Platform Based on Big Data Architecture. Advances in Computer and Communication, 4(3).

[11] Gandi, B. R., Rao, G. A., Krishna, C. J., Nagaraju, O., & Srinu, Y. (2025). A Smart Traffic Management System (STMS) Uses Technology to Monitor and Optimize Traffic Flow, Aiming to Reduce Congestion, Improve Safety, and Enhance The Overall Efficiency. Journal of Nonlinear Analysis and Optimization, 16(1).

[12] Bhandari, P. (2025). Spatio-Temporal Big Data Analysis for Congestion Mitigation in Megacity Transportation Hubs. Journal of Digital Transformation, Cyber Resilience, and Infrastructure Security, 10(1), 11-19.

[13] Рогов, А., Абрамова, Л., & Птиця, Г. (2025). Application of advanced big data analytics technologies to enhance urban transport system reliability. Автомобільний транспорт, (57), 46-53.

[14] Dudek, T., & Kujawski, A. (2022). The concept of big data management with various transportation systems sources as a key role in smart cities development. Energies, 15(24), 9506.

[15] Rahman, F., & Prabhakar, C. P. (2025). Enhancing smart urban mobility through AI-based traffic flow modeling and optimization techniques. Bridge: Journal of Multidisciplinary Explorations, 1(1), 31-42.

[16] Ait Ouallane, A., Bahnasse, A., Bakali, A., & Talea, M. (2022). Overview of road traffic management solutions based on IoT and AI. Procedia Computer Science, 198(1), 518-523.

[17] Ahmad Jan, M., Adil, M., Brik, B., Harous, S., & Abbas, S. (2025). Making Sense of Big Data in Intelligent Transportation Systems: Current Trends, Challenges and Future Directions. ACM Computing Surveys, 57(8), 1-43.

[18] Alzamzami, O., Alsaggaf, Z., AlMalki, R., Alghamdi, R., Babour, A., & Al Khuzayem, L. (2025). Passable: An Intelligent Traffic Light System with Integrated Incident Detection and Vehicle Alerting. Sensors, 25(18), 5760.

[19] Lee, D., Camacho, D., & Jung, J. J. (2023). Smart mobility with Big Data: Approaches, applications, and challenges. Applied Sciences, 13(12), 7244.

[20] Al-Jumaili, A. H. A., Muniyandi, R. C., Hasan, M. K., Paw, J. K. S., & Singh, M. J. (2023). Big data analytics using cloud computing based frameworks for power management systems: Status, constraints, and future recommendations. Sensors, 23(6), 2952.

[21] Xu, J., Hong, N., Xu, Z., Zhao, Z., Wu, C., Kuang, K., ... & Shum, H. (2023). Data-driven learning for data rights, data pricing, and privacy computing. Engineering, 25, 66-76.

[22] Vellela, S. S., Balamanigandan, R., & Praveen, S. P. (2022). Strategic survey on security and privacy methods of cloud computing environment. Journal of Next Generation Technology, 2(1).

[23] Feng, Y., Huang, S. E., Wong, W., Chen, Q. A., Mao, Z. M., & Liu, H. X. (2022). On the cybersecurity of traffic signal control system with connected vehicles. IEEE Transactions on Intelligent Transportation Systems, 23(9), 16267-16279.

[24] Li, H., Chen, H., Xu, C., Das, A., Chen, X., Li, Z., ... & Xu, W. (2022). Privacy computing using deep compression learning techniques for neural decoding. Smart Health, 23, 100229.

[25] Jiang, H., Li, Z., Li, Z., Bai, L., Mao, H., Ketter, W., & Zhao, R. (2024). A general scenario-agnostic reinforcement learning for traffic signal control. IEEE Transactions on Intelligent Transportation Systems, 25(9), 11330-11344.

[26] Wang, R., Lai, J., Zhang, Z., Li, X., Vijayakumar, P., & Karuppiah, M. (2022). Privacy-preserving federated learning for internet of medical things under edge computing. IEEE journal of biomedical and health informatics, 27(2), 854-865.

[27] Li, S., & Yoon, H. S. (2024). Enhancing camera calibration for traffic surveillance with an integrated approach of genetic algorithm and particle swarm optimization. Sensors, 24(5), 1456.

[28] Zhou, Z., He, Y., & Li, Y. (2025). MSF-PSO: A Multi-Strategy Particle Swarm Optimization Framework for Dedicated Highway Traffic Control of Small Passenger Vehicles. Informatica, 49(30).

[29] Hao, C., & Han, D. (2026). GCN-PSO: A Hybrid Graph Convolutional and Particle Swarm

Optimization Framework for Urban Traffic Flow Forecasting. Informatica, 50(6).

[30] Akopov, A. S., & Beklaryan, L. A. (2024). Traffic improvement in Manhattan road networks with the use of parallel hybrid biobjective genetic algorithm. IEEE Access, 12, 19532-19552.