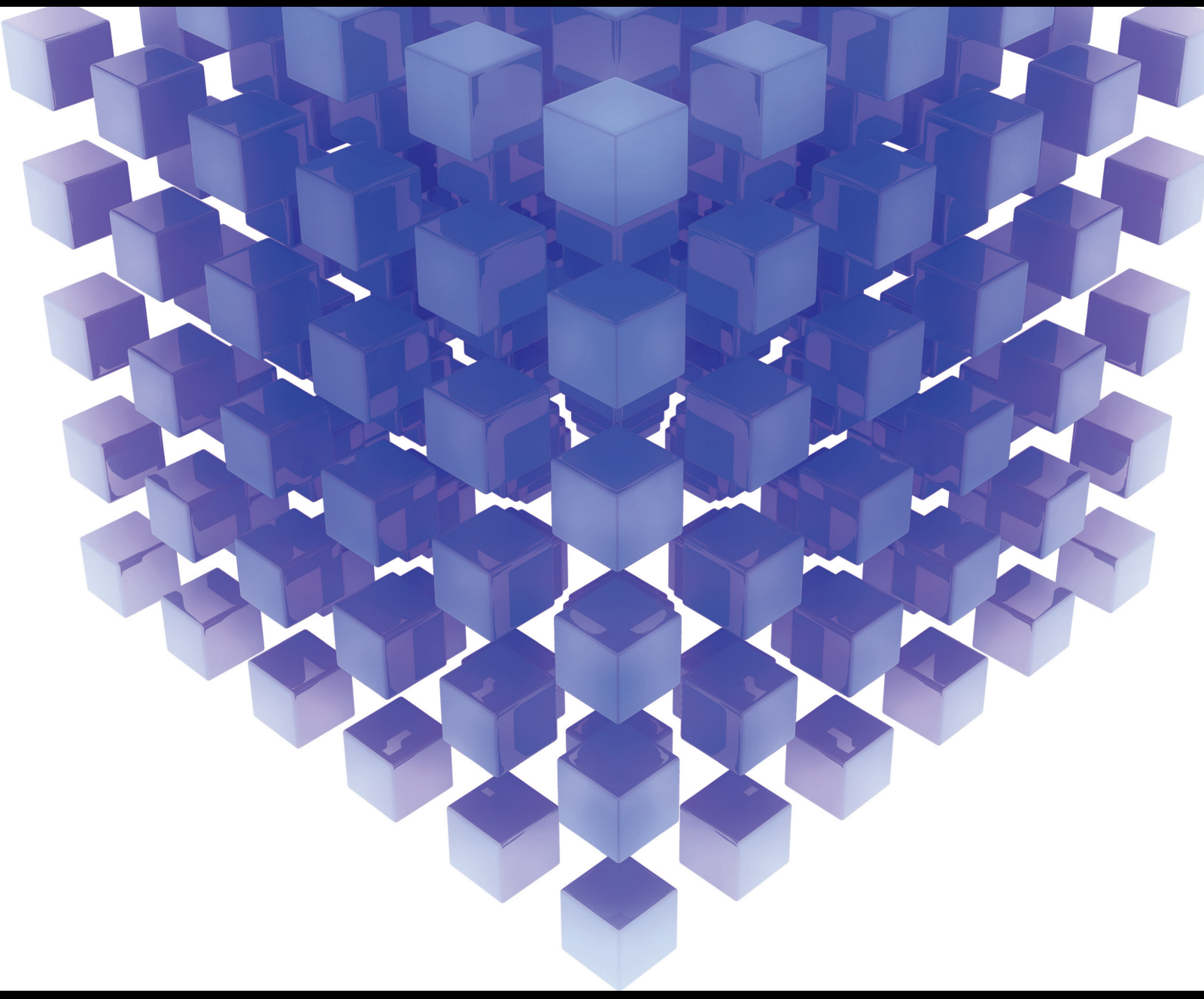
Mathematical Problems in Engineering

Control of Networked Systems with Engineering Applications 2021

Lead Guest Editor: Hou-Sheng Su
Guest Editors: Michael Z. Q. Chen, Zhiwei Gao, Xiao Ling Wang, and Sahar Yazdani



Control of Networked Systems with Engineering Applications 2021

Mathematical Problems in Engineering

Control of Networked Systems with Engineering Applications 2021

Lead Guest Editor: Hou-Sheng Su
Guest Editors: Michael Z. Q. Chen, Zhiwei Gao, Xiao Ling Wang, and Sahar Yazdani

Chief Editor

Academic Editors

Contents

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9954599, 13 pages
https://doi.org/10.1155/2021/9954599

Research Article

Blockchain-Based Authentication with Optional Privacy Preservation for Internet of Vehicles

Jinxin Zhang and Meng Wu

School of Computer Science & Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Correspondence should be addressed to Meng Wu; wum@njupt.edu.cn

Received 30 March 2021; Accepted 10 June 2021; Published 19 October 2021

Academic Editor: Hou-Sheng Su

With the rapid development of the mobile internet and intelligent technology of in-vehicle equipment, the Internet of Vehicles (IoV), centered on intelligent connected cars, has gradually entered people's lives. However, these technologies also bring serious privacy risks and security issues in terms of data transmission and storage. In this article, we propose a blockchain-based authentication system to provide vehicle safety management. The privacy and security attributes of various vehicle authentication transactions are based on high-level cryptographic primitives, realizing temporary and formal authentication methods. At the same time, a fair blockchain consensus mechanism Auction of block generation Rights (AoR) is proposed. To demonstrate the feasibility and scalability of the proposed scheme, security and performance analyses are presented. The relevant experimental results show that the scheme can provide superior decentralized management for IoV.

1. Introduction

In recent years, as a potential technology, the Internet of Vehicles [1] has attracted great attention, bringing a better life to human beings. It can be applied in the fields of logistics and transportation. According to a report [2] from the World Health Organization on road safety, in 2018, a total of 1.35 million people died in traffic accidents worldwide, and traffic accidents are the number one killer in the 5–29 age group. The Internet of Vehicles can solve such problems well. This is just one example of many applications of the Internet of Vehicles.

The Internet of Vehicles usually consists of a large number of static basic settings and dynamic vehicles as the main participants in the system. It is equipped with advanced in-vehicle sensors, controllers, actuators, and other devices and integrates modern communication and network technologies to realize the information exchange and sharing between vehicles and vehicles, roads, and service providers. It has complex environmental perception, intelligent decision-making, and collaborative control and performs other functions. The purpose of the Internet of Vehicles is to avoid unnecessary traffic accidents and congestion. The ultimate goal is to provide a comfortable driving experience including autonomous driving and in-vehicle entertainment.

The Internet of Vehicles technology system mainly includes the automotive sensor, automotive wireless communication, automotive navigation, electronic map and positioning, vehicle-mounted Internet of Things terminal, intelligent control, massive data processing, data integration, intelligent transportation technology, video surveillance, and mobile communication network. It differs from other types of network application scenarios; thus, all these technologies complement each other and cooperate to achieve. The future of the Internet of Vehicles system will face the requirements of system function integration, dataset quantification, and high transmission rate.

With the gradual evolution of closed-loop information services to IoV services and the close integration of vehicle, road, and surrounding environmental data, the Internet of Vehicles will more effectively reduce the incidence of car accidents and provide a safer, more economical, and more convenient travel service. On the Internet of Vehicles, the

vehicle must periodically broadcast the identity, current location, speed, and other related information of the traffic vehicle to all vehicles around it. Malicious vehicles can obtain the private information (identity, location, etc.) of the vehicle driver by analyzing the relationship between the message and the sender. This poses a potential threat to the privacy of vehicle users, increases the risk of data leakage and may potentially affect the safety of vehicle occupants [3, 4].

Authentication is considered the first line of defense against malicious vehicles and messages [5–7]. It is the basis for the security of all other applications of the Internet of Vehicles in the open traffic environment. Identity authentication includes the verification of the legitimacy of the identity of the connected vehicles, to ensure the authenticity of the identity of the communicating parties. At the same time, it is necessary to protect the privacy of users through anonymity [8–10]. Therefore, the automotive industry needs to establish more secure authentication methods to avoid this risk.

The characteristics that need to be considered in the identity authentication of the car network are as follows:

(1) Due to the fast-moving speed of the vehicle and the limited coverage of the roadside unit (RSU), high real-time identity verification is required.

(2) The environments in the IoV vary widely, with diverse application scenarios and performance differences between devices. Transaction processing may be delayed. Therefore, such realistic factors need to be considered to build a fair authentication environment so that vehicles can have equal opportunities to be authenticated.

The Internet of Vehicles currently uses centralized facilities, Public Key Infrastructure (PKI), or trust authority (TA) to perform the authentication mechanism. The disadvantage is that the centralization of the authentication node leads to heavy tasks and attacks on the central node. These shortcomings can cause data leakage of sensitive user information. Besides, the centralized facility is not suitable for the wide and complex geographically distributed IoV and high real-time requirements.

To address the issue, some scholars [11–13] use blockchain [14] to develop a decentralized scheme, which provides a secure way of managing vehicle registration. Some review articles [15–18] believe that blockchain can provide a boost to the Internet of Things. The essence of the blockchain is a distributed ledger database of a peer-to-peer (P2P) network. A complete blockchain system includes technologies such as data encryption, digital signatures, and timestamps, as well as consensus algorithms to support P2P and maintenance systems, mining, and anonymous transactions. The blockchain can also be applied in many fields [19–23] with its unique security mechanism.

In this paper, we propose an identity authentication system for IoV. Our system utilizes the following advantages of the blockchain:

(i) We use the common private key, public key, and address based on elliptic curve cryptography (ECC) in the blockchain to achieve anonymous security authentication, without the need to construct other cryptographic primitives.

(ii) Based on the decentralization nature of blockchain, trust management can be conducted among distributed RSUs, which can effectively avoid the problem of centralization. To sink a large number of vehicle registration functions, reduce the delay of neutralization, and reduce security risks.

(iii) We assign different node roles to vehicles and facilities and make the best use of them and enable RSUs to work together and maintain a consistent blockchain, and the vehicle node ensures that the basic information of the vehicle is maintained.

Thus, the key contributions of this paper can be summarized as follows:

(1) For high real-time requirements, we propose a decentralized vehicle registration with TA providing authentication keys for each RSU.

(2) We achieve optional privacy-preserving authentication for diverse scenarios. It includes short-term temporary authentication and long-term formal authentication.

(3) We propose a fair consensus mechanism to equalize the opportunity to process things on different devices and ensure that vehicle information can be treated equally.

The rest of this paper is organized as follows: Section 2 describes the related work. Section 3 describes the overall design. We present the system performance evaluation results in Section 4. Contrastive analysis is discussed in Section 5. Section 6 concludes this paper and presents some future work.

2. Related Work

In the traditional solution, a completely trusted neutral server is required. Such a central server is easily a target for attackers. Moreover, it is not suitable for deployment in scenarios where a large number of vehicles participate, and a large amount of data will bring high latency or even congestion. To deal with such problems, decentralized solutions have gradually become a research focus.

By using the tamper-proof, hash-encrypted features of blockchain technology, a decentralized and trusted identity can be defined as a set of keys used to prove the source and validity of the information. These keys are directly controlled by the principal with this identity. With the help of the consensus mechanism, this identity based on the blockchain can be effectively published and recorded.

2.1. Asymmetric Encryption Technology. In 1976, Diffie and Hellman proposed the concept of an asymmetric cryptosystem [24], that is, a public-key cryptosystem, which created a new direction in cryptography research. The asymmetric encryption algorithm is relative to the

symmetric algorithm. The difference between the two is reflected in whether the key can be disclosed. The symmetric key requires the same key to be used in the encryption and decryption process, while asymmetric encryption can provide a pair of keys. The private key is kept by yourself, and the public key can be made public. The common symmetric encryption algorithms are DES, 3DES, AES, and IDEA, and the common asymmetric encryption algorithms are RSA, D-H, and ECC [25].

The security of the blockchain is provided by cryptography. In many blockchain projects, asymmetric encryption algorithms are mainly used at the account level. In the symmetric encryption algorithm, because both parties need to share the key in advance, there are many inconveniences in the use process. The emergence of asymmetric algorithms solves this problem. Take Bitcoin [26] as an example. The Bitcoin address is converted from the public key, and the public key is converted from the private key. All user information is protected by a randomly generated private key.

Since the content of each block in the blockchain is open to the entire network, privacy protection is an important issue. The blockchain represented by Bitcoin uses the wallet address generated by the public key hash to externally represent the input and output process of the transaction which brings such a benefit: the public key is generated by a random private key. Only by relying on the public key hash, it is impossible to know who caused the transaction. The relationship between blockchain keys and addresses is shown in Figure 1. The owner of the private key is the only representative of the transaction generated by the corresponding address, but no one knows who the true private key holder is.

2.2. Consensus. The word consensus comes from Latin, meaning "agreement, accord," which in turn comes from consentire, meaning "feel together." Its meaning and usage relate to both a generally accepted opinion and the conclusion of a decision based on a collective agreement.

In an isolated system, the system always develops in disorder, and the same is true for information systems. However, this feature is not suitable for its use. To deal with this feature, the consensus mechanism has become an important choice. Consensus issues require multiple processes (or agents) to agree on single data [27]. Some processes (agents) may fail or be unreliable in other ways, so the consensus protocol must be fault-tolerant or flexible. The process must propose its candidate values in some way, communicate with each other, and reach a consensus on individual data.

The consensus problem is basic in controlling multiagent systems [28]. One way to reach consensus is to get a majority of all processes (agents) to agree. In this case, the majority requires at least more than half of the available votes (where each process is voted). However, one or more wrong processes may bias the final result, which may lead to failure to reach a consensus or a wrong consensus.

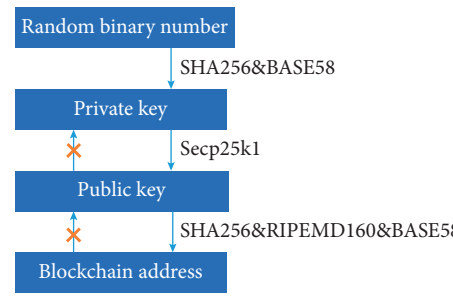


Figure 1: The relationship between blockchain keys and addresses.

2.2.1. Evolution of Consensus in the Internet. From the perspective of the development of the consensus mechanism, its development process is closely related to the process of network evolution. This section will sort out the development process of the network and consensus mechanism, clarify the relationship between the two issues, and find out the possible development direction of the future consensus mechanism. The Internet is a complex that has been evolving rapidly. From different perspectives such as history and technology, it can be divided into different stages. After synthesizing the scale and structure of the network, Table 1 divides the development of the Internet into five phases and reveals the development trajectory of the consensus mechanism.

The surge of intelligent Internet between 5G and artificial intelligence (AI) will be achieved in the 2020s, which is the most worthy of imagination. With the emergence of new business types and requirements such as the Internet of Vehicles, Internet of Things, Industrial Internet, 4K/8K, and AR/VR, the future network is appearing in a ubiquitous trend. It expects that the intelligent Internet will build an essential framework for a smart society in the future.

For the consensus mechanism, it is necessary to combine the specific requirements and application scenarios to implement adaptive settings for specific services [29, 30]. At this stage, increasingly or increasing various devices will become the main body of consensus. Decentralization makes the relationship between nodes equal. How to reach a fair and effective consensus is an inevitable problem. The following content of this article will clarify our consensus mechanism, which combines incentives and penalties to ensure its stable operation.

2.2.2. Blockchain Consensus Algorithm. In the decentralized network structure, there are no fixed nodes to manage transactions on the blockchain. Therefore, it is necessary to regularly allocate accounting rights from all mining nodes according to certain rules. At the same time, to achieve order from unnecessary to order and achieve entropy reduction, a certain amount of energy is inevitable. Since the blocks in the blockchain are generated at a certain time interval, in the interval phase, the transaction needs to be temporarily stored in the transaction pool. After the accounting node is selected in the network, the transaction is retrieved and

Table 1: The evolution of consensus in the internet.

Era of origin	Scale	Participants	Implementation process	Content
The late 1960s	Small scale	The network structure is simple, the differences between nodes are small, and there is no consensus requirement. Propose relevant problem assumptions		
In the mid-1980s	Medium scale	Distributed databases	Assuming there is no Byzantine fault node	Database agreement
In the mid-1990s	Large scale	Distributed databases	Tolerating Byzantine error nodes	Solving the problem of the efficiency of the Byzantine fault tolerance
The 2010s	Ultralarge scale	Nodes in the network	Tolerating Byzantine error nodes and focusing on performance and safety	Achieving open and transparent public accounts
The 2020s	Giga scale	People-machine-thing, full time and space, intelligent terminal	Intelligent, humanized interactive node	Configurable, fine-grained value delivery

packaged according to certain rules. After the packaging is completed, the results will be broadcasted and passed to other nodes for verification, and, finally, a consensus will be reached. The more classic blockchain consensus algorithms are shown in Table 2.

On this basis, many new algorithms based on the aforementioned consensus mechanism have emerged. Proof of activity [31] is a combination of PoW and PoS. It has been developed in the wake of an assumption based on an economic phenomenon called "Tragedy of the Commons." Practical Byzantine Fault Tolerance (PBFT) [32] is designated as more efficient than PoW in terms of latency and energy consumption, and it can only tolerate up to 33% of malicious nodes. Liu et al. [33] developed a novel consensus mechanism called Proof of Collaborative Work.

In the Internet of Vehicles, participants have a large number of devices with substantial differences, and they are easily affected by external environments such as regions and communication environments. Traditional consensus algorithms have insufficient transaction processing capabilities. There has also been some new consensus mechanism for the Internet of Vehicles [34, 35]. Nodes or devices with superior resources are in an advantageous position in the blockchain, which is prone to the phenomenon of the winner winning forever.

3. System Overview

In the ecosystem of the Internet of Vehicles, it is necessary to consider a large number of participants and the susceptibility to real factors such as geographic area and communication environment. The traffic system in the environment of the Internet of Vehicles is a complex system in which various parts are intertwined and interact with each other. It is necessary to pay attention to related safety issues such as the reliability and effectiveness of the system.

The factors that need to be considered for the safety of the Internet of Vehicles are shown in Figure 2.

3.1. System Model. In this section, we will demonstrate the system model.

As shown in Figure 3, the system consists of the following parts:

(1) Traffic management center (TC) is the highest authority on the Internet of Vehicles. It is connected to the roadside unit (RSU) and is mainly responsible for the registration of traffic participants and core information processing. The permanent identity and temporary identity of the vehicle are directly or indirectly related to the TC. It is considered completely credible and able to complete the work according to the design without being compromised by the adversary. As a complete node of the blockchain, it generates genesis blocks. It is responsible for mining and issuing and uploading various equipment official certificates to the blockchain ledger.

(2) Roadside units (RSUs) are distributed at intersections and on both sides of the road to generate vehicle access, identity verification, and other related matters. RSUs are considered as the edge computing node [36] that hosts the blockchain.

(3) Vehicle (V) is a general term for all types of vehicles driving on the road and represents the main participants of the network, namely, intelligent vehicles. Each vehicle can maintain its blockchain account and accompanying public and private key pairs and addresses on the blockchain. However, there are differences in computing power between different manufacturers and different types of in-vehicle devices.

(4) The operation of the blockchain relies mainly on four components: encryption algorithm, transaction processing, consensus algorithm, and distributed ledger technology. Using the encryption algorithm in the blockchain, a unique public-private key pair and wallet address are generated for each participant of the Internet of Vehicles.

Assume that all devices can protect their private keys from being obtained by others.

For vehicle information, we classify it as follows:

Table 2: Classic blockchain consensus.

Consensus mechanisms	Core concept	Pros	Cons
PoW	Computing power competition	It is completely decentralized to avoid the risks of concentration	The computing resources waste a lot, and the process of reaching a consensus takes a long time
PoS	Financial power competition	It shortens the time for a consensus to reach and avoids wasting power	Under the control of a few wealthy nodes, there is the possibility of unfairness
DPoS	Election and voting	It drastically reduces the time for a consensus to reach the second level	It is easy to be dominated by some nodes, and the voting representative may have doubts

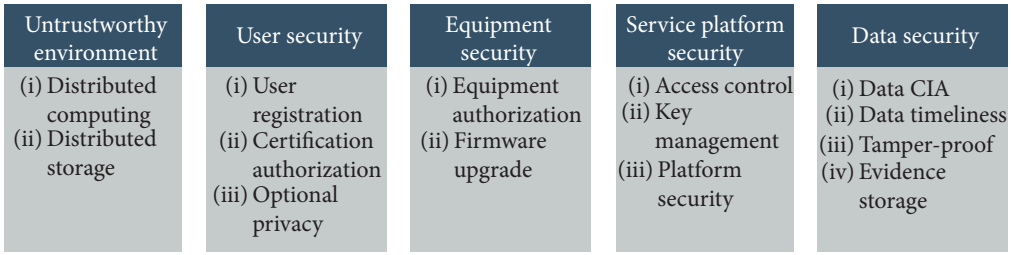


Figure 2: The factors for the safety of the Internet of Vehicles.

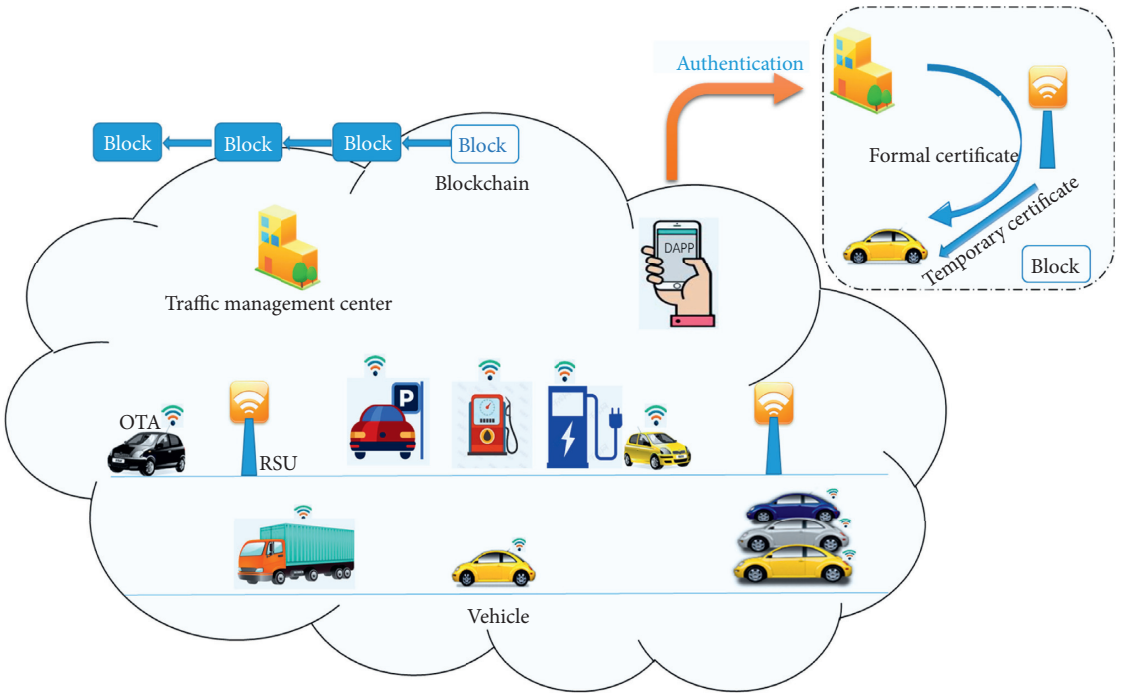


Figure 3: System model of IoV.

(i) General information

(ii) Sensitive information

Among them, general information includes license plates, models, brands, and colors, and sensitive information includes car owner information and vehicle identification number (VIN).

3.2. Proposed Overall Architecture. In this scheme, the main objects are devices on the Internet of Vehicles, including the aforementioned traffic management center, roadside units, and vehicles. Assume that every device can run the blockchain client and can protect its account.

In the initialization phase of the system, each device needs to run a blockchain client, complete the initialization

of the blockchain, and generate basic blockchain information (public key, private key, address, etc.).

The specific process is as follows.

3.2.1. Transportation Equipment Establishment Stage. After TC and RSU enter the Internet of Vehicles system supported by blockchain, the device will generate unique public and private key pairs and wallet addresses as part of its information. Then, Roadside unit RSU_i encrypts its attribute information (device serial number, location coordinates, etc.) with pk_{TC}, signs on it, and sends it to TC to apply for a device certificate.

When Tc receives an identity application request from RSU_i, it first verifies the digital signature of the application information. After verification, TC designed the authentication key for RSU_i $k_i = \{u_i | (u_i, v_i) = pk_{TC}\, sk_{RSUi} \bmod p\}$, encrypting ($k_i$, $address_i$) with the public key of RSU_i and writing it to the blockchain as the certificate of the RSU_i after TC signature.

It is worth noting that a timestamp is attached to the transaction after it is written to the blockchain, which ensures the traceability of the transaction. This can also be used to manage and maintain certificates.

3.2.2. Vehicle Equipment Establishment Stage. When a vehicle enters the Internet of Vehicles ecosystem for the first time, the registration needs to be completed for easy management. Therefore, it is necessary to apply to the Internet of Vehicles to obtain identification. Considering that the process of writing a transaction to the blockchain takes a nonnegligible time, in response to this, we design a type of temporary certificate to facilitate the timely connection of vehicles before the vehicle is officially certified.

The process of generating a temporary certificate is as follows:

(1) RSU_i broadcasts its signed identity certificate periodically.

(2) The vehicle j receives the broadcast information; it uses the public keys of TA and RSU_i to verify the legitimacy of the identity.

(3) If the signature is legal, it is determined that the RSU_i is credible, and the vehicle encrypts its attributes (attributes consist of two parts: general information authentication is encrypted with pk_{RSUi}, and sensitive information is encrypted with pk_{TC}).

(4) The vehicle signs the information and sends it to RSU_i. RSU_i verifies the signature, and it can decrypt the general information, perform preliminary authentication, and issue a lightweight temporary certificate ($address_{V_j}$ || timeliness || $addressRSU_i$ || RSU_i signature). The validity period of the certificate must be greater than the time it takes for things to be written on the blockchain.

(5) After receiving the temporary certificate, the vehicle verifies the signature and uses it as a temporary identity in the system to participate in the operation and maintenance of the system.

(6) RSU_i encrypts the general attributes of the received vehicle with the key k_i and attaches the encrypted sensitive information and a temporary certificate as a blockchain transaction. After completing the consensus agreement, it will be written into the blockchain.

As the complete nodes of the blockchain, RSU and TA are jointly maintained and can read and write data on the blockchain.

The process of generating a formal certificate is as follows.

(1) TA gets the transaction generated by the RSU_i in the blockchain.

(2) TA completes the decryption of the data and verifies the data signature. What we need to highlight here is

(3) $k_i = \{u_i | (u_i, v_i) = pk_{TC}\, sk_{RSUi} \bmod p\} = \{u_i | (u_i, v_i) = pk_{RSUi}\, sk_{TC} \bmod p\}$.

(4) This ensures that TC and RSU can calculate the same key without revealing their private key.

(5) Decrypt and verify the data. Then, TC issues a formal certificate ($address_{V_j}$ || timeliness || TA's signature) to the vehicle to complete the formal registration and certification. TC writes it into the blockchain as a transaction.

The notations in this paper are summarized in Table 3.

3.3. Fair Consensus Mechanism. We propose a consensus mechanism AoR, dedicated to the consortium blockchain. The commissioners maintain the blockchain.

3.3.1. Role Classification of Blockchain Nodes. There are three roles in the consensus mechanism:

Ordinary node (U): as the fundamental aspect of blockchain, they use cryptography to verify their identity and use the signature to verify their information sent. Ordinary nodes can join or leave the network without restriction. They are not permitted to participate in the block generation process directly but can "observe" the entire consensus process. They also can be involved in the process of block distribution and message forwarding and get a small number of online credit rewards.

Commissioners (C): different commissioners form the committee and maintain a consortium blockchain jointly. The commissioners exercise equal rights and obligations, reviewing bidders and organizing auctions, and verifying and forwarding blocks and transactions. A new block generated in the blockchain will be sent to all commissioners for verification signature. When a block receives at least 51% of the commissioners' approval, it will add to the blockchain as valid. The

Table 3: Summary of notations.

RSU_i	Roadside unit i
V_j	Vehicle j
p	Prime number in elliptic curve cryptography
k_i	The authentication key for RSU_i
pk_x	The public key of device x in the blockchain
sk_x	The private key of device x in the blockchain
$address_x$	The address of device x in the blockchain
N_u	The number of ordinary nodes
N_c	The number of commissioners
T_w	Each round of auction interval
T_a	The time of the vendee is determined
T_b	The time of new block be packaged
$T_{newblock}$	A new block's deadline
$b_{i,t}$	The credit of node i at time t

commissioners will be rewarded. The result of the vote will be the choices of all commissioners.

Bidder (B): in each round of auction, the ordinary node is required to submit a deposit within the specified time. The bidder competes for the right to generate the new block. Within the specified time, the bidder is ranked by the auction mechanism to determine the vendee (V). At the same time, the vendee's credit for the transaction price is used to reassign to the online nodes, and a block is generated in the blockchain. After completing the task, the vendee receives the corresponding reward.

3.3.2. Consensus Process. The AoR consensus assumes that the number of ordinary nodes is N_u, the number of commissioners appointed from ordinary nodes is N_c, and each auction interval is T_w. The vendee is determined within time T_a, and a new block needs to be generated within the time T_b, $T_a + T_b \ll T_w$. The valid block records the whole process of the auction transaction and gets $(N_c/2) + 1$ commissioners' signature at least. This process is known as a round of consensus. If no valid block is generated within T_b, it means that the original vendee has given up the right to generate the new block and the second highest bidder will generate the block as a winner and so on. As long as one bidder can work properly, the network can achieve a consensus finally.

Generating a new block requires the following steps:

S1. At the beginning of the chain, all nodes are assigned a certain amount of participation credit, and all nodes can trade each other and generate signed transaction data. At the same time, they verify the transaction data. If the transaction data are valid, they will forward the transaction data to the commissioners. The online task will be assigned a credit bonus after each round of the auction.

S2. All commissioners monitor the transaction data and store the legal data in the transaction pool.

S3. The vendee takes the valid transactions from the transaction pool and packs them into a block, sending the block to all commissioners. The block's deadline is

$$T_{newblock} = \text{PreviousBlockTime} + T_a + T_b. \quad (1)$$

S4. After receiving a raw block, the commissioner verifies the data in the block. If the commissioner approves this block, it shall sign for confirmation. After receiving at least $(N_c/2) + 1$ signatures, the vendee obtains the timestamp information of the NTP server. If the timestamp is earlier than $T_{newblock}$, the block will be signed by the vendee and published on the network. The committee accepts the vendee's credit and allocates it. If the timestamp is later than $T_{newblock}$, it means that the new block cannot be generated efficiently. The vendee will be replaced by the auction mechanism, and the specified task should be completed within the new deadline $T_{newblock} = T_{newblock} + T_b$. The former vendee who failed to complete the task will be liable. This mechanism prevents nodes from poor performance and malicious motives.

S5. After receiving the valid block, the vendee deletes the illegal transactions from the transaction pool. Moreover, all nodes wait for the time T_w to start the next round of the auction.

In particular, if $T_{newblock} > T_w$, it means that all bidders cannot complete the generation of the new block in time, or the network has truncated. In this article, we assume that the above situation does not transpire. The various relationships of the characters are given in Figure 4.

4. Experiments and Performance Analysis

In the scheme mentioned in this article, the system security is based on the difficult problem of the elliptic curve discrete logarithm problem (ECDLP). Each device node participating in the Internet of Vehicles can generate its private key, public key, and address according to the protocol and can use TA's public key to generate an encryption key.

4.1. Safety Analysis

(i) The adversary cannot get the private key of the IoV device:

Any node, including the adversary, can collect the public keys and system parameters of all nodes in the scheme, which is also a prerequisite guarantee for the use of blockchain. As mentioned earlier, we assume that any node can protect its private key from being acquired by an adversary. The adversary cannot calculate the encryption key. If the adversary calculates the private key of the node through the public parameters and public key, it violates the difficult problem of ECDLP. Therefore, the probability of the adversary compromising in our scheme is negligible.

(ii) The anonymity of the vehicle:

Whether it is a temporary certificate or a formal certificate, it only contains the blockchain address of the vehicle and does not contain any information

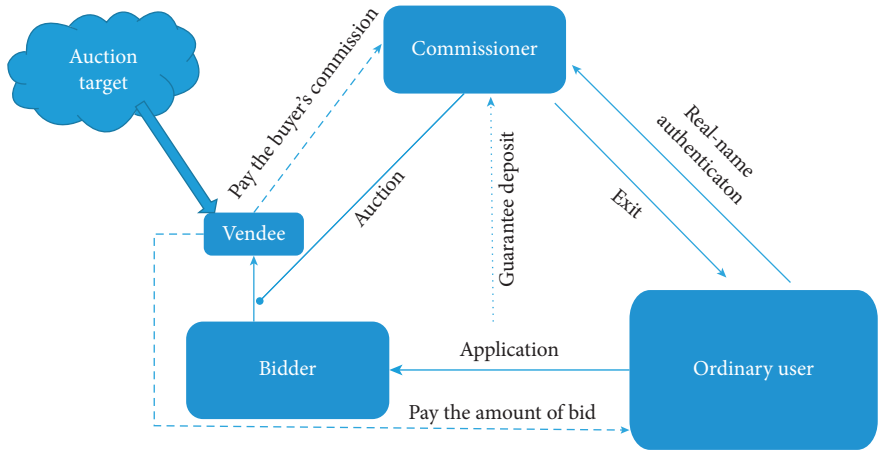


Figure 4: Interaction of various roles in the network.

related to the identity of the vehicle. The private information of the vehicle is encrypted with the public key of the TC. As long as the private key of the TC is not exposed, the private information of the vehicle can only be obtained by the TC.

(iii) Sybil attack protection:

In a network, the attacker simulates the existence of multiple entities (devices), that is, a single node has multiple identities. These entities send erroneous information to the server or management application weakening the role of redundant backup. In our design, each object can only have one formal certificate at a given time, and each identity can only have one key pair. Every communication message must be signed by the private key associated with this identity. Moreover, all formal identities must be approved by the traffic management center, so attackers cannot use forged identities.

(iv) DoS/DDoS protection:

Denial of service (DoS) or distributed DoS (DDoS) attacks are characterized by an explicit attempt by the attacker to prevent legitimate use of the service [37]. Since DoS/DDoS does not need to identify and utilize protocol or service flaws, they are highly efficient for any type of service and are therefore the most dangerous network attack. The decentralized architecture of the blockchain makes it powerful against DoS/DDoS attacks. The data on the blockchain is redundant and distributed on different nodes. Even if an attacker manages to stop one node, it cannot stop all other nodes. In addition, transactions in the blockchain require a certain cost, which prevents an attacker from sending a large number of transactions to attack.

4.2. Consensus Fairness Analysis. Based on the auction mechanism and the consortium blockchain, this paper proposes a new consensus model. The consensus mechanism needs a trade-off between performance, fairness, and security. Our model will guarantee fairness on the premise of blockchain security. In the following, we will analyze the fairness and effectiveness of AoR.

Suppose that N nodes share B credits $N > 3$, and Nc represents the number of committees considered; $t = 0, 1, 2 \ldots, T$ is the index of the period. At time t, the credit of node i is represented by $b_{i,t}$. Assume that the total credits are distributed equally among the participating nodes in the initial stage. That is, for any node, $b_{i0} = B/N$.

During the operation of the blockchain, the credit updates of different nodes are as follows:

$$\text{Ordinary nodes: } b_{i,t+1} = b_{i,t} + (1/N - 1 - Nc)\alpha x_{\mathrm{v,t}}$$
$$\text{Vendee: } b_{v,t+1} = b_{v,t} - x_{\mathrm{v,t}} \quad \text{s.t.} \quad x_{\mathrm{v,t}} < b_{v,t}$$
$$\text{Commissioners: } b_{c,t+1} = \alpha b_{c,t} + (1/Nc)\beta x_{\mathrm{v,t}}$$

$x_{\mathrm{v,t}}$ represents the amount of credit used by Vendee v at a time t, α is the proportion of the auction price, and β is the proportion of Commissioner's commissions and satisfies $\alpha + \beta = 1$.

Proposition 1. *At any given moment t, the sum of credits assigned to each node is certain.*

Proof. At $t = 0$, the sum of node credits is initialized to $\sum_0^N b_{i0} = B$. Considering the previous assumptions, no node in the system carries credits away, and no new credits are generated in the life cycle. Therefore, the sum of credits of each node remains constant, B.

This method can prevent some nodes from occupying the generation right for a long time. Because spending credit $x_{\mathrm{v,t}}$ means that other nodes will have a higher credit budget at a time $t + 1$. This movement of credit guarantees fairness among nodes.

Proposition 2. *At any time t, for any two nodes i and j,* $|\sum_0^t x_{i\tau} - \sum_0^t x_{j\tau}| < (N - 2/N - 3)B$. *N is the total of nodes, and* $x_{i\tau}$ *is the cost of node i at time* τ.

Proof. Consider the extreme case. At time $t = 0$, all credits belong to node i, $b_{i0} = B$ (actually $b_{i0} < B$). Other nodes have zero credit.

After the node i completes the first auction, the x_i transaction value is at most B, and node j gets the credit $b_j = (B/N)$. To minimize x_j, it assumes that node j never bid. Node i can collect up to $b_i = (N - 2/(N-1)^2)B$ before the next bid.

After the second round, x_i receives up to $(N - 2/(N-1)^2)B$.

And so,

$$\begin{aligned}\sum_0^t x_{i\tau} &< B\left(1 + \frac{N-2}{(N-1)^2} + \frac{(N-2)^2}{(N-1)^4} + \frac{(N-2)^3}{(N-1)^6} + \cdots \frac{(N-2)^t}{(N-1)^{2t}}\right)\\ &< B\left(1 + \frac{N-2}{(N-2)^2} + \frac{(N-2)^2}{(N-2)^4} + \frac{(N-2)^3}{(N-2)^6} + \cdots \frac{(N-2)^t}{(N-2)^{2t}}\right)\\ &< B\left(\frac{1-\left(1/(N-2)^t\right)}{1-(1/N-2)}\right)\\ &< \frac{N-2}{N-3}B.\end{aligned} \tag{2}$$

Thus, $|\sum_0^t x_{i\tau} - \sum_0^t x_{j\tau}| \le |\sum_0^t x_{i\tau}| < (N - 2/N - 3)B$.

The time-averaged expenditure credit difference between any two nodes is

$$\lim_{t\longrightarrow\infty} \frac{1}{t}\left|\sum_0^t x_{i\tau} - \sum_0^t x_{j\tau}\right| < \lim_{t\longrightarrow\infty} \frac{1}{t}\left|\frac{N-2}{N-3}B\right| = 0. \tag{3}$$

This means that the sum of credits for all nodes' expenditures tends to be the same. No node can spend more credits and take the dominant position. The aforementioned ensures fairness between nodes.

4.3. Performance Analysis

4.3.1. Encryption and Signature Performance. We conducted experiments on a 1.9 GHz processor, Intel Core i5 with 12 GB RAM, and Windows 10 running in python3.7 to study operating costs. Figure 5 depicts the time-consuming signature and verification, encryption, and decryption. The average time consumption of signature and verification is 9.6 ms and 0.34 ms. The average time consumption of encryption and decryption is 182.6 ms and 62.8 ms. It can be found that signature and encryption take more time than verification and decryption. However, the above operations are tolerable in terms of time consumption.

4.3.2. Consensus Algorithm Analysis. In this part, the simchain (https://github.com/YaoyaoBae/simchain) is used to evaluate the time consumption. As described in Section 3, we simulated 200, 500, and 1000 nodes participating in the generation of blockchain 40 times and compared the consensus mechanism we designed with PoW under the same conditions.

Figure 6 depicts the time it takes for nodes to reach a consensus. Figures 6(a)–6(c) depict the consensus system time consumption for 200, 500, and 1000 node scales, respectively. In PoW, it takes a long time to solve the hash puzzle, so we assume that 20%–60% of the nodes randomly participate in each round. However, in the AoR simulation, only a brief bid is submitted in each round of consensus, and we assume that all nodes participate. For each scale, we perform 40 experiments. In all three cases, AoR takes much less time than PoW. Figure 6(d) describes the consequences of AoR in networks with various node quantities. The collection and validation of transactions and the selection of the bid order by the ledger manager consume some time costs. Besides, these tasks are influenced by the size of the network.

5. Comparison with Related Work

The advancement of the Internet of Things technology has promoted the development of the Internet of Vehicles with autonomous vehicles and roadside infrastructure as the main components. IoV aims to provide innovative services for different traffic equipment through adaptive traffic management and to improve traffic safety and efficiency. In this section, we compare the functions of our scheme with the existing schemes.

A paper [4] proposes an efficient and practical pseudonymous authentication protocol with conditional privacy preservation. It expects an honest-but-curious behavior from otherwise fully trusted authorities. The proposed protocol protects a user's privacy until the user honestly follows the protocol. In case of malicious activity, the true identity of the user is revealed to the appropriate authorities.

The authors proposed a new identity-based (ID) signature based on the elliptic curve cryptosystem (ECC) and employed it to propose a new conditional privacy-preserving authentication scheme based on the ID-based signature they invented [38]. The scheme provides a secure authentication process for the information transmitted between the vehicle and the RSU.

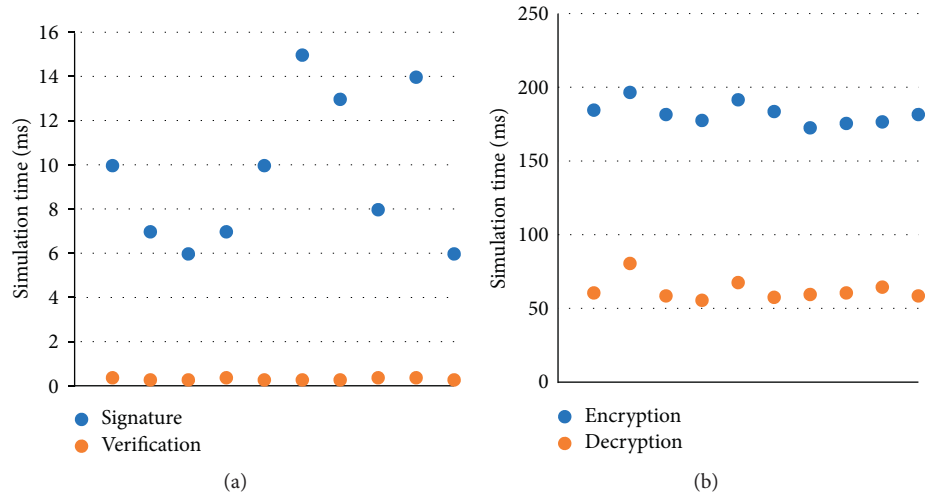


Figure 5: Time consumption of the basic operations. (a) Signature and verification. (b) Encryption and decryption.

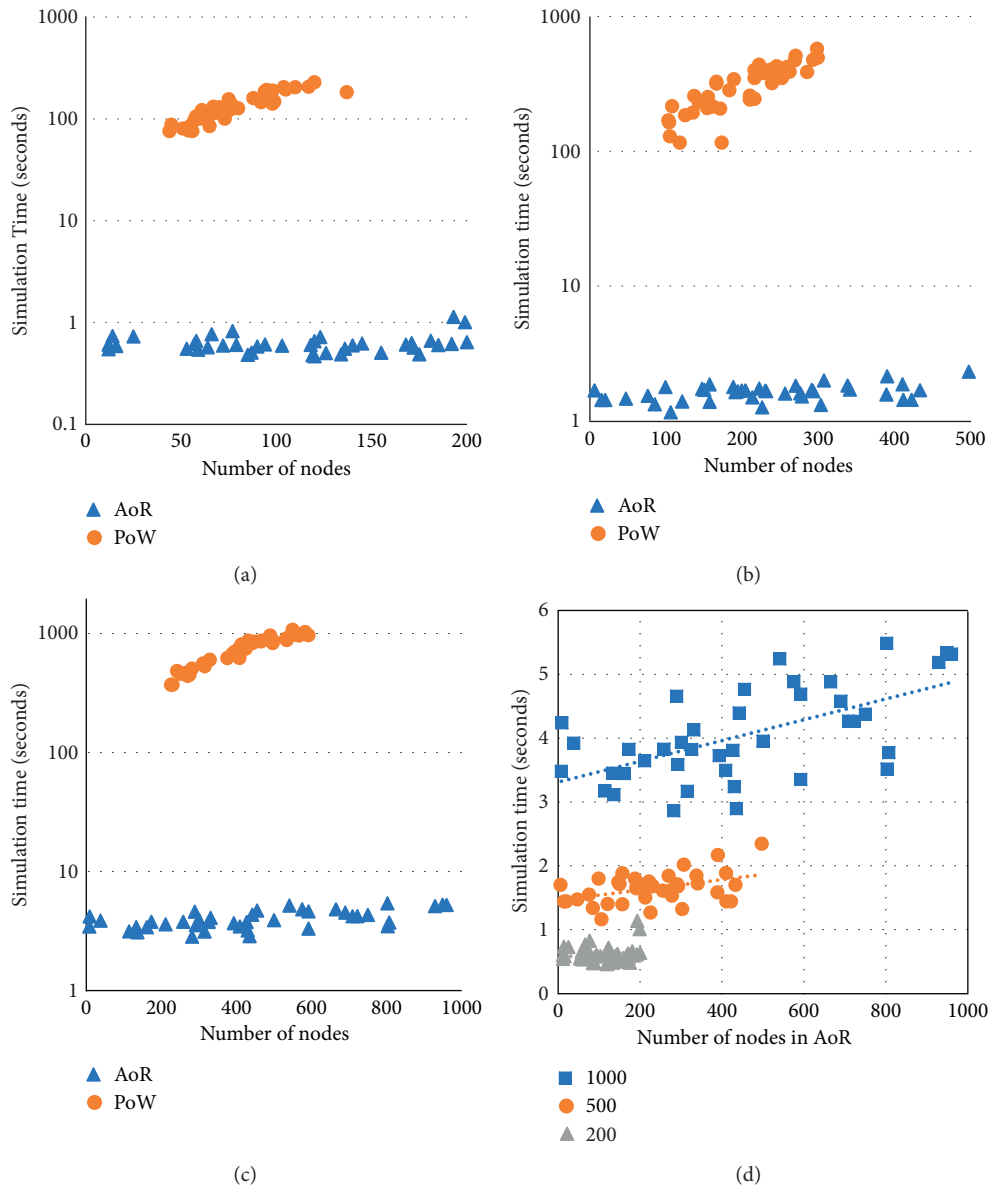


Figure 6: Time consumption of AoR and PoW.

Table 4: Comparisons of the related work.

Scheme	Anonymity	Authentication	Traceability	Decentralization	Certificate permanency
Lo's scheme [38]	Yes	Yes	Yes	No	Yes
Rajput's scheme [4]	Yes	Yes	Yes	No	Yes
Azees's scheme [3]	Yes	Yes	Yes	No	Yes
Wang et al. [7]	Yes	Yes	Yes	Yes	Yes
Pu's scheme [13]	Yes	Yes	Yes	Yes	Yes
Our scheme	Yes	Yes	Yes	Yes	Optional

In [3], Azees et al. propose an efficient anonymous authentication scheme to avoid malicious vehicles entering the VANET. Besides, the proposed scheme offers a conditional tracking mechanism to trace the vehicles or roadside units that abuse the VANET. As a result, the scheme revokes the privacy of misbehaving vehicles to provide conditional privacy computationally efficiently, through which the VANET entities will be anonymous to each other until they are revoked from the VANET system.

Javaid et al. [12] proposed a blockchain-based protocol for IoV using smart contracts, physical unclonable functions (PUFs), certificates, and a dynamic Proof-of-Work (dPoW) consensus algorithm. The blockchain with smart contracts provides a secure framework for registering trusted vehicles and blocking malicious ones. PUFs are used to assign a unique identity to each vehicle via which trust is established. Certificates are issued by roadside units that preserve the privacy of vehicles, whereas the dPoW consensus allows the protocol to scale according to the incoming traffic generated by the vehicles.

The authors introduced an efficient, reliable, and privacy-preserving scheme based on blockchain for VSNs [13]. In their scheme, a pseudonym mechanism is employed to achieve individual anonymization by concealing the vehicles' identity. To encourage vehicles to report trustworthy information, incentive punishment mechanism is proposed. Meanwhile, they propose a multifactor and single-factor weight-based evaluation mechanism to evaluate the reliability of the message. Practical Byzantine Fault Tolerance (PBFT) and blockchain are also employed to achieve consensus and store records, respectively, which can prevent malicious entities from manipulating vehicles' reward scores and credit scores.

To summarize, Table 4 describes the studied works. From the table, we can find that our scheme is used in a variety of scenarios and more comprehensive and suitable for IoV.

6. Conclusions

The latest development of the Internet of Things has promoted the evolution of the Internet of Vehicles. IoV aims to provide new and innovative services for different modes of transportation through adaptive traffic management to improve traffic safety and efficiency. However, due to the actual untrusted environment, establishing trust in IoV is a critical security issue. Therefore, we proposed an authentication with optional privacy preservation. In our scheme, we mainly consider two situations. One situation is that the vehicle temporarily enters the system, and the scheme is designed for temporary authentication. Another situation is that the vehicle has been participating in the Internet of Vehicles for a long time, and the program has been designed for formal certification. In the authentication process, the scheme inherits the decentralized, credible, and tamper-proof characteristics of the blockchain. At the same time, the certificate uses the cryptographic primitives in the blockchain, without using a new one. Furthermore, we design a new consensus mechanism that provides a fair chance for nodes in the blockchain to obtain bookkeeping rights. Security analysis and experimental results show that our scheme is efficient and secure for IoV devices.

In future work, we consider using aggregation technology [39] and coordination control [40] to process data to facilitate the use and transmission of data. For a wider range of Internet of Vehicles applications, we will consider upgrading the chain structure [41, 42] of the blockchain to reduce the resource consumption and transaction processing time of the blockchain.

Data Availability

No additional data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was funded by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX18_0903.

References

[1] F. Yang, S. Wang, J. Li, Z. Liu, and Q. Sun, "An overview of internet of vehicles," *China Communications*, vol. 11, no. 10, pp. 1–15, 2014.

[2] WHO Organization, *Global Status Report on Road Safety 2018: Summary*, World Health Organization, Geneva, Switzerland, 2018.

[3] M. Azees, P. Vijayakumar, and L. J. Deboarh, "EAAP: efficient anonymous authentication with conditional privacy-preserving scheme for vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 9, pp. 2467–2476, 2017.

[4] U. Rajput, F. Abbas, and H. Oh, "A hierarchical privacy preserving pseudonymous authentication protocol for VANET," *IEEE Access*, vol. 4, pp. 7770–7784, 2016.

[5] P. Vijayakumar, M. Azees, A. Kannan, and L. J. Deborah, "Dual authentication and key management techniques for secure data transmission in vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1015–1028, 2015.

[6] M. Li, L. Zhu, and X. Lin, "Efficient and privacy-preserving carpooling using blockchain-assisted vehicular fog computing," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4573–4584, 2018.

[7] K. Wang, J. Yu, X. Liu, and S. Guo, "A pre-authentication approach to proxy re-encryption in big data context," *IEEE Transactions on Big Data*, 2017.

[8] M. Du, K. Wang, Y. Chen, X. Wang, and Y. Sun, "Big data privacy preserving in multi-access edge computing for heterogeneous internet of things," *IEEE Communications Magazine*, vol. 56, no. 8, pp. 62–67, 2018.

[9] S.-C. Cha, J.-F. Chen, C. Su, and K.-H. Yeh, "A blockchain connected gateway for BLE-based devices in the internet of things," *IEEE Access*, vol. 6, pp. 24639–24649, 2018.

[10] R. Sharma and S. Chakraborty, "Blockapp: using blockchain for authentication and privacy preservation in IoV," in *Proceedings of the 2018 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, IEEE, Abu Dhabi, UAE, December 2018.

[11] C. Xu, H. Liu, P. Li, and P. Wang, "A remote attestation security model based on privacy-preserving blockchain for V2X," *IEEE Access*, vol. 6, pp. 67809–67818, 2018.

[12] U. Javaid, M. N. Aman, and B. Sikdar, "A scalable protocol for driving trust management in internet of vehicles with blockchain," *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11815–11829, 2020.

[13] Y. Pu, T. Xiang, C. Hu, A. Alrawais, and H. Yan, "An efficient blockchain-based privacy preserving scheme for vehicular social networks," *Information Sciences*, vol. 540, pp. 308–324, 2020.

[14] M. Swan, *Blockchain: Blueprint for a New Economy*, O'Reilly Media, Inc., Sebastopol, CA, USA, 2015.

[15] M. Wu, K. Wang, X. Cai, S. Guo, M. Guo, and C. Rong, "A comprehensive survey of blockchain: from theory to IoT applications and beyond," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8114–8154, 2019.

[16] J. Sengupta, S. Ruj, and S. Das Bit, "A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT," *Journal of Network and Computer Applications*, vol. 149, Article ID 102481, 2020.

[17] M. A. Ferrag, M. Derdour, M. Mukherjee, A. Derhab, L. Maglaras, and H. Janicke, "Blockchain technologies for the internet of things: research issues and challenges," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2188–2204, 2018.

[18] S. Chen, L. Yang, C. Zhao, V. Varadarajan, and K. Wang, "Double-blockchain assisted secure and anonymous data aggregation for fog-enabled smart grid," *Engineering*, 2020.

[19] N. Kshetri, "Can blockchain strengthen the internet of things?" *IT Professional*, vol. 19, no. 4, pp. 68–72, 2017.

[20] P. Treleaven, R. Gendal Brown, and D. Yang, "Blockchain technology in finance," *Computer*, vol. 50, no. 9, pp. 14–17, 2017.

[21] J. Zhang and M. Wu, "Blockchain use in IoT for privacy-preserving anti-pandemic home quarantine," *Electronics*, vol. 9, no. 10, p. 1746, 2020.

[22] H. Li, K. Wang, T. Miyazaki, C. Xu, S. Guo, and Y. Sun, "Trust-enhanced content delivery in blockchain-based information-centric networking," *IEEE Network*, vol. 33, no. 5, pp. 183–189, 2019.

[23] Y. Liu, K. Wang, Y. Lin, and W. Xu, "LightChain: a lightweight blockchain system for industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3571–3581, 2019.

[24] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, 1976.

[25] R. Yegireddi and R. K. Kumar, "A survey on conventional encryption algorithms of cryptography," in *Proceedings of the 2016 International Conference on ICT in Business Industry & Government (ICTBIG)*, pp. 1–4, IEEE, Indore, India, November 2016.

[26] S. Nakamoto, Bitcoin: a peer-to-peer electronic cash system, 2008.

[27] X. Wang, H. Su, X. Wang, and G. Chen, "Fully distributed event-triggered semiglobal consensus of multi-agent systems with input saturation," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5055–5064, 2017.

[28] X. Wang, G.-P. Jiang, H. Su, and Z. Zeng, "Consensus-based distributed reduced-order observer design for LTI systems," *IEEE Transactions on Cybernetics*, 2020.

[29] H. Li, K. Wang, X. Liu, Y. Sun, and S. Guo, "A selective privacy-preserving approach for multimedia data," *IEEE Multimedia*, vol. 24, no. 4, pp. 14–25, 2017.

[30] V. Ortega, F. Bouchmal, and J. F. Monserrat, "Trusted 5G vehicular networks: blockchains and content-centric networking," *IEEE Vehicular Technology Magazine*, vol. 13, no. 2, pp. 121–127, 2018.

[31] I. Bentov, C. Lee, A. Mizrahi, and M. Rosenfeld, "Proof of activity," *ACM SIGMETRICS Performance Evaluation Review*, vol. 42, no. 3, pp. 34–37, 2014.

[32] M. Castro and B. Liskov, "Practical Byzantine fault tolerance and proactive recovery," *ACM Transactions on Computer Systems*, vol. 20, no. 4, pp. 398–461, 2002.

[33] Y. Liu, K. Wang, K. Qian, M. Du, and S. Guo, "Tornado: enabling blockchain in heterogeneous internet of things through a space-structured approach," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1273–1286, 2019.

[34] Y.-T. Yang, L.-D. Chou, C.-W. Tseng, F.-H. Tseng, and C.-C. Liu, "Blockchain-based traffic event validation and trust verification for VANETs," *IEEE Access*, vol. 7, pp. 30868–30877, 2019.

[35] R. Shrestha, R. Bajracharya, and S. Y. Nam, "Blockchain-based message dissemination in VANET," in *Proceedings of the 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*, pp. 161–166, IEEE, Kathmandu, Nepal, October 2018.

[36] C. Xu, K. Wang, P. Li et al., "Making big data open in edges: a resource-efficient blockchain-based approach," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 4, pp. 870–882, 2018.

[37] J. Mirkovic and P. Reiher, "A taxonomy of DDoS attack and DDoS defense mechanisms," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 2, pp. 39–53, 2004.

[38] N.-W. Lo and J.-L. Tsai, "An efficient conditional privacy-preserving authentication scheme for vehicular sensor networks without pairings," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1319–1328, 2015.

[39] J. Yu, K. Wang, D. Zeng, C. Zhu, and S. Guo, "Privacy-preserving data aggregation computing in cyber-physical social systems," *ACM Transactions on Cyber-Physical Systems*, vol. 3, no. 1, pp. 1–23, 2018.

[40] X. Wang, X. Wang, H. Su, and J. Lam, "Coordination control for uncertain networked systems using interval observers,"

IEEE Transactions on Cybernetics, vol. 50, no. 9, pp. 4008–4019, 2020.

[41] S. Popov, "The tangle," *White Paper*, vol. 1, p. 3, 2018.

[42] M. Du, K. Wang, Y. Liu et al., "Spacechain: a three-dimensional blockchain architecture for IoT security," *IEEE Wireless Communications*, vol. 27, no. 3, pp. 38–45, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9978791, 9 pages
https://doi.org/10.1155/2021/9978791

Research Article

Application of Characteristic Model-Based Principal Component Analysis in Optimization of Flowmeter Parameters

Wenping Jiang,[1] **Zhencun Jiang**,[1] **Lingyang Wang**,[1] **Jun Min**,[2] **Yi Zhu**,[3] **Zhonghui Wang**,[3] **Lizhuang Tang**,[3] **and Limin Sun**[3]

[1]*Shanghai Institute of Technology, School of Electrical and Electronic Engineering, No. 100 Haiquan Road, Fengxian District, Shanghai 201418, China*
[2]*Tongji University, No. 1239 Siping Road, Yangpu District, Shanghai 200082, China*
[3]*Bitobar Company, No. 265 Lingdong Street, Tieling, Liaoning 112000, China*

Correspondence should be addressed to Wenping Jiang; jiangwenping@sit.edu.cn

Received 31 March 2021; Revised 8 June 2021; Accepted 4 August 2021; Published 20 August 2021

Academic Editor: Zhiwei Gao

In complex industrial processes, it is necessary to perform modeling analysis on some industrial systems and find and optimize the factors that have the greatest impact on the results, in order to achieve the optimization of the industrial systems. However, due to the high-level nature or complex working mechanism of complex industrial systems, traditional principal component analysis methods are difficult to apply. Therefore, this paper proposes a characteristic model-based principal component analysis (CMPCA) to perform principal component analysis on complex industrial systems. The differential pressure flowmeter is taken as an example to verify the effectiveness of the method. Flowmeter is an indispensable instrument in measurement, and its accuracy depends on its own structural parameters. However, the measurement accuracy of some flow meters is not high, and the measurement error is large, which affects the normal industrial production process. This method is used to analyze the influence of the structural parameters of the flowmeter on its measurement accuracy, and the four most important structural parameters are found and optimized. The measurement error of the Bitoba flowmeter is reduced from 1% to 0.2%, and the measurement repeatability is reduced from 0.3 to 0.06, which proves the effectiveness of the method.

1. Introduction

In the industrial production process, with the increase of testing means, the production line can provide a large number of production data for the application of engineers. Meanwhile, with the upgrade of the CPU, data-driven modeling and control become possible.

For complex industrial systems [1], traditional dynamics modeling is difficult to apply due to its high order or complex working mechanism, and even if it is applied, mathematical methods are needed for downscaling modeling: For the simplification of single-input single-output models of large power systems, a lot of people study model simplification method based on SVD-Krylov subspace projection methods, which aims to rely on singular value decomposition and Krylov subspace methods for order reduction system. The intelligent method is mainly applied in neural networks, fuzzy logic, genetic algorithm, etc. This method can solve the partial open-loop control problem without the need for an accurate dynamic model. It mainly constructs the model form by experience and then obtains the model through parameter identification. This method is limited in industrial production because it cannot guarantee the effect of online control. Yichuan Fu and Zhiwei Gao et al. proposed classifying the actuator and sensor faults of wind power equipment based on fast Fourier transform (FFT) and uncorrelated multilinear principal component analysis (UMPCA), which were tested and verified on wind turbines [2]. Due to the complex working mechanism of intermittence and continuity in the industrial system, the variables are often nonlinearly correlated. The traditional data-based PCA has a large calculation error in this case, because of the

linear relationship between the assumed variables, so it is not suitable for dynamic nonlinear working mechanisms.

The characteristic modeling method was proposed by Wu Hongxin Academician of the Chinese Academy of Sciences in the 1980s [3], which broke through the original framework of modeling controlled objects and provided a new idea for the modeling of high-order complex systems with unknown parameters and orders. It has laid a theoretical foundation for the realization of low-order controller design for some high-order systems. After more than 30 years of research and accumulation, important progress has been made in both theory and application. At present, the problem of characteristic modeling for linear time-invariant systems and affine nonlinear systems has been basically solved [4]. This method has been successfully applied to hundreds of systems in aerospace and industrial control fields and has important practical application prospects.

Flow measurement is an important guarantee for industrial automation. As the requirements for measurement accuracy in the industrial production process gradually increase, people's demand for flow measurement is also increasing. A lot of investment has been used to improve the automation level of the plant, which is more reflected in the field data collection and real-time monitoring. In the energy industry or in the chemical-pharmaceutical industry, high-precision flow meters are required [5–7]. Taking Bitoba Flowmeter as an example, it is a differential pressure flowmeter; this kind of flowmeter has a good application prospect in petroleum, metallurgy, water supply, and other fields, especially in the large-diameter industry. However, there are still two problems. (1) The flowmeter cannot fully adapt to the real situation of the factory due to the calibration of the flow coefficient is completely determined by the factory test. (2) The measurement accuracy needs to be further improved and the measurement reproducibility is expected to be reduced to an ideal level [8–12].

To address these two problems, a characteristic model-based principal component analysis method is proposed in this paper, and the structural parameters of the flowmeter are optimized using this method. A three-dimensional simulation model was performed on the flowmeter, and the optimized structural parameters were tested in the simulation environment and verified in actual industrial production. The second part introduces the principle of the characteristic model, the principal component analysis method based on the characteristic model, and the flowmeter simulation design method. The third part is the result of the flowmeter simulation modeling and the results of the principal component analysis method based on the characteristic model on flowmeter parameter analysis. The fourth part is the conclusion.

2. Materials and Method

2.1. Characteristic Model-Based Principal Component Analysis. For the process control systems that widely existed in industrial processes and cannot be accurately modeled, the characteristic model method can identify the nonlinear relations existing in the system through data identification and decouple the coupling relations between various variables. For the same structure, no matter how the initial value is selected, the system parameters will converge to a relatively fixed value. This is the difference between the characteristic model method and intelligent algorithms such as neural networks. Therefore, the characteristic model method can accurately determine the system's principal components, thus providing a practical and effective method for the modeling and control of the system.

The so-called characteristic model-based principal component analysis is the model built according to the dynamic characteristics, environmental characteristics, and control performance requirements of the object, and it finds the principal components of the object, not only using the precise dynamic equations of the object to model. The characteristic model is different from the higher-order model in order reduction, but the relevant information of the higher-order model is compressed into several characteristic parameters without losing. In the dynamic process, the output of the characteristic model is equal to that of the real object. In the stable state, the outputs are more equal, and the range of coefficients of the characteristic model can be determined in advance.

The main features of the characteristic model-based principal component analysis include the following aspects:

(1) The characteristic model compresses the main information of the higher-order model into several characteristic parameters. In general, the characteristic model can be described by the slow time-varying difference equation

(2) The obtained characteristic model expression must have lower order than the original dynamic model, which is easy to realize in the industrial field

(3) The order and form of the characteristic model mainly depend on the control performance requirements, in addition to the need to consider the object characteristics

(4) In the stable state, the output of the object characteristic model and the actual object are equal. If the system is given the same input signals, they are also equivalent in terms of output; that is, they can be kept within the allowable output error in an unstable state

For systems whose physical characteristics are clear but difficult to describe with dynamic equations, we can directly establish the characteristic model expression between the control variables represented by physical characteristics and their physical characteristics output and extract the principal components.

The transfer function of the high-order linear time-invariant system commonly seen in engineering is as follows:

$$G(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \cdots + b_1 s + b_0}{s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0}. \quad (1)$$

It can be simplified into the following form:

$$G(s) = \frac{k_v}{s^2} + \sum_{i=1}^{m} \frac{k_i}{s+\lambda_i} + \sum_{i=1}^{m} \left(\frac{k_{p+i}}{s+\lambda_{p+i}} + \frac{\overline{k}_{p+i}}{s+\overline{\lambda}_{p+i}} \right) + \sum_{i=1}^{p} \frac{k_{\omega_i}}{(s+\omega_i)^2}. \tag{2}$$

Equation (2) can also be described by the higher-order time-varying difference equation, as shown in equation (3). However, if the coefficients of the difference equation change in real time, then the higher-order time-varying difference equation can be rewritten into a time-varying difference equation of order 2. The order of the difference equation is determined according to the actual situation.

For any n-order linear time-invariant system $G(s)$ that can be decomposed according to equation (2) to achieve position retention or position tracking control, under the condition that a certain sampling period is satisfied, it can be expressed in the form of the second-order time-varying difference equation as follows:

$$y(k+1) = \alpha 1(k)y(k) + \alpha 2(k)y(k-1) + \beta 0(k)u(k) + \beta 1(k)u(k-1). \tag{3}$$

When the object $G(s)$ is stable or contains integral links, then one has the following:

(1) The coefficients of equation (3) are slow-varying

(2) The range of coefficients can be determined in advance

(3) Assuming that the sampling period can ensure that the error of the output is within the allowed range, the output of the characteristic model can be used to replace the output of the actual system under the condition that the input is the same, and the two are equal; in the stable state, the static gain of equation (3) is equal to the static gain of equation (2)

(4) When the static gain $D = 1$, the system tends to be stable and the sum of the coefficients of the difference equation is equal to 1; then

$$\alpha 1(k) + \alpha 2(k) + \beta 0(k) + \beta 1(k) = 1. \tag{4}$$

(5) The object has integral link time:

$$\alpha 1(k) + \alpha 2(k) = 1. \tag{5}$$

For the establishment of first-order and second-order characteristic models for a continuous system, its mathematical description can be expressed by the following equation:

$$\dot{y} = f(y, u). \tag{6}$$

In sample equation (6), the sampling period is Δt. The general second-order characteristic model is as follows:

$$y(k+1) = f_1(k)y(k) + f_2(k)y(k-1) + g_0(k)u(k). \tag{7}$$

The first-order Taylor expansion can also be used to obtain

$$y(k+1) = f_1(k)y(k) + g_0(k)u(k). \tag{8}$$

For the general dynamic equations, the unknown system can also be directly constructed according to the characteristic mechanism, as shown in equations (7) and (8).

2.2. Flowmeter Structure. The Bitobar flowmeter's physical diagram is shown in Figure 1.

Its composition principle is shown in Figure 2.

The most important factor which affects the measurement accuracy is the part of the pressure head. The three-dimensional diagram of the pressure head is shown in Figures 3(a)–3(c).

The structural parameters of the indenter are mainly as follows:

(1) The angle of full pressure surface

(2) Static pressure plane angle

(3) The distance from the centerline to the full pressure hole

(4) The distance from centerline to static pressure hole

(5) The height of the static hole

We build a mathematical model between the data and the test model based on the data provided by the manufacturer. The mathematical model includes the following parameters: standard volume, standard mass, flow rate, pressure, Reynolds number, test current, flow coefficient, error, and repeatability.

2.3. Working Principle of Flowmeter. The calculation formula of the volume flow of fluid is shown in

$$qv = \sqrt{\frac{2\Delta P}{\rho}}, \tag{9}$$

where qv is the volume flow rate and its unit is m^3/s. ΔP is the pressure difference between total pressure and static pressure. The density of a medium fluid is ρ and kg/m^3 is its unit. Let $A = \pi D^2/4$; equation (9) can be derived as equation (10). The cross-sectional area of the pipe's unit is m^2 and is denoted as A. The unit of the inner diameter of the pipe is mm, and it is expressed as D. a is a dimensionless flow coefficient.

$$\alpha = 7.90484 \frac{qv}{D^2} \sqrt{\frac{\rho}{\Delta P}}. \tag{10}$$

In industry, the flow coefficient a must be measured by the flowmeter using the flow standard device, so when the flowmeter becomes the main body of this study, the flow coefficient α is a known quantity. The volume flow in the pipeline can be calculated by calculating the measured pressure difference, density, and cross-sectional area as equation (9). The equation for calculating the repeatability of verification point I is shown in

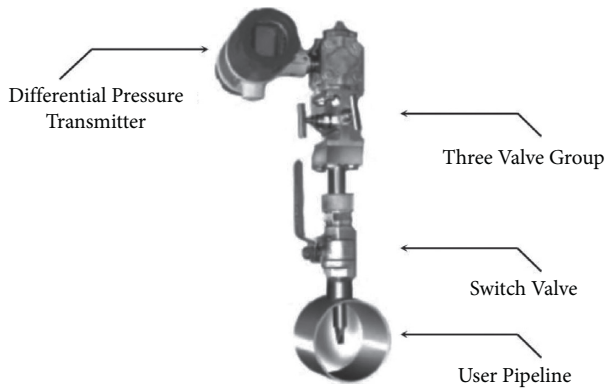


Figure 1: Flowmeter physical map.

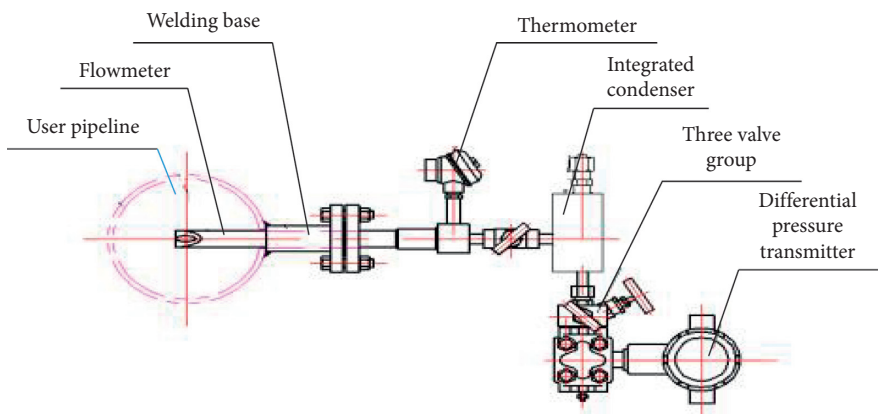


Figure 2: Flowmeter composition schematic.

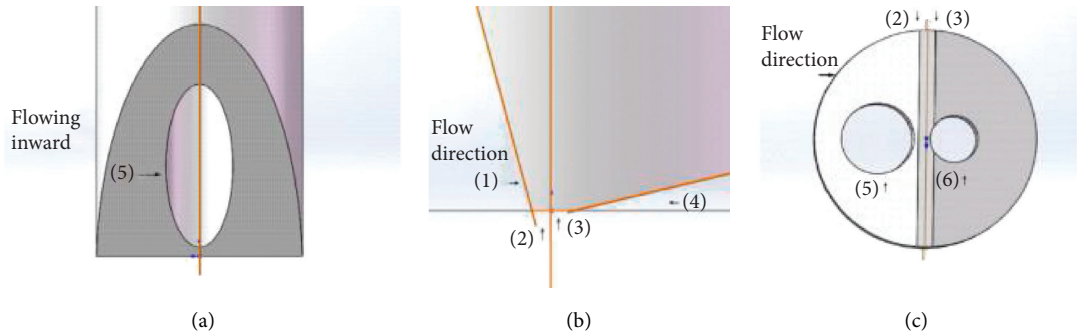


Figure 3: Three views of the flowmeter head section. (a) The flowmeter pressure head's front view. (b) The flowmeter pressure head's end view. (c) The flowmeter pressure head's vertical view.

$$(E_r)_i = \frac{1}{\alpha i}\left[\frac{1}{(n-1)}\sum_{i=n}^{n}\left(\alpha_{ij}-\alpha_i\right)^2\right]^{1/2} \times 100\%. \quad (11)$$

The equation used to calculate the flow coefficient of the flow meter is shown in

$$E = \frac{(\alpha i)\max - (\alpha i)\min}{(\alpha i)\max + (\alpha i)\min} \times 100\%. \quad (12)$$

The calculation equation of the linear error of the flow meter is shown in

$$\alpha = \frac{(\alpha i)\max + (\alpha i)\min}{2}. \quad (13)$$

Flow coefficient α_i reflects the characteristics of the device. Its size and linear characteristics are related to the shape, size, industrial conditions, and pipe diameter of the equipment. For the Bitobar differential pressure flowmeter, the flow coefficient must be determined first. After the determination of the flow coefficient, the real flow value can be obtained and the indication error can be determined.

2.4. Flowmeter Simulation. Working conditions have a great influence on the accuracy of the flowmeter, so are the structural parameters. Now, the flowmeter's accuracy will change with different pipe diameters, liquids, or pipe filling degrees. It is necessary to design the structural parameters previously and simulate a test for the flowmeter [13–18].

(1) SolidWorks: SolidWorks, the most wildly used drawing software in engineering, can build a flowmeter's three-dimensional model. What's more, the software has powerful functions, simple operations, and a rich conversion interface. We need to use it for modeling

(2) HyperMesh: HyperMesh can be used for grid dividing. Nowadays, there are many kinds of software for grid diving, such as ICEM and ANSA. Compared with HyperMesh, ICEM and ANSA both have disadvantages. Firstly, ICEM cannot generate high-quality prism meshes in the boundary layer. Secondly, the degree of freedom of ANSA in generating meshes is not as high as HyperMesh. HyperMesh can manipulate the grid precisely. And it can delete and merge nodes, so HyperMesh is selected to divide the grid

(3) Gambit: Gambit is used to process the model's boundaries. After being processed by HyperMesh, the model was cut into many small squares. Using gambit software, the boundary of these grids can be assigned, and the parameters inputted should be consistent with the actual data. Then, we define the flow rate, pressure, and other parameters to prepare for the following fluid calculations

(4) Fluent: At present, Fluent is the most widely used software in fluid calculation. It supports discontinuous mesh, hybrid mesh, dynamic deformable mesh, and sliding mesh. It has powerful performance and can meet the needs of different conditions and can perform complex fluid calculations

In this paper, the κ-ε RNG model is used instead of the standard κ-ε model [19–21]. RNG theory provides an analytical formula to consider viscous flow at a low Reynolds number. In a wide flow range, the κ-ε model is more reliable than the standard κ-ε model and more accurate than the κ-ε model. The κ-ε RNG model is suitable for various flow types, especially in the simulation of circular jet and plane jets, which can make the fluid flow simulation more reasonable. The κ-ε RNG model is used to simulate the outflow coefficient. The average error of numerical simulation is about 5%, and the maximum value is less than 8%.

At the same time, the SIMPLE algorithm [22–25] is used and it takes the pressure as the basic variable. There are five steps in the process of the SIMPLE algorithm, shown in Figure 4.

3. Experimental Results and Analysis

3.1. Flowmeter Model Simulation Results. Firstly, the simulation model is validated by the factory test data. Secondly, more than 70 groups of data can be obtained when simulating each feature number from 10 to 15 times. Thirdly, the influence of each characteristic parameter of the flowmeter is analyzed when the preliminary database is established. Thus, the optimal model is determined. Finally, the simulation block diagram is shown in Figure 5.

Let the pressure be equal to 185000 Pa and the mass velocity is 84.0277 kg/s. The κ-ε RNG model is selected, iterated 50 times until convergence. The final simulation result is shown in Figure 6. The value of the simulated pressure difference ΔP is 2.568104, while the real data given by the manufacturer is 2.5423104. Therefore, the error between the algorithm adopted in this experiment and the actual results is shown in equation (14) with the first set of parameters.

$$E = \frac{2.568 - 2.5423}{2.5423} = 1.1089\%. \tag{14}$$

The result shows that the simulation error is 1.1%. Similarly, multiple sets of real data on-site are used to verify the accuracy of the simulation method, and the maximum error is 8%. The smaller the diameter of the static pressure hole is, the smaller the measurement error will be. However, the static pressure hole will be blocked by impurities which need to be cleaned up frequently when the static pressure hole is too small. Therefore, different diameters are formulated to ensure the accuracy and continuity of production according to different media. For different velocities, the position of the average velocity is 0.24 times the radius of the pipe wall, diameter, and roughness in the process of measurement.

3.2. Characteristic Model-Based Principal Component Analysis Application Results. Through the analysis of the flowmeter structure, five structural parameters that affect the measurement accuracy of the flowmeter are selected, and characteristic model-based principal component analysis is used to further principal component analysis of these five parameters.

Structural fitting data is based on the following characteristic model.

$$\begin{aligned} y(k+1) = {} & f_1(k)u_1(k) + f_2(k)u_2(k) \\ & + f_3(k)u_3(k) + f_4(k)u_4(k) \\ & + f_5(k)u_5(k), \end{aligned} \tag{15}$$

where $f_i(k)$ is the characteristic model coefficient, $u_i(k)$ is the structural parameters, and $y(k+1)$ is the system output ΔP. The identification method we selected is the gradient method. The main ideas of the gradient method are as follows. Firstly, start at one point and the direction where the function drops fastest is taken as the search direction. Then, the direction where the function value at any point drops fastest is its negative gradient direction. Finally, the N-dimensional problem is transformed into a one-dimensional problem, and the one-dimensional search method is used to find the optimization along the negative gradient direction. The gradient method can be used to solve the nonlinear equations with real coefficients, such as a group of roots belonging to $f_i(x_1, x_2, \ldots x_n) = 0$, $i = 1, 2, \ldots, n$. The gradient method needs to define an objective function firstly:

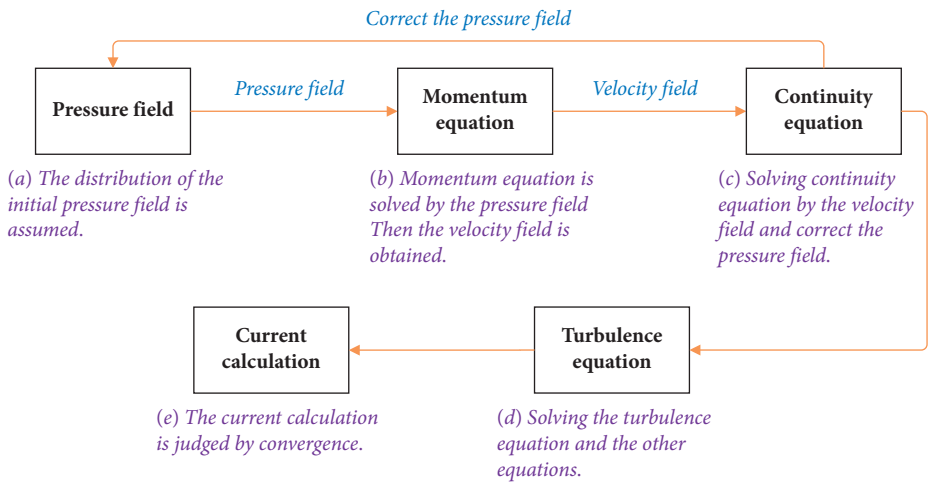


Figure 4: The process of the SIMPLE algorithm.

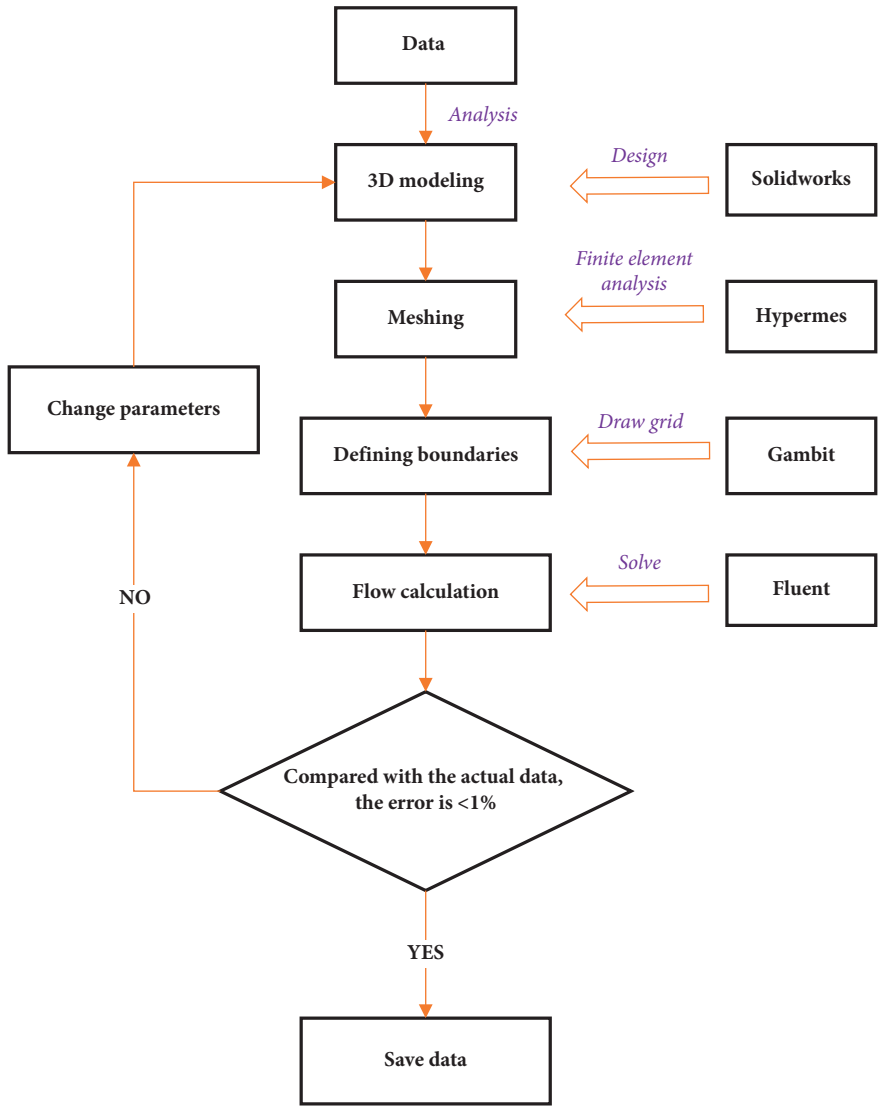


Figure 5: Simulation block diagram.

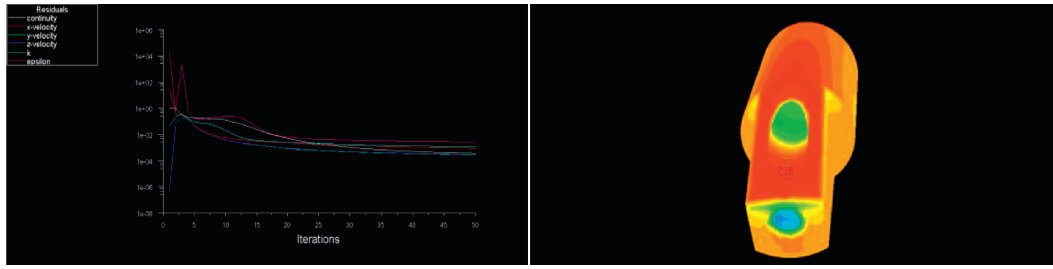


Figure 6: Simulation results.

$$\Phi(x_1, x_2, \Lambda, x_n) = \sum_{i=1}^{n} f_1^2(x_1, x_2, \Lambda, x_n). \tag{16}$$

Secondly, we look for the minimum value of the objective function $\phi = \sum_{i=1}^{n} f_1^2$ among $x_1, x_2, \ldots x_n$. If we get a set of solutions: ${x_1}^k, {x_2}^k, \ldots {x_n}^k$ at the k step ($k = 0, 1, 2, \ldots$), then making $\Phi = ({x_1}^k, {x_2}^k, \ldots {x_n}^k) < \varepsilon$, where ${x_1}^k, {x_2}^k, \ldots {x_n}^k$ is a set of efficient solutions of the original nonlinear equations under certain precision conditions. The iterative steps of the gradient method are as follows.

Step 1. Give an initial value $x_1^0, x_2^0, \ldots x_n^0$ (the initial value cannot be 0 completely) and set up the roots obtained in step k which are ${x_1}^k, {x_2}^k, \ldots {x_n}^k$.

Step 2. Calculate the value of the objective function $\Phi = (x_1^k, x_2^k, \ldots x_n^k)$.

Step 3. If $\Phi = (x_1^k, x_2^k, \ldots x_n^k)$, we can think that $x_1^k, x_2^k, \ldots x_n^k$ is a set of solutions that meet a certain precision. Otherwise, $x_i^{k+1} = x_i^k - a^k (\partial\phi/\partial x_i)|_{x_i = x_i^k}$.

$$\left.\begin{aligned}
a^k &= \frac{\Phi\left(x_1^k, x_2^k, \Lambda, x_n^k\right)}{\sum_{i=1}^{n} \left(\partial\Phi/\partial x_j\right)^2 |_{x_j = x_j^2}}, \\
\frac{\partial\Phi}{\partial x_1} &= \frac{\Phi\left(x_1^k + h^1, x_2^k, \Lambda, x_n^k\right) - \Phi\left(x_1^k, x_2^k, \Lambda, x_n^k\right)}{h_1}, \\
\frac{\partial\Phi}{\partial x_2} &= \frac{\Phi\left(x_1^k, x_2^k + h_2, \Lambda, x_n^k\right) - \Phi\left(x_1^k, x_2^k, \Lambda, x_n^k\right)}{h_2}, \\
&\Lambda\Lambda \\
\frac{\partial\Phi}{\partial x_n} &= \frac{\Phi\left(x_1^k, x_2^k, \Lambda, x_n^k + h_n\right) - \Phi\left(x_1^k, x_2^k, \Lambda, x_n^k\right)}{h_n}, \\
h_i &= H * x_i^k, \quad i = 1, 2, \Lambda, n.
\end{aligned}\right\} \tag{17}$$

Step 4. Correct x_i^{k+1} repeatedly until

$$\left(\Phi_1^{k+1}, \Phi_2^{k+2}, \Lambda, \Phi_n^{k+1}\right) < \varepsilon. \tag{18}$$

The block diagram is shown in Figure 7.

After performing principal component analysis on the structural parameters of the flowmeter through characteristic model-based principal component analysis, the order of the influencing factors of differential pressure ΔP of the flowmeter is obtained, and the parameter convergence curve is shown in Figure 8.

The specific results are as follows:

(1) The height of Static hole, B

(2) The distance from the centerline to the full pressure hole, C

(3) The angle of full pressure surface, A

(4) The distance from centerline to static pressure hole, H

3.3. Industrial Process Practical Application Verification. According to the characteristic model-based principal component analysis, the results of principal component analysis for the structural parameters of the flowmeter are tested in the flowmeter model built by the simulation software and verified in specific industrial applications.

We use the characteristic model-based principal component analysis method to find the four main influencing factors of the flowmeter pressure difference ΔP and optimize these four parameters to achieve the purpose of improving the measurement accuracy of the flowmeter and reducing the measurement repeatability. We verify the effectiveness of the algorithm.

(1) Increase the height of the static hole from 1.5 to 6

(2) Increase the distance from the centerline to the full pressure hole from 1.5 to 1.875

(3) Reduce the angle of full pressure surface from 70° to 56°

(4) Increase the distance from the centerline to the static pressure hole from 1.5 to 1.875

The results of the actual industrial production process show that after the optimization of the flowmeter structure parameters, the measurement error of the flowmeter is reduced to 0.2%, and the measurement repeatability is reduced to 0.06. The specific actual data is shown in the table. In addition, it was verified in subsequent production that the optimized flowmeter structure parameters added 2 million yuan in economic benefits to the manufacturer. It proves the

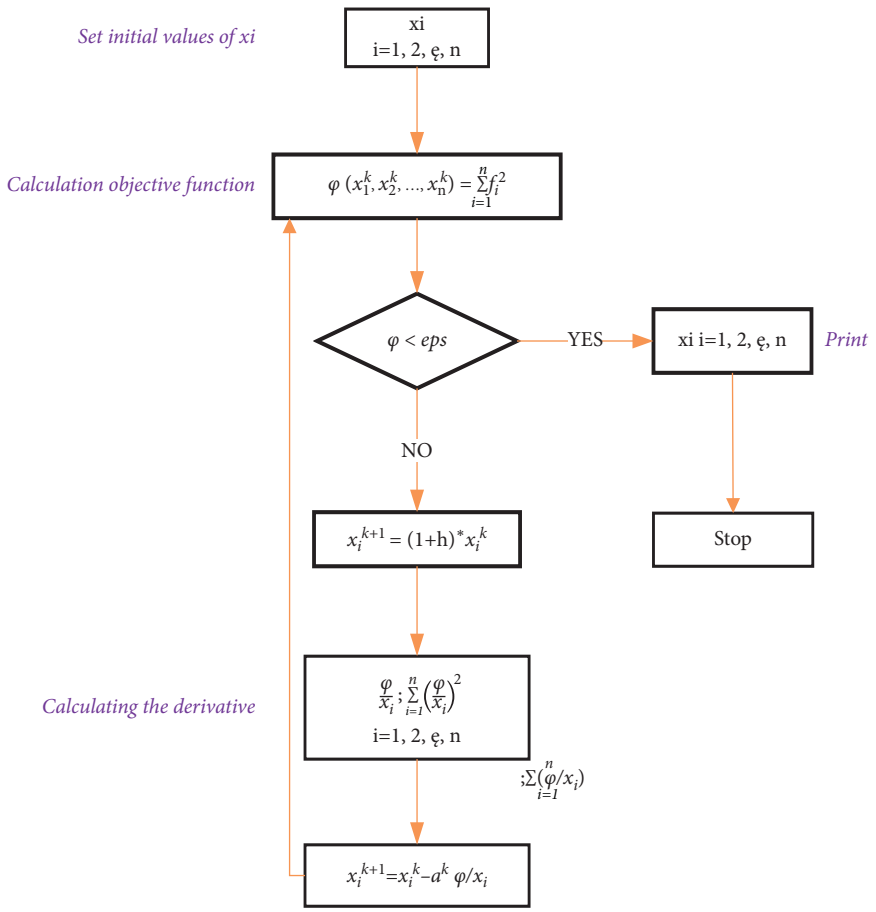


Figure 7: Block diagram of gradient method.

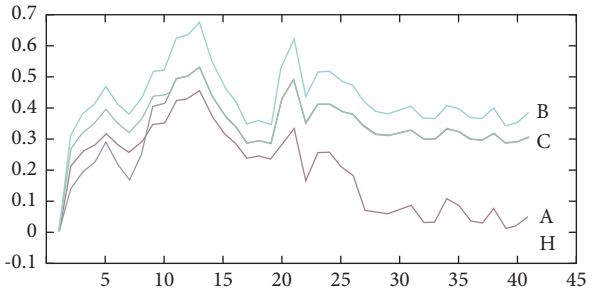


Figure 8: The parameter convergence curve.

effectiveness of characteristic model-based principal component analysis for the optimization of flowmeter parameters.

4. Conclusion

(1) This paper proposes a characteristic model-based principal component analysis method, introduces the characteristic model theory, and applies it to the optimization of flowmeter structure parameters.

(2) Take Bitoba flowmeter as an example, analyze its structural characteristics, use simulation software to simulate Bitoba flowmeter, and build a simulation model of Bitoba flowmeter.

(3) Use characteristic model-based principal component analysis to perform principal component analysis on the five structural parameters that affect the measurement accuracy of the flowmeter, and obtain the four structural parameters that have the greatest impact on them. After optimization of these four structural parameters, the Bitoba flowmeter's measurement error is reduced from 1% to 0.2%, and the measurement repeatability is reduced from 0.3 to 0.06, which increases economic benefits in industrial production.

It is proved that the characteristic model-based principal component analysis can well identify the nonlinear relationship existing in the system, and for the same structure, the system parameters will converge to a relatively fixed value, thus clearly distinguishing the principal component and its effect on the system and the size of the weight. For process control systems that cannot be accurately modeled widely in industrial processes, this method can decouple the coupling relationship between various variables through data identification and accurately find the principal elements of the system, providing an effective method for modeling such systems.

Data Availability

Data are unavailable.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[2] K. Fu, "Learning control systems and intelligent control systems: an intersection of artificial intelligent and automatic control," *IEEE Transactions on Automatic Control*, pp. 70–72, 1971.

[3] Y. Fu, Z. Gao, Y. Liu, A. Zhang, and X. Yin, "Actuator and sensor fault classification for wind turbine systems based on fast fourier transform and uncorrelated multi-linear principal component analysis techniques," *Processes*, vol. 8, no. 9, p. 1066, 2020.

[4] Z. Liu, G. Du, and L. Liu, "Study on internal flow field characteristics and measurement performance of momentum gas flowmeter," *Acta Metrologica Sinica*, vol. 40, no. 1, pp. 124–129, 2019.

[5] H. Yang, Y. Tu, and Y. Mao, "The ap-Hilbert method for phase difference estimation of Coriolis flowmeter," *Chinese Journal of Scientific Instrument*, vol. 40, no. 1, pp. 35–42, 2019.

[6] D. Park and J. S. Go, "Design of cyclone separator critical diameter model based on machine learning and CFD," *Processes*, vol. 8, p. 1521, 2020.

[7] W. Liu, K. Xu, J. Le et al., "Determination of two key parameters in the amplitude control of Coriolis flowmeter," *Journal of Electronic Measurement and Instrument*, vol. 32, no. 10, pp. 183–189, 2018.

[8] L. Sun, S. Dang, T. Zhang, and Y. Liu, "Experiment of flowmeter performance affected by flow stability," *Journal of Tianjin University*, vol. 51, no. 4, pp. 422–432, 2018.

[9] L. Wu, W. Yin, and C. Wang, "Development and experimental research of liquid fertilizer torque target flowmeter," *Journal of Transduction Technology*, vol. 30, no. 2, pp. 224–229, 2017.

[10] H. M. Abd, O. R. Alomar and I. A. Mohamed, Effects of varying orifice diameter and Reynolds number on discharge coefficient and wall pressure," *Flow Measurement and Instrumentation*, 2019.

[11] Z. Wang, E. Andiroglu, G. Wang, and L. Song, "Accuracy improvement of virtual pump water flow meters using calibrated characteristics curves at various frequencies," *Energy and Buildings*, vol. 191, pp. 143–150, 2019.

[12] B. Oblak, S. Babnik, V. Erklavec-Zajec, B. Likozar, and A. Pohar, "Digital twinning process for stirred tank reactors/separation unit operations through tandem experimental/computational fluid dynamics (CFD) simulations," *Processes*, vol. 8, no. 11, p. 1511, 2020.

[13] J. Zhang, Di Wang, B. Xu, Su Qi, Z. Lu, and W. Wang, "Flow control of a proportional directional valve without the flow meter," *Flow Measurement and Instrumentation*, vol. 67, 2019.

[14] G. Wang, Z. Wang, and Li Song, "Uncertainty analysis for different virtual pump water flow meters," *Science and Technology for the Built Environment*, vol. 25, no. 3, 2019.

[15] D. V. Gradov, A. Saren, J. Kauppi, K. Ullakko, and T. Koiranen, "Auto-aspirated daf sparger study on flow hydrodynamics, bubble generation and aeration efficiency," *Processes*, vol. 8, p. 1498, 2020.

[16] S. Nasiruddin, S. N. Singh, S. V. Veeravalli, and S. Hegde, "Effect of vertex angle and vertex tip radius on the performance of V-cone flow meter using CFD," *Measurement*, vol. 138, pp. 536–544, 2019.

[17] G. Cortellessa and L. Iacomini, "A novel calibration system for heat flow meters: Experimental and numerical analysis," *Measurement*, vol. 144, pp. 105–117, 2019.

[18] S. Z. Islami rad and R. Gholipour Peyvandi, "A simple and inexpensive design for volume fraction prediction in three-phase flow meter: single source-single detector," *Flow Measurement and Instrumentation*, vol. 69, Article ID 101587, 2019.

[19] M. Guilizzoni, G. Salvi, G. Sotgia, and L. P. M. Colombo, "Numerical simulation of oil-water two-phase flow in a horizontal duct with a Venturi flow meter[J]," *Journal of Physics: Conference Series*, vol. 1224, no. 1, 2019.

[20] G. Chiesa, C. Bertani, N. Falcone, A. Bersano, M. D. E. Salve, and B. Panella, "Horizontal air-water two-phase flow measurement using an electrical impedance probe and a venturi flow meter," *Journal of Physics: Conference Series*, vol. 1224, no. 1, 2019.

[21] Y. Liu, F. Yi, C. Zhou, and L. Tao, "Study on the influence of reflecting channel position on flow velocity characteristics of ultrasonic heat meter," *IOP Conference Series: Materials Science and Engineering*, vol. 544, no. 1, 2019.

[22] D. Duda, V. Yanovych, and V. Uruba, "An experimental study of turbulent mixing in channel flow past a grid," *Processes*, vol. 8, p. 1355, 2020.

[23] M. H. Roudi, J. Malekzadeh, M. Ebrahimi, A. Mirhaghi, and M. Shakeri, "Comparison between Emergency Severity Index plus peak flow meter and Emergency Severity Index in the dyspneic patients with chronic obstructive pulmonary disease: a randomized clinical trial," *Turkish journal of emergency medicine*, vol. 19, no. 2, pp. 68–72, 2019.

[24] M. Fujita, K. Nagashima, S. Takahashi, K. Suzuki, T. Fujisawa, and A. Hata, "Handheld flow meter improves COPD detectability regardless of using a conventional questionnaire: A split-sample validation study," *Respirology (Carlton, Vic.)*, vol. 25, no. 2, pp. 191–197, 2019.

[25] E. J. Jang, B. Nandram, Y. Ko, and D. H. Kim, "Small area estimation of receiver operating characteristic curves for ordinal data under stochastic ordering," *Statistics in Medicine*, vol. 39, no. 10, pp. 1514–1528, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6256785, 13 pages
https://doi.org/10.1155/2021/6256785

Research Article

Parameters' Identification of Vessel Based on Ant Colony Optimization Algorithm

Chen Zhao and Xiaojian Li

School of Information Science and Engineering, Northeastern University, Shenyang, China

Correspondence should be addressed to Chen Zhao; 20181492@stu.neu.edu.cn

Received 13 May 2021; Accepted 9 July 2021; Published 26 July 2021

Academic Editor: Xiao Ling Wang

In this paper, the ant colony optimization (ACO) method is used to identify the parameters of a 3-DOF nonlinear vessel model. Identifying the parameters is abstracted as a nonlinear optimization problem to solve through the ant colony optimization algorithm. The identification procedure is divided into two parts. The first part of the identification procedure is to identify the parameters related to surge motion. The second part of the identification procedure is to identify the rest parameters of the vessel's kinetics model. In the surge model identification procedure, the transient motor speed is used to generate the training data, and in the sway and yaw motion identification procedure, the zigzag maneuvering with different motor speeds is used to generate the training data. All the parameters are identified by the ACO method and the least-square (LS) method based on the training data and then validated on the validation data. The prediction performance of parameters identified by different methods is compared in the simulation to demonstrate the effectiveness of the ACO algorithm.

1. Introduction

Unmanned technology has developed rapidly in recent years and attracted more and more attention from academia and industry. Many kinds of unmanned products such as unmanned vehicles, unmanned aerial vehicles, unmanned surface vessel, and unmanned submersibles have been widely deployed as a network in various situations such as scientific research, environmental missions, ocean resource exploration, military use, and other applications [1] to help people improve work efficiency. As an intelligent device that works on the water, the unmanned surface vessel can work as a node of the overall unmanned network and extend the working range of the entire unmanned network to the surface of the water and underwater. The level of intelligence of the unmanned network is related to the level of intelligence of every node in the network. A sufficiently intelligent unmanned surface vessel contains a lot of techniques such as navigation techniques [2, 3], guidance techniques [4, 5], and control techniques [6–8].

The modeling of the unmanned surface vessel is imperative for both control method design and simulation study purposes [1]. To describe the vessel motion in the surge, sway, and heave, a nonlinear 6-DOF model with different dimension motions coupled together is imperative. However, the nonlinear 6-DOF model has lots of parameters which are very difficult to identify, the vessel algorithms designed based on the nonlinear model are very difficult, and the real-time performance of the algorithms is hard to guarantee. For algorithm design, some simple models such as the first-order Nomoto model proposed by Nomoto in 1957 [9, 10] are used. The Nomoto model describes vessel motion approximately under the assumption that the forward speed of the vessel changes slowly. Except for the Nomoto model, the 3-DOF model, including surge velocity, sway velocity, and yaw angular velocity, is also widely used in vessel control [11–13].

The prediction results of the Nomoto model and 3-DOF model cannot meet the requirements of the simulation study. How to identify the parameters of the nonlinear 6-

DOF model is a challenging problem. The towing carriage can identify the parameters of the vessel model such as CyberShip II [14], but the size of the towing carriage limits the size of the identified vessel. Some research use CFD [15] to simulate the ship motion and to identify the parameters [16], but the accuracy of the CFD in a complex situation is still worthy of further research. Many research studies choose full-scale vessel trials to identify the vessel model; the approach to model identification involved adapting model parameter value which was proposed in [17]. In [18], a new transformed multi-innovation least-squares (TMILS) algorithm is developed; the model structure is identified first and then the parameters are identified with full-scale trial data. An artificial neural network was used in [19] to model a high-speed craft with sea trial data.

As for the identify method, the least-squares method [20–23] has been a widely used method in system identification procedure, but the least-squares method can only deal with the linear problem; some identified procedure are abstracted as a nonlinear optimization problem. Support vector machines have the advantage in solving the nonlinear optimization problem, and it was used to identify the parameters of the vessel in [24–27]. Except for the least-square method and support vector machines method, some other methods are also used to identify the parameters. In [28], the identification procedure was divided into two steps: the first step was to determine the structure of the nonlinear model and the second step was parameter estimation refinement by using a nonlinear prediction error method with the unscented Kalmen filter. In [29], a sensitivity analysis and SQP method were used to identify the hydrodynamic coefficients of the Esso Bernicia tanker.

In this paper, the ant colony optimization method is used to identify the parameters of the vessel kinetics model. The parameter-identified problem is summarized as a nonlinear least-square problem, and the ant colony optimization method is used to solve the nonlinear least-square problem. The solution to the problem is the parameters to be identified. The identification procedure is divided into two parts [23]: the parameters related to surge motion were identified first by an experiment and then the rest of the parameters related to sway and yaw motion were identified by the zigzag test.

The organization of this paper is as follows. In Section 2, a nonlinear discrete 3-DOF state-space model of USV with surge speed, lateral speed, and yaw angle as state variables is established. Section 3 discusses the ant colony optimization method. Section 4 illustrates and analyzes the identification results. Section 5 summarizes the conclusions.

2. Problem Formulation

2.1. 3-DOF Nonlinear Model. In this section, the 3-DOF nonlinear symmetric model is established. By neglecting the heave, roll, and pitch motion, we consider the maneuvering models proposed by [17, 30]

$$\mathbf{M}\dot{\nu} + \mathbf{C}(\mathbf{v})\mathbf{v} + \mathbf{D}(\mathbf{v})\mathbf{v} = \boldsymbol{\tau} + \boldsymbol{\tau}_{\text{wind}} + \boldsymbol{\tau}_{\text{wave}}. \tag{1}$$

(i) Vector $\nu = [u \;\; v \;\; r]^{\top}$ denotes vessel surge velocity, sway velocity, and yaw angular velocity, respectively.

(ii) Matrix $\mathbf{M}$ accounts for inertial effects:

$$\mathbf{M} = \begin{bmatrix} m - X_{\dot{u}} & 0 & 0 \\ 0 & m - Y_{\dot{v}} & mX_g - Y_{\dot{r}} \\ 0 & mX_g - Y_{\dot{r}} & I_z - N_{\dot{r}} \end{bmatrix}. \tag{2}$$

(iii) Matrix $\mathbf{C}(\nu)$ accounts for centrifugal and Coriolis effects:

$$\mathbf{C}(\mathbf{v}) = \begin{bmatrix} 0 & 0 & c13 \\ 0 & 0 & c23 \\ c13 & -c23 & 0 \end{bmatrix}, \tag{3}$$

where

$$\begin{aligned} c13 &= -m\left(X_g r + v\right) + Y_{\dot{v}} v + \frac{1}{2}\left(N_{\dot{v}+Y_r}\right) r, \\ c23 &= mu - X_{\dot{u}} u. \end{aligned} \tag{4}$$

(iv) Matrix $\mathbf{D}(\nu)$ accounts for viscous and dissipative effects:

$$\mathbf{D}(\mathbf{v}) = \begin{bmatrix} d11 & 0 & 0 \\ 0 & d22 & d23 \\ 0 & d32 & d33 \end{bmatrix}, \tag{5}$$

where

$$\begin{aligned} d11 &= -X_u - X_{|u|u}|u| - X_{uuu}u^2, \\ d22 &= -Y_v - Y_{|v|v}|v| - Y_{|r|v}|r|, \\ d23 &= -Y_r - Y_{|v|r}|v| - Y_{|r|r}|r|, \\ d32 &= -N_v - N_{|v|v}|v| - N_{|r|r}|r|, \\ d33 &= -N_r - N_{|v|r}|v| - N_{|r|r}|r|. \end{aligned} \tag{6}$$

(v) Vector τ denotes the force and torques generated by actuators.

(vi) τ_{wind} denotes force and torques caused by wind

(vii) τ_{wave} denotes force and torques caused by wave.

The vessel model used in this paper is CyberShip II [14], which has a bow thruster, two thrusters, and two rudders. For the convenience of modeling force and torques, the bow thruster is excluded from the actuator model, and two thrusters and rudders are equivalent to one. Thus, the actuator model can be written as

$$\boldsymbol{\tau} = \mathbf{B}\boldsymbol{\tau}_{\text{act}}(\mathbf{v}, \mathbf{n}, \boldsymbol{\delta}), \tag{7}$$

where matrix $\mathbf{B}$ is the actuator configuration matrix and vector $\tau_{\text{act}}(\nu, \mathbf{n}, \delta)$ is the force and torques related to speed of the motor n and rudder angle δ:

$$\mathbf{B}(\nu) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & -2|l_{xR}| \end{bmatrix}, \boldsymbol{\tau}_{\text{act}}(\mathbf{v}, \mathbf{n}, \boldsymbol{\delta}) = \begin{bmatrix} T & L \end{bmatrix}^{\top}, \tag{8}$$

with

$$\begin{aligned} T &= T_{|n|n}|n|n - T_{|n|u}|n|u, \\ L &= \left(L_{\delta}\delta - L_{\delta}\delta|\delta|\delta\right)|u|u. \end{aligned} \tag{9}$$

2.2. Discrete Nonlinear Model. In the vessel experiments, the experiment data are collected in the discrete form. Also, vessel motion controllers are usually designed in the discrete zone. Therefore, the continuous-time model equation (1) is approximated as a discrete model, and the parameters of the discrete model are identified in this paper. The external force is all ignored for the convenience of discretizing the continuous-time model, and equation (1) can be rewritten as

$$\mathbf{M}\dot{\nu} = \mathbf{A}_c\boldsymbol{\alpha}, \tag{10}$$

where the matrix $\mathbf{A}_c$ and vector α denote the time-varying terms and constant parameters in the vessel model:

$$\mathbf{A}_c\alpha = -\mathbf{C}(\mathbf{v})\mathbf{v} - \mathbf{D}(\mathbf{v})\mathbf{v} + \mathbf{B}\boldsymbol{\tau}_{\text{act}}(\mathbf{v}, \mathbf{n}, \boldsymbol{\delta}),$$

$$\boldsymbol{\alpha} = \begin{bmatrix} u & |u|u & u^3 & r^2 & vr & n^2 & nu & 0 & 0 & 0 \\ uv & ur & v & r & |v|v & |v|r & |r|v & |r|r & |u|u\delta & |u|u|\delta|\delta \\ uv & ur & v & r & |v|v & |v|r & |r|v & |r|r & |u|u\delta & |u|u|\delta|\delta \end{bmatrix}^{\top},$$

$$\mathbf{A}_c = \begin{bmatrix} X_u & X_{|u|u} & X_{uuu} & mx_g - \frac{1}{2}\left(N_{\dot{v}} + Y_{\dot{r}}\right) & m - Y_{\dot{v}} & 2T_{|n|n} & -2T_{|n|u} & 0 & 0 & 0 \\ 0 & X_{\dot{u}} - m & Y_v & Y_r & |Y_{|v|v} & |Y_{|v|r} & |Y_{|r|v} & Y_{|r|r} & 2L_{\delta} & -2L_{|\delta|\delta} \\ Y_{\dot{v}} - X_{\dot{u}} & \frac{1}{2}\left(N_{\dot{v}+Y_{\dot{r}}}\right) - mx_g & N_v & N_r & N_{|v|v} & N_{|v|r} & N_{|r|v} & N_{|r|r} & -2|l_xR|L_{\delta} & 2|l_xR|L_{|\delta|\delta} \end{bmatrix}^{\top}. \tag{11}$$

The sample time of collecting data in simulation is denoted as T_s; equation (10) can be approximated by the back Euler integration method at time k as

$$\begin{aligned} \mathbf{v}(\mathbf{k}+1) &= \mathbf{v}(\mathbf{k}) + T_s\mathbf{M}^{-1}\mathbf{A}_c\boldsymbol{\alpha}(\mathbf{k}) \\ &= \mathbf{B}_c\boldsymbol{\alpha}(\mathbf{k}) + T_s\mathbf{M}^{-1}\mathbf{A}_c\boldsymbol{\alpha}(\mathbf{k}), \end{aligned} \tag{12}$$

where $\mathbf{B}_c = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^{\top}$.

Because the matrix $\mathbf{B}_c$ and $T_s\mathbf{M}^{-1}\mathbf{A}_c$ are both constant matrix, thus, equation (12) can be rewritten as

$$\begin{aligned} u(k+1) = {} & a_1u(k) + a_2|u(k)|u(k) + a_3u(k)^3 \\ & + a_4r(k)^2 + a_5v(k)r(k) + a_6n(k)^2 + a_7n(k)u(k), \end{aligned} \tag{13}$$

$$\begin{aligned} v(k+1) = {} & b_1u(k)v(k) + b_2u(k)r(k) + b_3v(k) \\ & + b_4r(k) + b_5|v(k)|v(k) + b_6|v(k)|r(k) \\ & + b_7|r(k)|v(k) + b_8|r(k)|v(k) \\ & + b_9|u(k)|u(k)\delta(k) + b_{10}|u(k)|u(k)|\delta(k)|\delta(k), \end{aligned} \tag{14}$$

$$\begin{aligned} r(k+1) = {} & c_1u(k)v(k) + c_2u(k)r(k) + c_3v(k) \\ & + c_4r(k) + c_5|v(k)|v(k) + c_6|v(k)|r(k) \\ & + c_7|r(k)|v(k) + c_8|r(k)|v(k) \\ & + c_9|u(k)|u(k)\delta(k) + c_{10}|u(k)|u(k)|\delta(k)|\delta(k). \end{aligned} \tag{15}$$

The parameters $a_1, \ldots, a_7$, $b_1, \ldots, b_{10}$, and $c_1, \ldots, c_{10}$ are unknown parameters. In Section 3, an ACO method is used to identify parameters of equations (13)–(15).

3. Parameter Identification Based on the ACO Method

In this section, the parameter of discrete vessel nonlinear model equations (13)–(15) is identified based on the ACO method.

3.1. Parameter Identification Formulation. Consider that the vessel has two main motion states: one is the straight-line surge motion and the other is the turning motion. Therefore, parameter identification includes two steps: the first step is to identify the parameters related to surge motion and the

second step is to identify the residual parameters. In the straight-line surge motion, the lateral velocity and yaw angular velocity are both close to zero. Thus, the surge motion can be decoupled from the other motion, the parameters related to straight-line motion and thrust force can be identified. By ignoring the terms related to the lateral and yaw motion, equation (13) can be simplified as

$$\begin{aligned} u(k+1) &= a_1 u(k) + a_2 |u(k)| u(k) + a_6 n(k)^2 + a_7 n(k) u(k) \\ &= \boldsymbol{\theta}\boldsymbol{\beta}(u(k), n(k))^\top, \end{aligned} \tag{16}$$

where $\theta = [\, a_1 \;\; a_2 \;\; a_6 \;\; a_7 \,]^\top$ and $\beta(u(k), n(k))^\top = [\, u(k) \;\; |u(k)|u(k) \;\; n(k)^2 \;\; n(k)u(k) \,]^\top$.

Based on equation (16), the predicted vessel surge velocity at time $k+n$, denoted as $\widehat{u}(k+n+1)$, can be calculated by the motor speed at time $k+n$, denoted as $n(k+n)$, and predicted as surge velocity $\widehat{u}(k+n)$. Then, the surge motion model identification problem can be described as a nonlinear optimization problem:

$$\begin{aligned} &\min f(\boldsymbol{\theta}) = \frac{1}{2} \sum_{k=1}^{N} \|\widehat{u}(k) - u(k)\|_2, \\ &s.\text{t.} \quad \widehat{u}(1) = u(1), \\ &\qquad \widehat{u}(k) = \boldsymbol{\theta}\boldsymbol{\beta}(\widehat{u}(k-1), n(k-1))^\top, \quad k \geq 2, \end{aligned} \tag{17}$$

where $\widehat{u}$ denotes the predicted surge speed based on the parameter θ and u denotes the surge speed collected in the experiments.

Since the parameters θ are identified in equation (17), then the rest parameters can also be described as a nonlinear optimization problem to identify. Based on equations (13)–(15), the vessel state can be predicted by

$$\begin{aligned} \mathbf{v}(k+1) = {} & \boldsymbol{\theta}_2 \boldsymbol{\alpha}_1 (u(k), v(k), r(k), n(k), \delta(k))^\top \\ & + \boldsymbol{\alpha}_2 (u(k), n(k))^\top, \end{aligned} \tag{18}$$

where

$$\begin{aligned} \boldsymbol{\theta}_2 &= \begin{bmatrix} 0 & 0 & 0 & a_4 & a_5 & 0 & 0 & 0 & 0 & 0 \\ b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 & b_9 & b_{10} \\ c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 & c_9 & c_{10} \end{bmatrix}, \\ \boldsymbol{\alpha}_2 (u(k), n(k))^\top &= \begin{bmatrix} \boldsymbol{\theta}\boldsymbol{\beta}(u(k), n(k))^\top \\ 0 \\ 0 \end{bmatrix}, \\ \boldsymbol{\alpha}_1 &= \begin{bmatrix} 0 & 0 & 0 & r^2 & vr & 0 & 0 & 0 & 0 & 0 \\ uv & ur & v & r & |v|v & |v|r & |r|v & |r|r & |u|u\delta & |u|u|\delta|\delta \\ uv & ur & v & r & |v|v & |v|r & |r|v & |r|r & |u|u\delta & |u|u|\delta|\delta \end{bmatrix}^\top. \end{aligned} \tag{19}$$

Based on equation (18), the predicted vessel state at time $k+n$, denoted as $\widehat{v}(k+n+1)$, can be calculated by the motor speed $n(k+n)$, rudder angle $\delta(k+n)$, the predicted surge velocity $\widehat{u}(k+n)$, the predicted lateral velocity $\widehat{v}(k+n)$, and the predicted yaw angular velocity $\widehat{r}(k+n)$. Then, the parameters' identification problem can be described as a nonlinear optimization problem:

$$\begin{aligned} &\min f(\boldsymbol{\theta}_2) = \frac{1}{2} \sum_{k=1}^{N} \|\widehat{v}(k) - \mathbf{v}(k)\|, \\ &s.\text{t.} \quad \widehat{v}(1) = \mathbf{v}(1), \\ &\qquad \widehat{v}(\mathbf{k}+1) = \boldsymbol{\theta}_2 \boldsymbol{\alpha} 1 (\widehat{u}(k), \widehat{v}(k), \widehat{r}(k), n(k), \delta(k))^\top \\ &\qquad\qquad + \boldsymbol{\alpha}_2 (\widehat{u}(k), n(k))^\top, \quad k \geq 2. \end{aligned} \tag{20}$$

The parameters now can be identified by solving two nonlinear optimization problem equations (17) and (20).

3.2. ACO Formulation. As mentioned above, the key to identify the model parameters is solving an optimization problem as equation (17) or equation (20). Generally, the typical unconstrained continuous optimization problem is shown in equation (21), where $f(x)$ is the objective function and X is the decision variable ($D.V$) in the form of an N-dimensional vector in which the member changes continuously:

$$M\text{inf}(x), \quad X = [\, x_1 \;\; x_2 \;\; \cdots \;\; x_N \,]. \tag{21}$$

As far as equations (17) and (20) are concerned, $f(\theta)$ and $f(\theta_2)$ are objective functions of the two identification problems, respectively. Accordingly, θ and θ_2 are high-dimensional $D.$Vs.

Traditional optimization algorithms find it difficult to solve the nonlinear high-dimensional optimization problems described in equations (17) and (20). A new computational paradigm called "Ant System" is proposed in [31], which is used for stochastic combinatorial optimization and

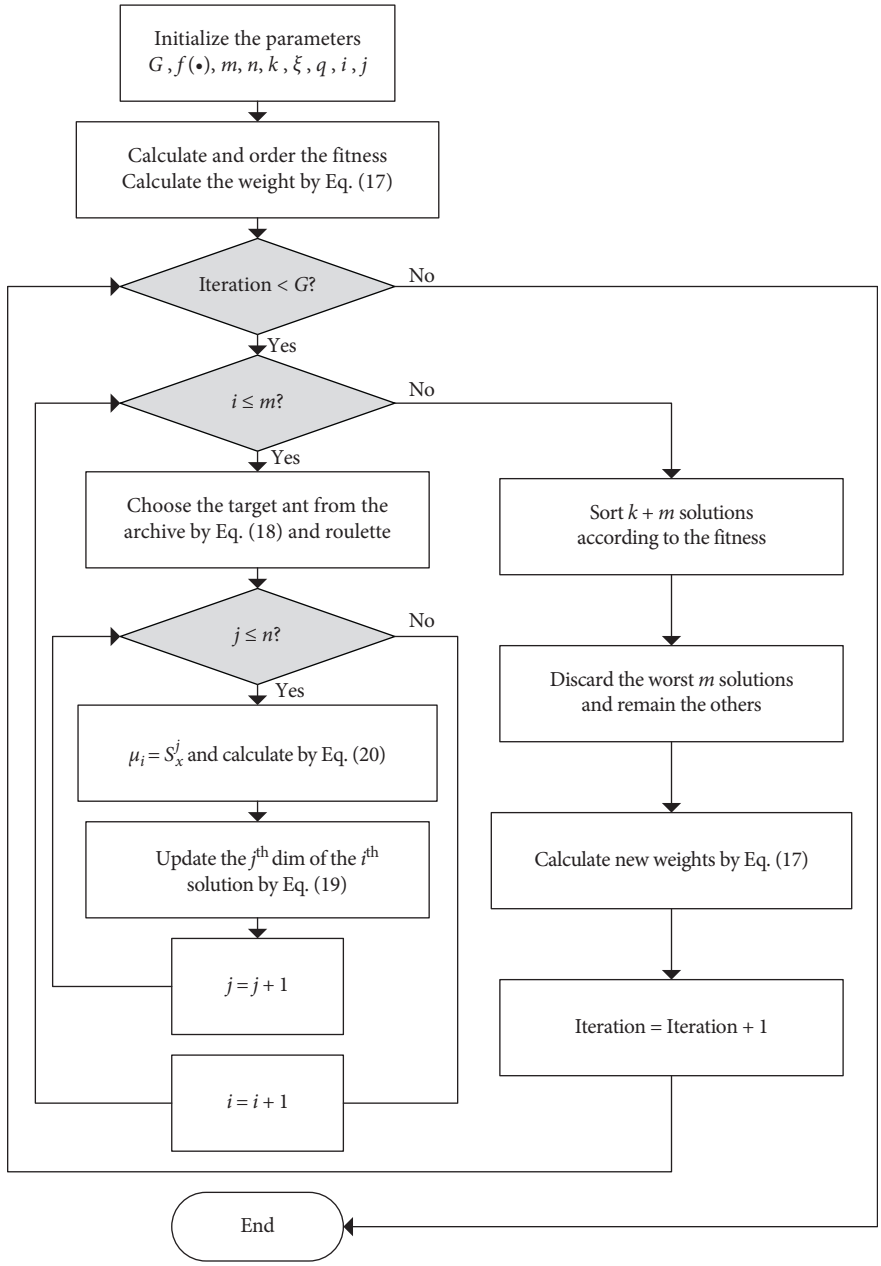


Figure 1: ACO method flowchart.

motivated amounts of interrelated solutions to improve the performances. The representative one is Ant Colony Optimization (ACO). The parameter identification problem we focused on, however, is a typical optimization problem of which decision variables ($D.V$) vary in continuous domains.

Despite the method that discretizes, the feasible region takes the effect to a certain extent; the accuracy and efficiency are severely restricted. Socha and Dorigo [32] extend ACO to continuous domains and keep the relevant conceptual structure invariable. The complete flowchart of the CACO is shown in Figure 1. In CACO, the fundamental work is initializing the elementary data structures and parameters to abstract the actual optimization problem and determine the initial state. Pheromone, inspired by the natural behavior of ants, which records the excellent solutions, is the significant feature of ACO. Different from the ACO which uses a high-dimensional matrix to represent the pheromone, CACO applies the continuous probability density function (CPDF), such as the Gaussian function. Furthermore, CACO maintains a solution archive of capacity k composed of k n-dimensional decision variables $(S_j^1, S_j^2, \ldots, S_j^n)$, which are constructed by ants, and the corresponding fitness $f(S_j)$ is to memorize the several former solutions. For the surge model identification problem, the variable n is 4 and the fitness is given by $f(\theta)$. As for the remaining parameters identification problem, the variable n is determined by θ_2, that is, 22, and the fitness is calculated by $f(\theta_2)$. Moreover, the solution vectors are supposed to be initialized using

random values, sampling from the search space $\mathfrak{R}$. The other related parameters which will be discussed below are supposed to be set reasonably at this stage.

As shown in Figure 1, once the initialization work is accomplished, the algorithm will step into searching for solutions in $\mathfrak{R}$ and further instruct to update the pheromone. In the course of searching for solutions for an individual ant, the first step is choosing a target; that is, select a solution from the archive based on the selected probability following equations (22) and (23) and the roulette method. The value l is the rank of the corresponding solution, k is the size of the archive, and q is an extra parameter to adjust the standard deviation, which has an effect on the shot probability of different ranks. Namely, q is a trade-off factor between the global and local optimal solutions:

$$\omega_1 = \frac{1}{qk\sqrt{2\pi}} e^{-\left((l-1)^2/2q^2k^2\right)}, \tag{22}$$

$$p_l = \frac{\omega_1}{\sum_{r=1}^{k} \omega_r}. \tag{23}$$

It is worth noting that the ants in ACO or CACO are not a container (i.e., a vector composed of decision variables) but a solution builder (i.e., a pathfinder in the search space to construct the solution vector). For an n-dimensional decision variable, an ant needs n steps to accomplish the construction per iteration. Once the target is selected, the destination of the i^{th} dimension variable is assigned by Gaussian sampling using the function described as

$$g_l^i\left(x, \mu_l^i, \sigma_l^i\right) = \frac{1}{\sigma_l^i\sqrt{2\pi}} e^{-\left((x-\mu_l^i)/2\mu_l^{i^2}\right)}. \tag{24}$$

The parameter $\mu_l^i = s_l^i$, and σ_l^i is calculated as equation (25), where ξ has equivalent effect with the pheromone evaporation rate which can influence the speed of convergence:

$$\sigma_l^i = \xi \sum_{j=1}^{k} \frac{\left|s_j^i - s_l^i\right|}{k-1}. \tag{25}$$

On account of the pheromone reflects on the solution archive, an update-pheromone operation may be accomplished by renewing the table. After all of the ants complete the construction of solutions, there will be $m + k$ solutions. For the sake of the superior solution to maintain the archive, all solutions should be ordered according to the evaluation index (e.g., the root mean square error between predicting results and sample data). In the subsequent step, the worst m solutions would be discarded and the others remain in the archive.

After updating the archive, the algorithm tests' termination conditions (such as the maximum iterations and convergence precision) are in accordance with the optimization results of the current generation. If the termination condition is not met, the optimization problem will repeat the searching, evaluating, and renewing process mentioned above until the condition is satisfied. Otherwise, the iteration should be broken up and the optimization results should be output, which are the optimized model parameters θ and θ_2 and the corresponding model performance.

4. Simulation Results

In this paper, the CyberShip II is used as a study example. The kinetics model structure of the vessel is already known, and the parameters of the kinetics model are unknown. The parameters are identified by the ACO method and the LS method, respectively, and the identification results are compared to demonstrate the effectiveness of the ACO method. Normally, in the vessel identification experiments, the vessel states and actuator states such as velocity, angular velocity, motor speed, and rudder angle are collected by the sensor with noise. Therefore, the Gaussian white noise is added in the whole identification simulations to simulate the sensor noise. The whole identification procedure is divided into two steps, the first step is to identify the parameters related to the surge motion and the second step is to identify the parameters related to the sway and yaw motion.

4.1. Surge Model Identification. In the vessel surge motion, sway velocity and yaw angular velocity are approximately equal to zero. Therefore, the couple terms of the vessel kinetics model related to sway and yaw motion can be ignored. The purpose of the surge model identification is to identify the parameters a_1, a_2, a_3, a_5, and a_6 in equation (13).

The vessel motor speed is used as the excitation input which varies randomly between 5 rpm and 25 rpm, and the vessel surge velocity as the response to the excitation input varies between 0.2 m/s and 0.8 m/s, as shown in Figure 2. The transient motor speed and surge speed are used as training data to identify the parameters. The parameters related to the ACO method in solving surge model identification problem are shown Table 1, and the ACO method convergence curve are shown in Figure 3. From Figure 3, we can see the fitness which is the value of $f(\theta)$ in equation (17) gradually decreases and becomes stable after 450 iterations. The identified results of the ACO method and LS method are shown in Table 2.

In Figure 2, the prediction results of parameters identified by the different methods are very close. In addition, another two different surge motions are used as validation data to compare the performance of different algorithms, as shown in Figures 4 and 5. In Figure 4, the input motor speed is a PRBS sequence. From the results, we can see the prediction results of different methods are very close, except that the prediction results of the LS method are slower than the simulation data when the surge speed is larger than 0.9 m/s.

In Figure 5, the input motor speed is maintained at 5 rpm, 10 rpm, and 20 rpm, so the surge speed can finally be maintained at around 0.16 m/s, 0.36 m/s, and 0.67 m/s. We can see from the results that the prediction results of different methods are almost the same when the motor speed is kept at 5 rpm and 10 rpm. However, the prediction surge speed of the LS method is larger than the simulation results when the motor speed is kept at 10 rpm.

In order to quantify the comparison result of different methods, the R-square is used as an index function. The closer the R-square to 1, the better the identified results, and

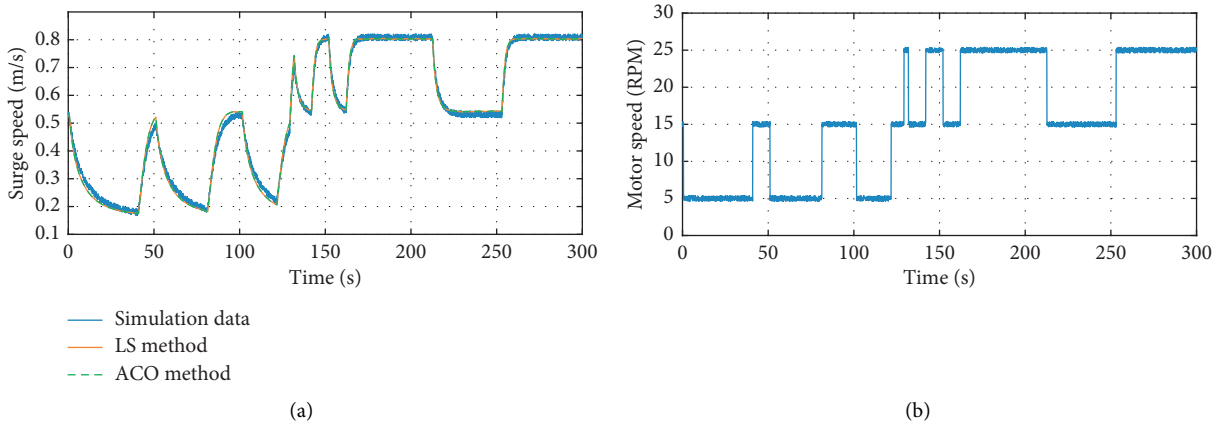


Figure 2: Identification data of the surge model.

Table 1: Parameters of the ACO method in surge model identification.

Parameters	n	m	k	q	ξ	G
Value	5	20	10	$1e-03$	0.75	500

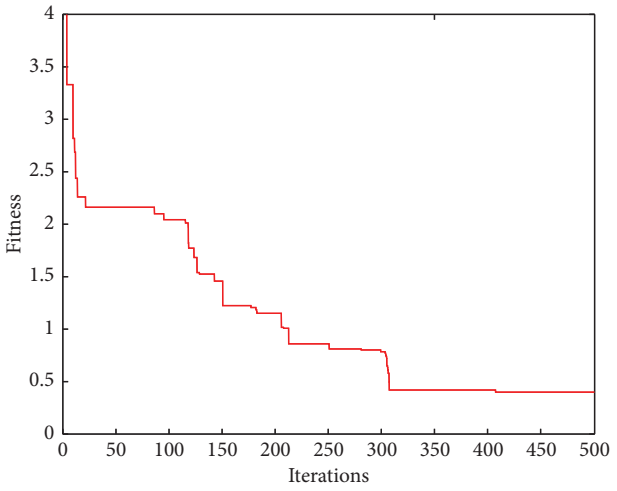


Figure 3: ACO-method convergence curve.

Table 2: Parameters of the surge model.

Parameters	a_1	a_2	a_3	a_6	a_7
LS	0.9953	−0.0060	−0.0388	$2.6357e-05$	$5.6225e-04$
ACO	0.9972	−0.0051	−0.0227	$2.8218e-05$	$-2.8661e-06$

in some situations, the R-square may be negative; then, we use the symbol − to indicate it. From the R-square results of different methods in Table 3, we can see the R-square values of the ACO method and the LS method are very close. However, the R-square value of the ACO method is better than the LS method in training data and validation data, so the ACO method has better performance than the LS method in the surge model parameters' identification simulation experiments.

4.2. Lateral and Yaw Model Identification. In the sway and yaw model identification simulation experiments, the rest parameters related to the sway and yaw motion in equations (13)–(15) are identified. Zigzag tests are widely used in vessel parameters' identification experiments [16, 24]. In this paper, a 20°/20° zigzag test contains constant motor speed which is used to generate the training data, as shown in Figure 6 and Table 4.

The convergence process of the ant colony algorithm is shown in Figure 7. The result shows that the optimal solution

Table 3: R-square values of different methods on the surge model.

R-square	Training data	Validation data 1	Validation data 2
LS method	0.9953	0.9959	0.9908
ACO method	0.9959	0.9982	0.9934

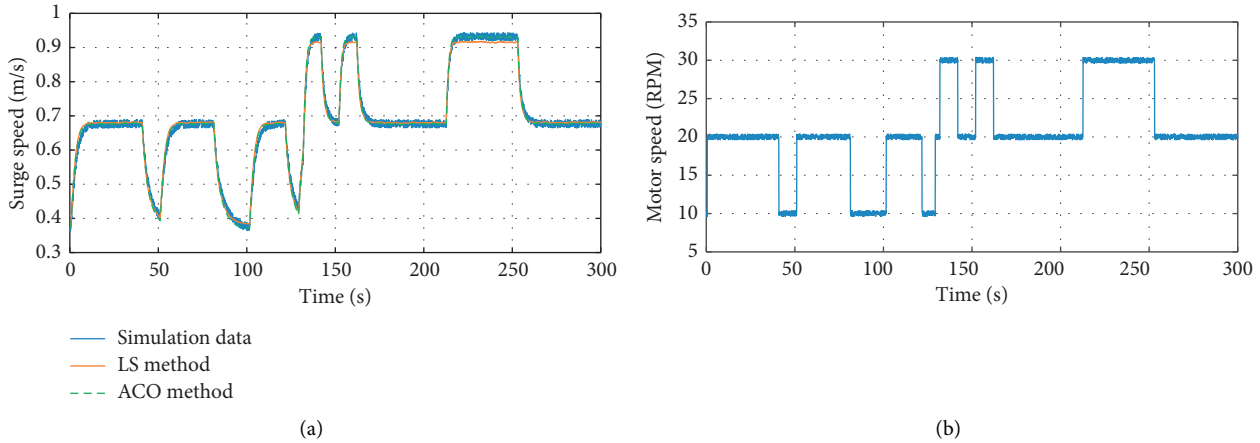


Figure 4: Validation data 1 (results of the surge model under PRBS motor speed).

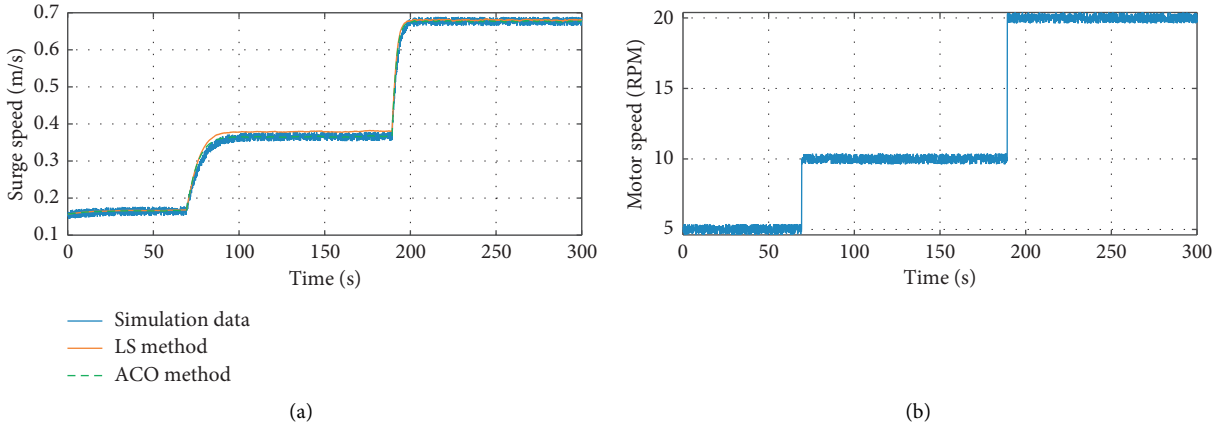


Figure 5: Validation data 2 (results of the surge model under different constant motor speeds).

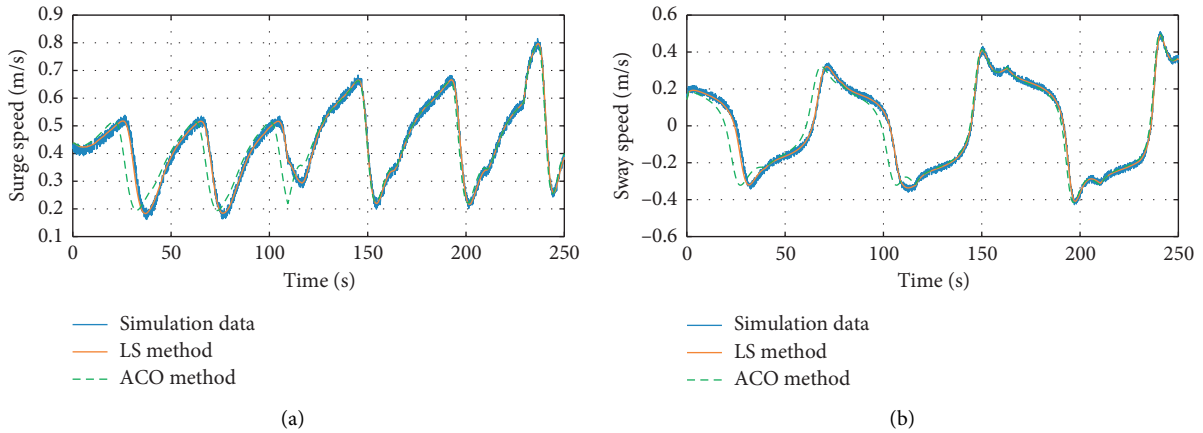


Figure 6: Continued.

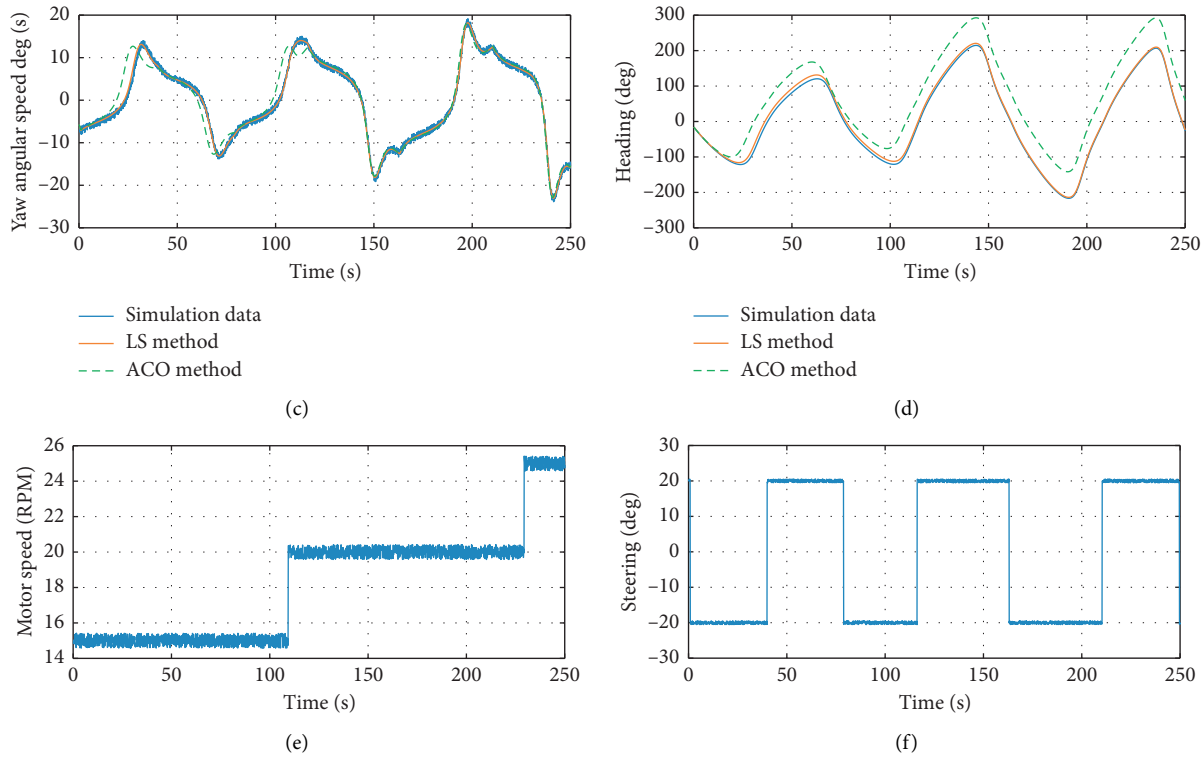


Figure 6: Identification results of vessel dynamic.

Table 4: Parameters of the ACO method in lateral and yaw model identification.

Parameters	n	m	k	q	ξ	G
Value	22	30	15	$1e-03$	0.75	1000

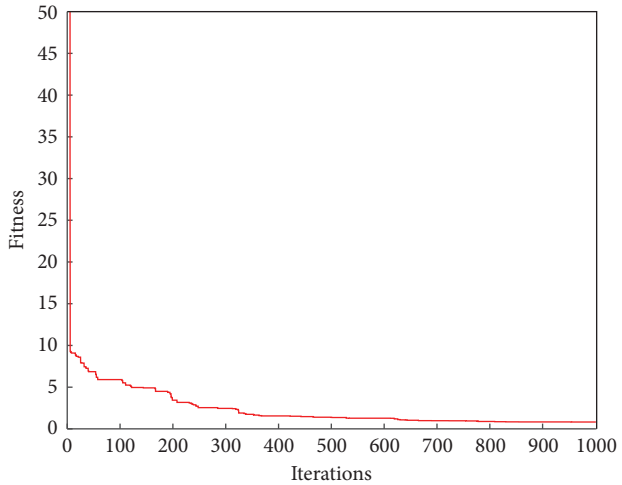


Figure 7: ACO method convergence curve.

almost converges to the expected range after 1000 generations. The parameters identified by the LS method and ACO method are shown in Table 5.

From Figure 6 we can see that the LS method has better prediction performance than the ACO method at the surge model, sway model, and yaw model. To further verify the effectiveness and the performance of the ACO method, we use another two different excitation inputs to generate the validation data which are shown in Figures 8 and 9, and for the convenience of comparison, the heading angle is not changed to the range from 0° to 360°. In addition, the R-square value is used to quantitatively compare the prediction performance for the two different methods, as shown Table 6.

In Figure 8, the motor speed maintains at 10 rpm, 20 rpm, and 30 rpm at different times, and the rudder angle is a PRBS sequence with different amplitudes. From the results, we can see the prediction results of ACO are slightly better than the LS method at a motor speed of 10 rpm, but the prediction results of the two methods are similar at a motor speed of 20 rpm and 30 rpm. The R-square value of the ACO method is closer to 1 than the LS method on the surge model, sway model, yaw model, and heading model.

In Figure 9, the motor speed maintains at 10 rpm, 20 rpm, and 30 rpm at different times, and the rudder angle maintains at 5°, −15°, and 25° at different times. From the results, we can see only at a motor speed of 5 rpm and rudder angle of 5°; the two methods have similar prediction results. From the R-square values, we can see the prediction results of the ACO method have very good performance, but the

Table 5: Parameters of lateral and yaw motion.

Parameters	a_3	a_4	b_1	b_2
LS	0.0024	0.1292	$-3.6851e-04$	−0.0789
ACO	0.0041	0.1302	0.0012	−0.0780
Parameters	b_3	b_4	b_5	b_6
Ls	0.9978	−0.0206	−0.1089	$-3.1260e-4$
ACO	0.9917	−0.0262	−0.1025	−0.0030
Parameters	b_7	b_8	b_9	b_{10}
Ls	−0.0026	−0.0126	0.0238	0.0083
ACO	−0.0019	−0.0090	0.0443	−0.0287
Parameters	c_1	c_2	c_3	c_4
Ls	−0.0413	0.0013	$-1.3609e-04$	0.9905
ACO	−0.0453	−0.0013	$3.2504e-04$	0.9864
Parameters	c_5	c_6	c_7	c_8
Ls	0.0277	−0.0017	0.0011	$-8.6515e-04$
ACO	0.0203	$5.3517e-04$	$7.6613e-04$	−0.0039
Parameters	c_9	c_{10}		
Ls	−0.0243	−0.0085		
ACO	−0.0342	0.0251		

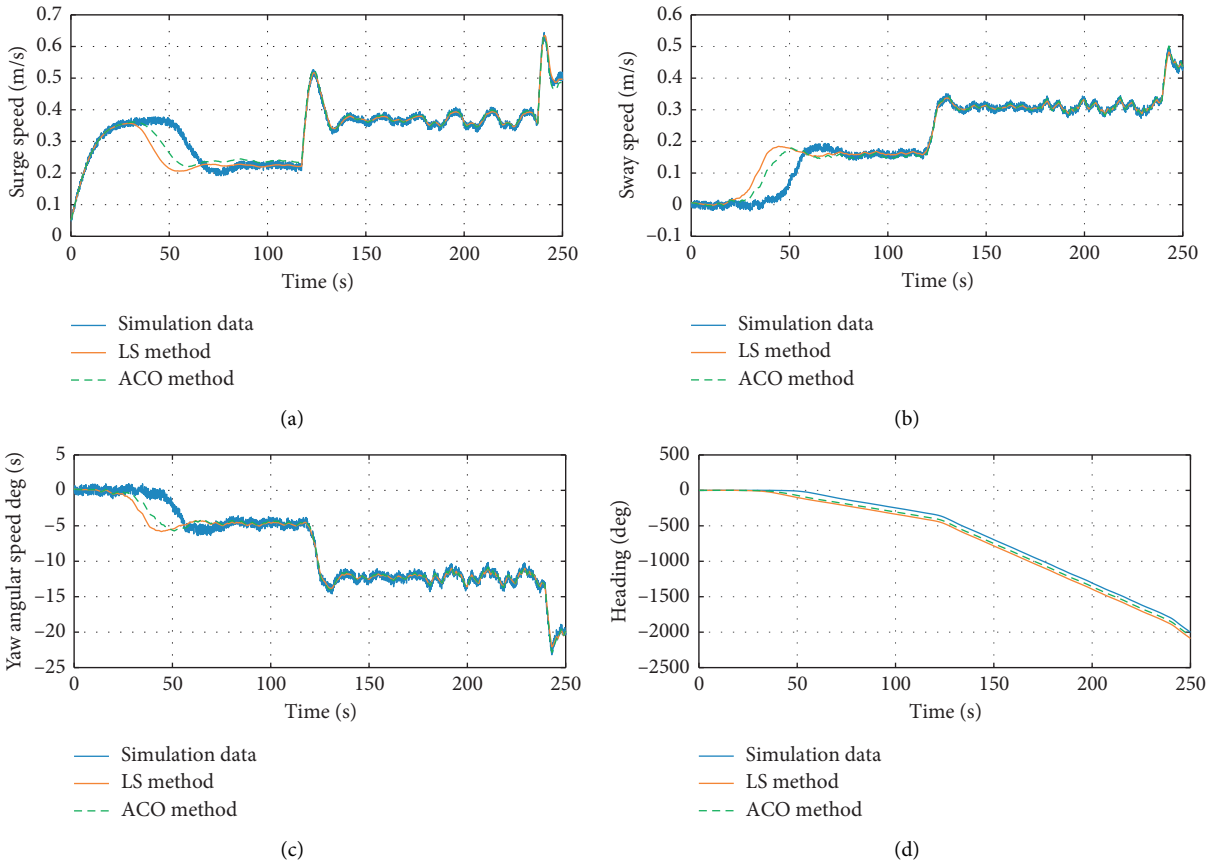


Figure 8: Continued.

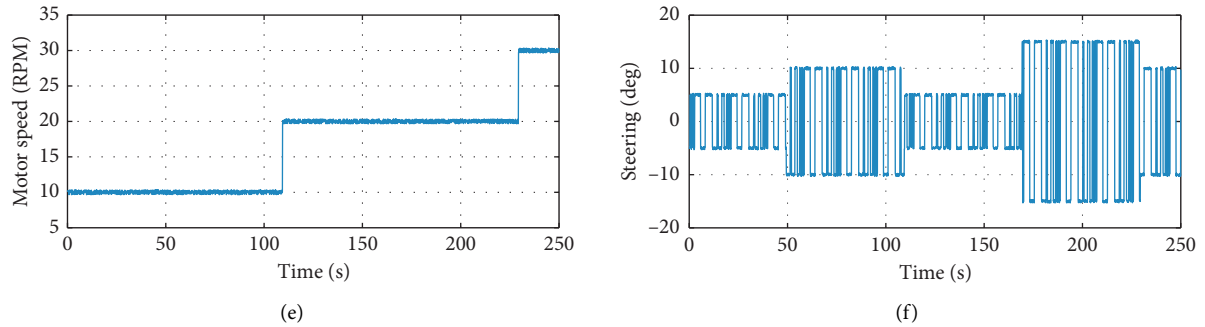


Figure 8: Validation results of vessel dynamic by the least-square method.

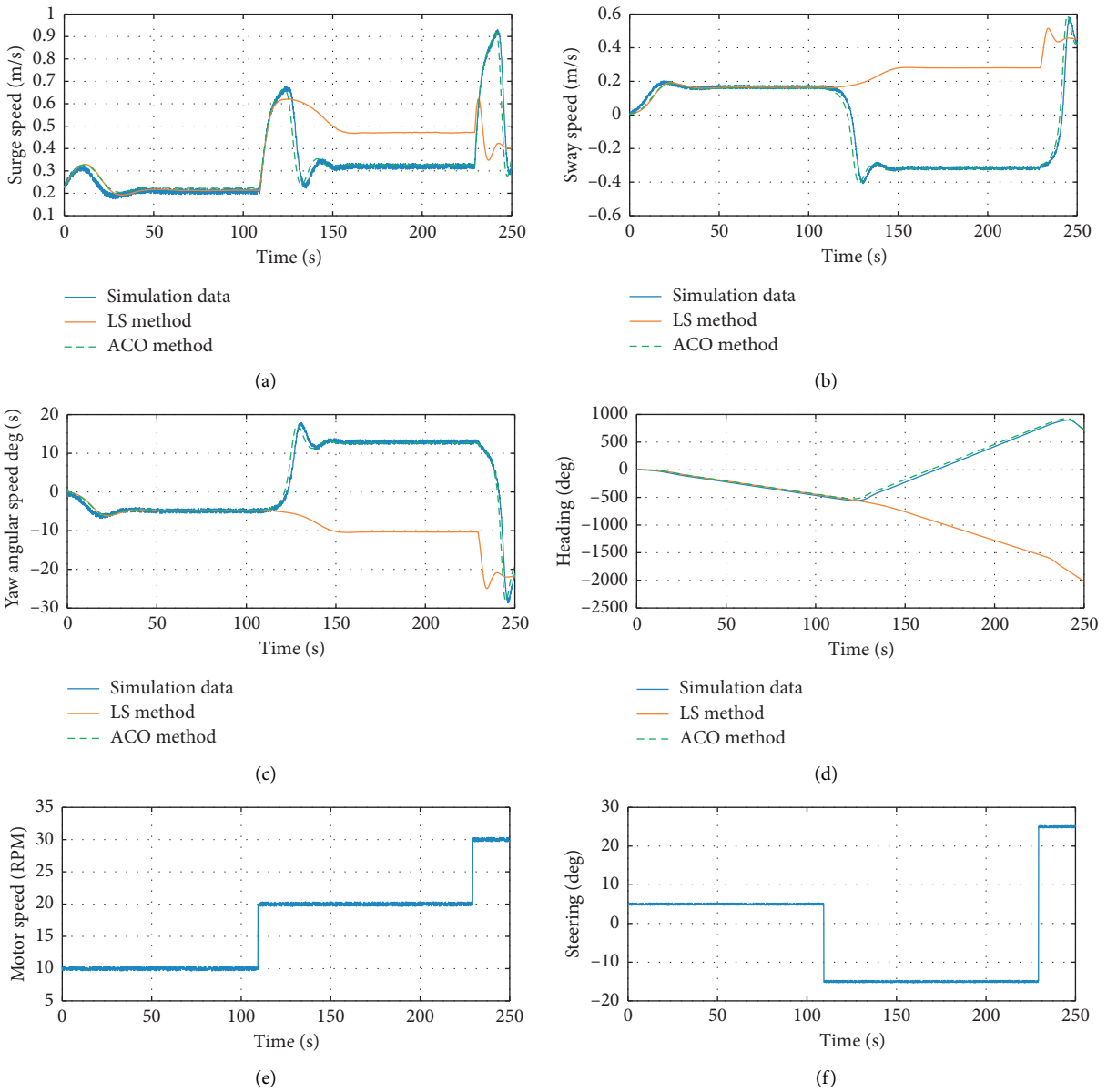


Figure 9: Validation results of vessel dynamic by the least-square method.

Table 6: R-square values of different methods on the vessel model.

Identification error	Surge model	Sway model	Yaw model	Heading model
LS	0.9890	0.9957	0.9952	0.9964
ACO	0.8478	0.9182	0.9417	0.6804
LS	0.9106	0.9458	0.9740	0.9955
ACO	0.9476	0.9672	0.9831	0.9982
LS	0.7050	—	—	—
ACO	0.9526	0.9745	0.9784	0.9942

prediction results of the LS method are only valid at low motor speed and small rudder angle.

In the identification results of the sway and yaw motion, although the prediction results of the LS method are better than the LS method in training data, the identified parameters of the LS method are invalid in predicting the surge, sway, and yaw motion when the motor speed and rudder angle maintain the constant value. However, the parameters identified by the ACO method have a good prediction result in the surge, sway, and yaw motion both in the training data and validation data.

5. Conclusions

This paper presents the parameters' identification under the known structure of the vessel kinetics model. The parameters' problem is summarized as a least-square problem; the least-square method and the ant colony optimization method are used to solve the least-square problem. The solution to the least-square problem is the parameters to be identified. The identification procedure is divided into two parts. The first part of the identification procedure is to identify the parameters related to surge motion. The second part of the identification procedure is to identify the rest parameters of the vessel kinetics model. In the identification procedure, the transient excitation inputs are used to generate the training data; the constant excitation inputs and different transient inputs are used to generate the validation data. The R-square value is used as an index function to quantitatively compare the prediction results of the ACO method and the LS method. The comparison of the identification results and maneuvering predictions demonstrate the effectiveness of the ACO method and reflect the performance advantage of the ACO method.

Data Availability

The data used to support the findings of this study are available within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Funds of the National Natural Science Foundation of China under Grant 61873050, in part by the Fundamental Research Funds for the Central Universities under Grant N2004010, in part by the Research Fund of State Key Laboratory of Synthetical Automation for Process Industries under Grant 2018ZCX14, and in part by LiaoNing Revitalization Talents Program XLYC1907088.

References

[1] Z. Liu, Y. Zhang, X. Yu, and C. Yuan, "Unmanned surface vehicles: an overview of developments and challenges," *Annual Reviews in Control*, vol. 41, pp. 71–93, 2016.

[2] S. Savitz, I. Blickstein, P. Buryk et al., *Us Navy Employment Options for Unmanned Surface Vehicles (Usvs)*, RAND National Defense Research Institute, Santa Monica, CA, USA, 2013.

[3] A. Annamalai and A. Motwani, "A comparison between lqg and mpc autopilots for inclusion in a navigation, guidance and control system," in *Marine and Industrial Dynamic Analysis*, vol. 1, pp. 1–19, School of Marine Science and Engineering, Plymouth University, Plymouth, England, 2013.

[4] P. Tang, R. Zhang, D. Liu, L. Huang, G. Liu, and T. Deng, "Local reactive obstacle avoidance approach for high-speed unmanned surface vehicle," *Ocean Engineering*, vol. 106, pp. 128–140, 2015.

[5] J. Li, G. Deng, C. Luo, Q. Lin, Q. Yan, and Z. Ming, "A hybrid path planning method in unmanned air/ground vehicle (uav/ugv) cooperative systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9585–9596, 2016.

[6] L. Liu, D. Wang, Z. Peng, and H. Wang, "Predictor-based los guidance law for path following of underactuated marine surface vehicles with sideslip compensation," *Ocean Engineering*, vol. 124, pp. 340–348, 2016.

[7] H. Zheng, R. R. Negenborn, and G. Lodewijks, "Trajectory tracking of autonomous vessels using model predictive control," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 8812–8818, 2014.

[8] Z. Liu, H. Jin, M. J. Grimble, and R. Katebi, "Ship forward speed loss minimization using nonlinear course keeping and roll motion controllers," *Ocean Engineering*, vol. 113, pp. 201–207, 2016.

[9] K. Nomoto, T. Taguchi, K. Honda, and S. Hirano, "On the steering qualities of ships," *International Shipbuilding Progress*, vol. 4, no. 35, pp. 354–370, 1957.

[10] Z. Lei and C. Guo, "Disturbance rejection control solution for ship steering system with uncertain time delay," *Ocean Engineering*, vol. 95, pp. 78–83, 2015.

[11] H. Zhou, L. Güvenç, and Z. Liu, "Design and evaluation of path following controller based on mpc for autonomous vehicle," in *Proceedings of the 2017 36th Chinese Control Conference (CCC)*, pp. 9934–9939, IEEE, Dalian, China, July 2017.

[12] L. Liu, D. Wang, and Z. Peng, "Eso-based line-of-sight guidance law for path following of underactuated marine surface vehicles with exact sideslip compensation," *IEEE*

Journal of Oceanic Engineering, vol. 42, no. 2, pp. 477–487, 2016.

[13] Z. Zheng, L. Sun, and L. Xie, "Error-constrained los path following of a surface vessel with actuator saturation and faults," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 10, pp. 1794–1805, 2017.

[14] R. Skjetne, Ø. Smogeli, and T. I. Fossen, "Modeling, identification, and adaptive maneuvering of cybership ii: a complete design with experiments," *IFAC Proceedings Volumes*, vol. 37, no. 10, pp. 203–208, 2004.

[15] J. Wang, L. Zou, and D. Wan, "Cfd simulations of free running ship under course keeping control," *Ocean Engineering*, vol. 141, pp. 450–464, 2017.

[16] M. Araki, H. Sadat-Hosseini, Y. Sanada, K. Tanimoto, N. Umeda, and F. Stern, "Estimating maneuvering coefficients using system identification methods with experimental, system-based, and cfd free-running trial data," *Ocean Engineering*, vol. 51, pp. 63–84, 2012.

[17] C. Sonnenburg, A. Gadre, D. Horner et al., "Control-oriented planar motion modeling of unmanned surface vehicles," in *Proceedings of the OCEANS 2010 MTS/IEEE SEATTLE*, pp. 1–10, IEEE, Sydney, Australia, May 2010.

[18] G. Zhang, X. Zhang, and H. Pang, "Multi-innovation auto-constructed least squares identification for 4 dof ship manoeuvring modelling with full-scale trial data," *ISA Transactions*, vol. 58, pp. 186–195, 2015.

[19] M. Ertogan, P. A. Wilson, G. T. Tayyar, and S. Ertugrul, "Optimal trim control of a high-speed craft by trim tabs/interceptors part i: pitch and surge coupled dynamic modelling using sea trial data," *Ocean Engineering*, vol. 130, pp. 300–309, 2017.

[20] H. K. Yoon, N. S. Son, and G. J. Lee, "Estimation of the roll hydrodynamic moment model of a ship by using the system identification method and the free running model test," *IEEE Journal of Oceanic Engineering*, vol. 32, no. 4, pp. 798–806, 2007.

[21] K. R. Muske, H. Ashrafiuon, G. Haas, R. McCloskey, and T. Flynn, "Identification of a control oriented nonlinear dynamic usv model," in *Proceedings of the 2008 American Control Conference*, pp. 562–567, IEEE, Seattle, WA, USA, June 2008.

[22] J. Xiong, Y. He, F. Gu, D. Li, and J. Han, "Quasi-lpv modeling and identification for a water-jet propulsion usv: an experimental study," in *Proceedings of the 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014)*, pp. 431–436, IEEE, Bali, Indonesia, December 2014.

[23] J. Han, J. Xiong, Y. He, F. Gu, and D. Li, "Nonlinear modeling for a water-jet propulsion usv: an experimental study," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 4, pp. 3348–3358, 2016.

[24] C. Jian, Z. Jiayuan, X. Feng et al., "Parametric estimation of ship maneuvering motion with integral sample structure for identification," *Applied Ocean Research*, vol. 52, pp. 212–221, 2015.

[25] W. Luo, C. Guedes Soares, and Z. Zou, "Parameter identification of ship maneuvering model based on support vector machines and particle swarm optimization," *Journal of Offshore Mechanics and Arctic Engineering*, vol. 138, no. 3, 2016.

[26] W. Luo and X. Li, "Measures to diminish the parameter drift in the modeling of ship manoeuvring using system identification," *Applied Ocean Research*, vol. 67, pp. 9–20, 2017.

[27] Z. Wang, Z. Zou, and C. Guedes Soares, "Identification of ship manoeuvring motion based on nu-support vector machine," *Ocean Engineering*, vol. 183, pp. 270–281, 2019.

[28] E. R. Herrero and F. J. V. Gonzalez, "Two-step identification of non-linear manoeuvring models of marine vessels," *Ocean Engineering*, vol. 53, pp. 72–82, 2012.

[29] P. Du, A. Ouahsine, K. T. Toan, and P. Sergent, "Simulation of ship maneuvering in a confined waterway using a nonlinear model based on optimization techniques," *Ocean Engineering*, vol. 142, pp. 194–203, 2017.

[30] T. I. Fossen, *Handbook of Marine Craft Hydrodynamics and Motion Control*, John Wiley & Sons, New Jersey, NY, USA, 2011.

[31] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 26, no. 1, pp. 29–41, 1996.

[32] K. Socha and M. Dorigo, "Ant colony optimization for continuous domains," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1155–1173, 2008.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 7240683, 9 pages
https://doi.org/10.1155/2021/7240683

Research Article

An Access Network Selection Algorithm for Terrestrial-Satellite Networks Based on a QoS Guarantee

Yongzhou Lu,[1] Yuanbao Chen,[2] Shuang Huang,[2] and Mingqi Zhang[3]

[1]*Naval Research Academy, Beijing 100161, China*
[2]*Wuhan Second Ship Design and Research Institute, Wuhan 430064, China*
[3]*School of Electronic Information, Wuhan University, Wuhan 430064, China*

Correspondence should be addressed to Mingqi Zhang; zmq199708@whu.edu.cn

Received 7 April 2021; Accepted 7 June 2021; Published 16 June 2021

Academic Editor: Xiao Ling Wang

Considering the communication requirements of different monitoring services and the performance of different access networks, to ensure the quality of service and reliability of sensor networks, we propose an access network selection algorithm based on a quality-of-service (QoS) guarantee. According to the service requirements, the attributes that affect the terminal access options are determined. Then, utility functions are designed for normalizing the attribute values of the network. In addition, we construct a weight calculation model that combines subjective and objective factors. Finally, the algorithm calculates the relative closeness between each pair of networks by the technique for order preference by similarity to an ideal solution (TOPSIS) and sorts the candidate networks to obtain the best candidate network of each monitoring terminal to access. Simulation results show that the proposed algorithm can effectively meet the communication needs of different monitoring services, ensures network reliability, and can improve the utilization rate of network resources.

1. Introduction

With the rapid development of the economy, transmission lines have more hidden dangers. A transmission line stretches for thousands of kilometers; hidden dangers may occur everywhere [1]. In some areas, the manual way is still used to maintain the lines. However, manual inspection has some disadvantages, such as high costs of resources, a heavy workload, untimely detection of hidden dangers, and limited inspection areas. Even though some transmission lines have been installed with online monitoring equipment, greatly improving their efficiency in detecting hidden dangers, due to high traffic fees, most areas adopt methods such as taking regular photos or watching short videos to monitor lines.

Compared with the existing transmission line monitoring technology, the number of terminals that will be used in the future is huge, and the monitoring content will tend to be video-based and possess high-resolution. Therefore, it will have higher requirements for the coverage, flexibility, and costs of communication methods than current content. With the construction of 5G networks, the coverage of mobile networks has been improved, as 5G networks provide reliable communication support for data streams. However, on the ground in poor areas, 5G networks capacity is limited, and network reliability and real-time performance cannot be guaranteed [2].

Aiming at transmission line monitoring in remote areas, the low Earth orbit (LEO) satellite network, whose coverage is not limited by terrestrial conditions and location, is introduced as a supplementary coverage method for the terrestrial mobile networks. Compared with other satellite networks, an LEO satellite has the characteristics of large bandwidth and low delay, so these satellites will be an important part of the integrated network in space and on Earth in the future [3, 4].

However, in the terrestrial-satellite network, the available network resources are relatively limited. There are many types of monitoring services, and their requirements vary greatly. Therefore, how to reasonably select access networks and maximize resource utilization rates, to meet the needs of different monitoring services, is a subject worth studying [5].

To solve the problem that traditional access networks selection algorithm is unable to correlate the service requirements with the network performance, this study focuses on an access network selection algorithm for a terrestrial-satellite network. Based on the analysis of transmission line monitoring in remote areas, the influencing factors of access network selection are determined, and the access network selection algorithm is studied to maximize the reliability of monitoring services.

The remainder of this article is organized as follows: Section 2provides an overview of the related work. Section 3 describes the proposed access network selection algorithm. A simulation is discussed in Section 4. Finally, Section 5concludes this article.

2. Related Work

Because LEO satellites and 5G networks are still in the early stage of construction, the study of terrestrial-satellite networks is rare. However, the terrestrial mobile networks, such as macro sites and base stations, are a hotspot of research; scholars have conducted many studies in this field. The terrestrial-satellite network used in this paper is similar to the existing heterogeneous wireless networks to some extent. The access network selection algorithms used for the sensor networks are analyzed and studied in the following subsections.

2.1. Access Network Selection Algorithm Based on the Received Signal Strength. An access network selection algorithm based on the received signal strength relies on a single index. Gwon [6] and Yan [7] set the threshold value for the received signal strength. If the access signal strength of the candidate network exceeds the threshold value of the received signal strength, it is considered that the network meets the needs of users and is regarded as an appropriate access network. Based on setting the threshold value for the received signal strength, Song [8] added network throughput as a reference index to comprehensively select access networks by combining the two indexes. Kunarak and Suleesathira [9] proposed a received signal strength prediction method based on fuzzy logic, and they took the signal strength predicted by a back propagation neural network (BPNN) at the next moment to select a network.

This kind of algorithm has low complexity and little difficulty in terms of implementation. However, the use of the received signal strength as the main indicator of access network selection does not meet the specific QoS demands of many users. Therefore, the network selected is often not the network that best meets the needs of these users.

2.2. Access Network Selection Algorithm Based on Multiattribute Decision Making. A single-attribute access network selection algorithm cannot fully reflect the business requirements in a heterogeneous network environment. Due to the differences in network performances and the diversity of user requirements, the available bandwidth, delay, jitter, packet loss rate, and cost may affect the selection of networks and should be considered, as they can be used to synthesize additional information and make comprehensive decisions.

Wu et al. [10] adopted a weight calculation method combining subjective and objective weights to select an access network. Zineb [11] combined fuzzy logic with a sorting method, where the input parameters and fuzzy reasoning rules were integrated for access network selection.

2.3. Access Network Selection Algorithm Based on a Utility Function. In economics, utility describes to consumers how to use their limited funds to obtain maximum satisfaction. A utility function can characterize the corresponding relationship between consumers' pay and benefits. The function value represents the degree of consumer satisfaction regarding the consumption process to a certain extent.

Park et al. [12] designed utility functions for network costs and users' QoS demands, respectively, so that users could access a network with a minimal cost while meeting their QoS demands. Roveri et al. [13] comprehensively considered the revenue balance between users and the network, and they classified the factors affecting network selection. The factors include the QoS, network priority, and weighting. Their model could be used to meet business requirements and improve network performance.

2.4. Access Network Selection Algorithm Based on Fuzzy Logic. In heterogeneous wireless networks, network parameters are ambiguous and cannot be used to accurately express network performances. At this time, users' judgements regarding these network parameters are also inaccurate. Fuzzy logic can be used to analyze a network by simulating human thinking. The fuzzy logic theory uses a "membership degree" to represent the element-set relations of the research object and analyzes the data that cannot be accurately represented.

Kustiawan [14] proposed a vertical switching algorithm based on fuzzification. This algorithm combined fuzzy logic with indicators, such as bandwidth and delay, to judge whether to switch. Guo [15] combined fuzzy logic with a neural network, and indicators such as the bandwidth and number of users were considered comprehensively to improve the communication quality of users.

In summary, several studies focused on terrestrial mobile networks. However, the LEO satellite and 5G networks are still in the early stages of construction, so studies on access network algorithms are rare. There are still considerable gaps in terms of the actual applications of a terrestrial-satellite network. Therefore, we proposed an access network selection algorithm in a terrestrial-satellite network, and the access network selection algorithm is studied to maximize the reliability of the network and guarantee the QoS of users.

3. Access Network Selection Algorithm for Terrestrial-Satellite Networks Based on a QoS Guarantee

3.1. Description of the Algorithm. When using a QoS-guaranteed algorithm for access network selection, how to normalize the attributes, how to determine the weights of

attributes, and how to combine the weighted attributes to obtain a comprehensive evaluation value are important components. A variety of mathematical models are commonly used to calculate the normalized methods, but they are based on a comparison between the candidate networks. Thus, these normalized values can only reflect the relative attribute values without considering the needs of businesses. Therefore, the subjective and objective weights should be calculated separately and then combined in a certain way to obtain the comprehensive weights.

According to the analysis above, this paper proposes a terrestrial-satellite access network selection algorithm based on a QoS guarantee. In the algorithm, the attributes that affect the choice of access network are first determined. A utility function is defined for each attribute, and the attribute values are normalized by establishing the corresponding relationships between the network attributes and business requirements. Then, the analytic hierarchy process (AHP) is used to calculate the subjective weights, and the entropy weight method is used to calculate the objective weights. The comprehensive weights are obtained by combining the subjective and objective weights through the comprehensive weighting method [16]. Finally, by weighting the normalized attribute values, the positive and negative ideal networks are constructed, and the relative closeness degrees between the candidate network and the positive and negative ideal networks are calculated to obtain the comprehensive evaluation value of the candidate network. The network with the highest relative closeness degree is selected as the best access network.

3.2. Network Attribute Normalization. Network attributes can be divided into benefit-oriented attributes and cost-oriented attributes. Among them, benefit-oriented attributes are attributes where higher values are better, such as the network bandwidth and received signal strength. Cost-oriented attributes are the opposite, such as the delay and packet loss rate.

In the algorithm proposed in this paper, to reflect the influences of terrestrial-satellite network performance on different service access network choices, four attributes, namely, the bandwidth, delay, packet loss rate, and cost, are selected as the deciding factors for access network selection. The roles of these attributes in remote transmission line monitoring scenarios are analyzed in detail below.

3.2.1. Bandwidth. Bandwidth is a huge bottleneck for current network applications in transmission line monitoring. In the future, line monitoring will include a large number of high-definition videos, images, and other services with large bandwidth requirements. Furthermore, the bandwidth resources allocated to individual users are also different to some extent.

3.2.2. Time Delay. Delay is an important part of the business demands, as it can reflect network connectivity. Time delay varies according to the type of service. In the transmission line monitoring service scenario, all services return data through the uplink. Due to the use of different return modes, the delays between base stations and the data center are quite different. Additionally, the processing delays of base stations also change dynamically according to different network loads.

3.2.3. Packet Loss Rate. The packet loss rate reflects the congestion degree of the network. When the load of a network is high, congestion occurs in the base station, and some data loss results in a decline of service performance. Among the various services in transmission line monitoring, real-time services, such as video services and voice services, have the highest requirements for the packet loss rate. In addition, the packet loss rate is also related to the return mode of the base station. Compared with a macro station, the packet loss rate of a small base station increases as the number of transmissions increases. Therefore, if the impact of packet loss caused by congestion is not considered, the packet loss rate also becomes an inherent attribute that can reflect different levels of network performance.

3.2.4. Cost. The cost reflects the use of different network resources. LEO satellite network resources are scarcer than those of a terrestrial network, so cost should be one of the deciding factors.

After determining the attributes that affect the access network selection decision, the terminal obtains the candidate network parameters. Assuming that there are m candidate networks at the mobile terminal and the number of attributes is n, the initial decision matrix X is as follows:

$$X = \left(x_{ij}\right)_{m\times n} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}. \tag{1}$$

Among the traditional normalization methods, the range transformation is most commonly used. According to the various attribute types, different calculation formulas are used to obtain the relative normalization value by comparing the attribute value of the current candidate network with the extreme value of the attribute from all candidate networks [17].

For the benefit attribute, the following equation exists:

$$b_{ij} = \frac{x_{ij} - \min_j x_{ij}}{\max_j x_{ij} - \min_j x_{ij}}. \tag{2}$$

For the cost attribute, the following equation exists:

$$b_{ij} = \frac{\max_j x_{ij} - x_{ij}}{\max_j x_{ij} - \min_j x_{ij}}, \tag{3}$$

where x_{ij} represents the parameter value of the jth indicator in the ith candidate network. $\max_j x_{ij}$ and $\min_j x_{ij}$ represent the maximum and minimum values of the jth candidate network, respectively. b_{ij} is the normalized value, which constitutes the normalized matrix $B = (b_{ij})_{n\times m}$.

Therefore, an attribute of the current network is characterized based on all candidate network attributes in the normalization process. The normalized attribute value of the largest network is 1, and the normalized attribute value of the smallest network is 0. When the performances of all candidate networks meet the business requirements, according to the normalization process above, the normalized value of the worst-performing network among the candidate networks is 0. At this time, it is most likely that a network with good performance is chosen, and this causes a waste of resources. Similarly, when the performances of the candidate networks are poor, the normalized value of the network with the best relative performance is 1. This causes the current service to occupy network resources but fail to meet its business requirements, resulting in an unreasonable allocation of resources.

To use as few network resources as possible to meet users' business needs, network performance should be linked to the users' business needs. Utility represents the users' satisfaction with the current service, and it can be expressed by using a utility function. By introducing a utility function to normalize the business attributes, the relationships between network performances and business requirements can be accurately reflected. For different attributes, user satisfaction with the network is different, so the utility function of each attribute is usually different. According to the characteristics of different attributes, scholars have conducted relevant studies and proposed various utility functions [18–20], as shown in Table 1.

When selecting an access network, different utility function models and model parameters are set according to the communication requirements of different services. In the following, to provide a basis for access network selection, utility functions are designed according to the characteristics of each attribute to replace the traditional normalization method, and the initial attribute value is normalized to a utility value that can accurately achieve business satisfaction.

(i) Bandwidth Utility Function. Bandwidth is a benefit attribute; that is, the bandwidth utility function is an increasing function. The greater the bandwidth is, the greater the utility. When the bandwidth provided by the network is lower than the minimum required bandwidth of the service, the normal communication of the service is seriously affected. When the bandwidth reaches the maximum required bandwidth of the service, the service can achieve the best communication, and further increase of the bandwidth does not improve the utility. The bandwidth utility function is designed to be an exponential type of function as follows:

$$U_B = \begin{cases} 0, & B \le B_{\min}, \\ 1 - e^{-\alpha B\left(B - B_{\min}\right)}, & B_{\min} \le B \le B_{\max}, \\ 1, & B > B_{\max}, \end{cases} \tag{4}$$

Table 1: Comparison table of various attribute and utility function types.

Attribute	Utility function type
Bandwidth	Exponential type, linear type, S type
Time delay	S Type
Packet loss probability	S type, exponential type
Cost	Linear type

where B is the bandwidth of the current network, $B_{\min}$ is the minimum bandwidth required by the service, $B_{\max}$ is the maximum bandwidth required by the service, and α_B is the shape parameter, which determines the increase rate of the bandwidth utility function. The larger α_B is, the higher the rising rate will be.

(ii) Delay Utility Function. The delay is a cost-type attribute; that is, the utility function is a decreasing function. When the delay is larger, the utility is smaller. For each business, there is a minimum delay required by the business and maximum delay tolerance. When the delay of the network is higher than the maximum tolerance, the network cannot provide access services for the business. When the delay is lower than the minimum delay required by the business, the utility does not change. The delay utility function is designed as an S-type function and is shown as follows:

$$U_D = 1 - \frac{1}{1 + e^{-\alpha_D\left(D - \beta_D\right)}}, \tag{5}$$

where D is the delay of the current network, $D_{\min}$ is the minimum delay required by the business, and $D_{\max}$ is the maximum delay tolerated by the business. α_D is the shape parameter that determines the rate of decline of the delay utility function. The larger α_D is, the higher the rate of decline will be.

(iii) Packet Loss Rate Utility Function. The packet loss rate can reflect the reliability of data transmission, and it is an important index of QoS. When the network packet loss rate is too high, data transmission errors are induced. Both the packet loss rate and delay are cost-type attributes, and their utility function models are the same:

$$U_L = 1 - \frac{1}{1 + e^{-\alpha_L\left(L - L_{\min}\right)}}, \tag{6}$$

where L is the packet loss rate of the current network, $L_{\min}$ is the minimum packet loss rate required by the service, and $L_{\max}$ is the maximum packet loss rate that the service can tolerate. α_L is the shape parameter for determining the drop rate of the utility function. The larger α_L is, the higher the drop rate will be.

(iv) Cost Utility Function. Cost is a cost-type attribute. The higher the access network cost is, the smaller its utility will be. Its utility function is designed as a linear function, as shown in

$$U_C = \begin{cases} 1, & C < C_{\min}, \\ 0.8 + 0.1 \times \dfrac{C_{\max} - C}{C_{\max} + C_{\min}}, & C_{\min} \le C < C_{\max}, \\ 0, & C \ge C_{\max}. \end{cases} \quad (7)$$

According to the business requirements of each terminal, the utility function of each selected attribute is calculated. The utility value on the interval [0, 1] is obtained to be as the normalized values of the initial decision values. The normalized decision matrix is shown as follows:

$$B = \left(b_{ij}\right)_{m\times n} = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}. \quad (8)$$

3.3. Comprehensive Weight Calculation

3.3.1. Subjective Weight Calculation Based on the Analytic Hierarchy Process. The analytic hierarchy process compares and judges the importance of attributes through the subjective judgement of humans and quantifies it. It is suitable for solving the problem that it is difficult to construct a mathematical model for a complete quantitative analysis, and it can be applied in the weight calculation for access network selection.

After the establishment of the hierarchical structure model, the relative importance between two attributes is obtained by comparing the importance of the two attributes, and a decision matrix is then constructed. The criterion layer contains n attributes, and the decision matrix $A = (a_{ij})_{n\times n}$ can be obtained by comparing two attributes. a_{ij} is the importance degree of attribute i compared with that of attribute j; it satisfies $a_{ij} = 1/a_{ij}$ and $a_{ij} = 1$. The subjective decision is quantified to obtain the value of a_{ij} by determining which attribute is more important.

Using (9), we calculate the geometric mean of the jth column.

$$o_j = \sqrt[n]{\prod_{k=1}^{n} a_{jk}}, \quad j = 1, 2, \ldots, n. \quad (9)$$

The weight value of the decision index is calculated from a subjective point of view, and the weight is calculated as shown in

$$sw_j = \frac{O_j}{\sum_{k=1}^{n} O_k}, \quad j = 1, 2, \ldots, n. \quad (10)$$

Due to subjective bias, decision-makers' judgements regarding attributes may be contradictory, so it is necessary to conduct a consistency check on the decision matrix to ensure that the judgements on the importance of attributes are reasonable [16].

3.3.2. Objective Weight Calculation Based on the Entropy Weight Method. The entropy weight method is an objective weight calculation method that calculates the information entropy of each attribute by observing the amplitude variation of each attribute in different candidate networks. The larger the range of variation an attribute has in each candidate network, the smaller the information entropy is, the greater the amount of information provided for the decision is, and the greater the corresponding weight should be [20]. The main steps of calculating objective weight by entropy weight method are as follows:

The proportion of the attribute value for the jth index in the ith candidate network p_{ij} is calculated:

$$p_{ij} = \frac{b_{ij}}{\sum_{i=1}^{m} b_{ij}}. \quad (11)$$

The entropy value E_j of the jth index in different candidate networks is calculated:

$$E_j = -\frac{1}{\ln m} * \sum_{i=1}^{m} p_{ij} \ln p_{ij}. \quad (12)$$

The objective weight ow_j of the jth index is calculated:

$$ow_j = \frac{1 - E_j}{\sum_{i=1}^{m} 1 - E_j}. \quad (13)$$

The objective weights calculated by the entropy weight method can highlight the attributes that change significantly in different candidate networks and weaken the attributes that change little. This can provide useful information for objective decision making.

3.3.3. Comprehensive Weight Calculation. Subjective weight sw_j and objective weight ow_j are obtained by the AHP and entropy weight method, respectively. The deviation minimization method is used to calculate the comprehensive weights.

α is used to represent the weight coefficient of the subjective weight vector, and β is used to represent the weight coefficient of the objective weight vector. Combined with moment estimation theory, α and β can be calculated as

$$\begin{cases} \alpha_j = \dfrac{sw_j}{sw_j + ow_j}, \\ \\ \beta_j = \dfrac{ow_j}{sw_j + ow_j}, \end{cases} \quad j = 1, 2, \ldots, n. \quad (14)$$

The weight coefficients of the subjective and objective weight vectors are used for weighted calculations, and the comprehensive weight W is calculated as

$$w_j = \frac{a_j ow_j + \beta_j sw_j}{\sum_{j=1}^{n} a_j ow_j + \beta_j sw_j}, \quad j = 1, 2, \ldots, n. \quad (15)$$

3.3.4. Selecting the Best Candidate Network. After the comprehensive weight set $W = \{w_1, w_2, \ldots, w_n\}$ is calculated in (15), it is multiplied by the normalized decision matrix in (8). The weighted normalized matrix can be obtained:

$$Z = \left(z_{ij}\right)_{m\times n} * W^T = \begin{pmatrix} w_1 b_{11} & w_2 b_{12} & \cdots & w_n b_{1n} \\ w_1 b_{21} & w_2 b_{22} & \cdots & w_n b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_1 b_{m1} & w_2 b_{m2} & \cdots & w_n b_{mn} \end{pmatrix}. \tag{16}$$

The candidate network is sorted by an approximate ideal sorting algorithm. The main steps are as follows:

According to the weighted normalized matrix, the positive ideal network Z^+ and the negative ideal network Z^- can be determined, where the positive ideal network can be expressed as

$$Z^+ = \left(z_1^+, z_2^+, \ldots, z_n^+\right). \tag{17}$$

The negative ideal network can be expressed as

$$Z^- = \left(z_1^-, z_2^-, \ldots, z_n^-\right). \tag{18}$$

The distance between the candidate network and the positive ideal network can be calculated as follows:

$$L_i^+ = \sqrt{\sum_{j=1}^{n} \left(z_{ij} - z_j^+\right)^2}, \quad i = 1, 2, \ldots, m. \tag{19}$$

The distance between the candidate network and the negative ideal network is calculated as follows:

$$L_i^- = \sqrt{\sum_{j=1}^{n} \left(z_{ij} - z_j^-\right)^2}, \quad i = 1, 2, \ldots, m. \tag{20}$$

The relative closeness between the candidate network and the ideal network can be calculated as follows:

$$C_{\text{TOPSIS}} = \frac{L_i^-}{L_i^+ + L_i^-}, \quad i = 1, 2, \ldots, m. \tag{21}$$

The candidate networks are sorted according to the degrees of relative closeness. The greater the degree of relative closeness is, the better the network will be. Then, the best access network is selected [21].

4. Simulation and Results

4.1. Simulation Scenario. The simulation scenario is shown in Figure 1. It consists of small conventional terrestrial stations and other base stations. The macro station is located in the center of the scene, and small conventional terrestrial base stations are randomly distributed within the coverage range of the macro station. The monitoring terminals are distributed in the overlapping coverage of the network. Three kinds of networks can provide communication transmission line monitoring terminal access. The simulation is mainly aimed at the overlapping coverage area of the networks. Since the network environments are dynamically changing, each attribute parameter in the simulation is randomly generated within the predefined interval. As the number of service access networks increases, the available load of each network decreases, and the delay and packet loss rate of each network also start to increase; that is, the network parameters change with the decrease of the available load.

In the simulation scenario shown in Figure 1, each monitoring terminal completes access control and executes the access network selection algorithm based on a QoS guarantee. Each monitoring terminal evaluates the networks and selects the best candidate access network by collecting network attribute information. Therefore, the access network selection algorithm based on a QoS guarantee is implemented by the distributed computing of each monitoring terminal. Each monitoring terminal independently completes the algorithm process, while the different types of networks in the terrestrial-satellite network only provide network access in the scenario; they do not participate in access control, and there are no algorithmic calculations.

Due to the particularity of the transmission line monitoring business, the business models that are considered in the scenario are uplink communications. The simulation is mainly aimed at the overlap area of the network. Considering the maximum reliability of the networks, it is assumed that all businesses retain their communication connections. Businesses include videos, images, data, and voice clips. For these four types of businesses, the number of business monitoring terminals is randomly generated. Because the network environment is dynamically changing, the parameters of each decision index in the simulation are generated randomly during the interval at a certain moment.

MATLAB is used to build the test scenarios of terrestrial-satellite networks. The proposed access algorithm and comparison algorithm are, respectively, run 10 times; the access network selection and data processing results are recorded. The average results of the two algorithms are taken for comparison. Because the experiments regarding the initial attribute and terminal business types of the scenario are randomly generated, to ensure that the scenes are the same, different random seed numbers are set for the experiment in MATLAB. When the random seed number remains unchanged, the random numbers generated by the program remain the same, and the execution of each experimental algorithm scenario is the same.

According to the preferences of a given service regarding the bandwidth, delay, packet loss rate, and cost, the comparison judgement matrix is obtained through a pair-to-pair comparison of the importance values. The comparison judgement matrix is verified by consistency, and the subjective weights calculated by the AHP are shown in Figure 2.

4.2. Analysis and Results. To test the performance of the algorithm proposed in this paper under the background of the transmission line monitoring application in remote areas, MATLAB is used to simulate the algorithm. In the access network selection scenario for multi-transmission

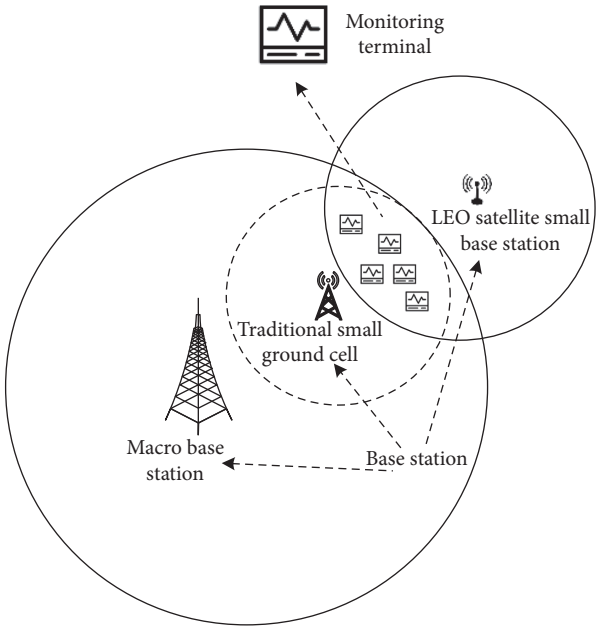


Figure 1: Schematic diagram of the simulation scenario.

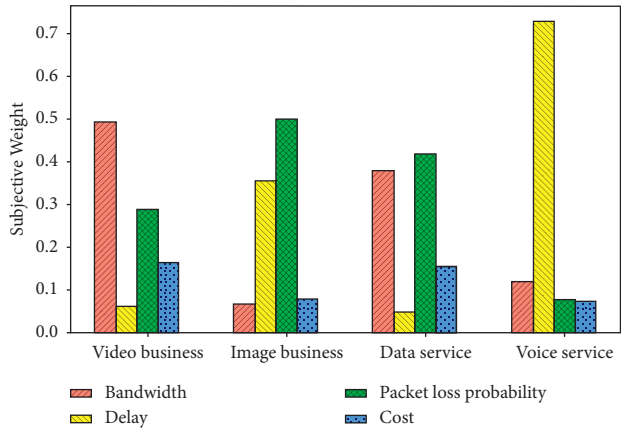


Figure 2: The subjective weight distributions of different businesses.

line monitoring services, the access network selection algorithm proposed in this paper based on a QoS guarantee and the access network selection algorithm based on comprehensive weights in the literature [10] are used to evaluate the candidate networks and select the best candidate access network. The performances of the algorithms are compared from three aspects: the traffic blocking rate, network selection rate, and network resource utilization.

4.2.1. Traffic Blocking Rate Analysis. The blocking rates of the two algorithms are shown in Figure 3. When the terminals use the access network selection algorithm based on comprehensive weight, most of the monitoring terminals give priority to the macro station based on the performance comparison between the networks. Due to the limited number of connected terminals in a macro station, the traffic blocking rate increases. When the terminal uses the access

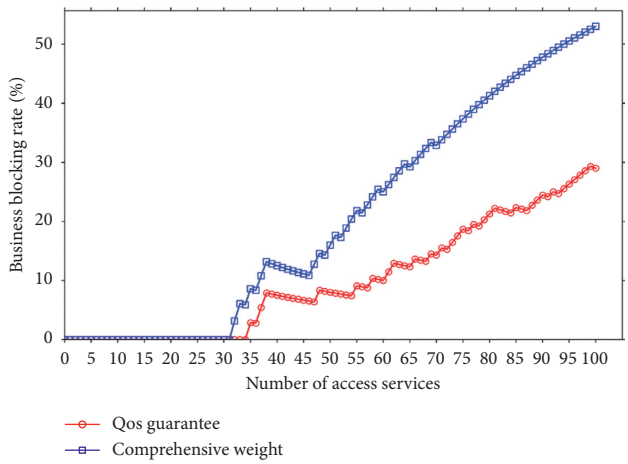


Figure 3: Comparison of the traffic blocking rates of two algorithms.

network selection algorithm based on a QoS guarantee, the utility function replaces the traditional range transformation method for normalizing the network parameters. Rather than blindly choosing the network with the best performance, the evaluation value of the network is closer to the service requirements, so the traffic blocking rate is lower.

4.2.2. Network Selection Rate Analysis. The network selection rates are shown in Figures 4 and 5. It is assumed that each network is in a no-load state at the beginning, and the performance of each candidate network remains relatively good. The goal of the business is to choose the network with the best performance, but as the number of businesses increases, the available loads of access networks gradually decrease, and other networks start to be chosen. This finally balances the selection rates of all networks. With the use of the access network selection algorithm based on the comprehensive weight, the changes in network selection rate are more dramatic. If the number of access businesses is small, the network with better performance is preferred. As the network performance degrades, businesses gradually begin to choose other networks. By using the QoS guarantee-based access network selection algorithm, when the number of businesses is small, each network has a probability of being selected, and the selection rate is flat.

4.2.3. Network Resource Utilization Analysis. The utilization rate of network resources changes as shown in Figure 6. When using the access network selection algorithm based on the comprehensive weight, because the business does not evaluate the candidate networks in combination with business demand, it prefers to choose a network with better performance. As a result, the number of accessible businesses on this network will soon reach the upper limit. By considering the needs of the businesses, the access network selection algorithm proposed in this paper with a QoS guarantee chooses the most appropriate network. The simulation results show that the algorithm improves the network resource utilization of the whole system to a certain extent.

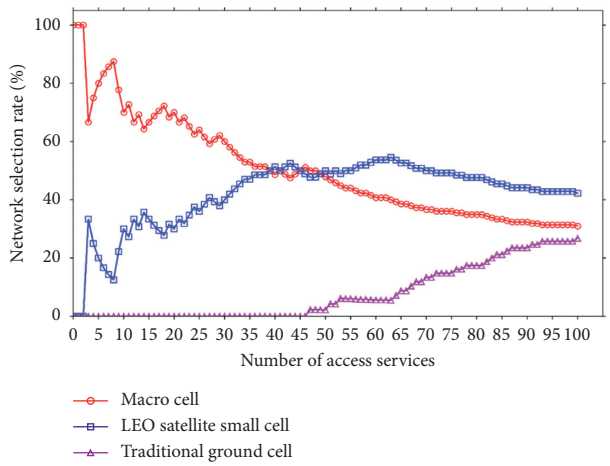


Figure 4: Network selection rate of access network selection algorithm based on service QoS guarantee.

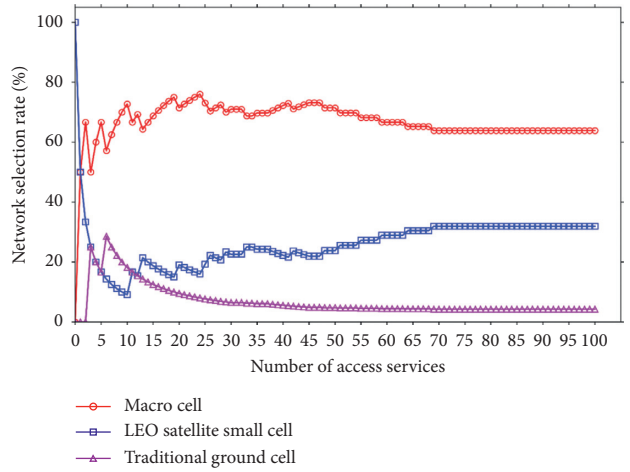


Figure 5: Network selection rate of the access network selection algorithm based on comprehensive weight.

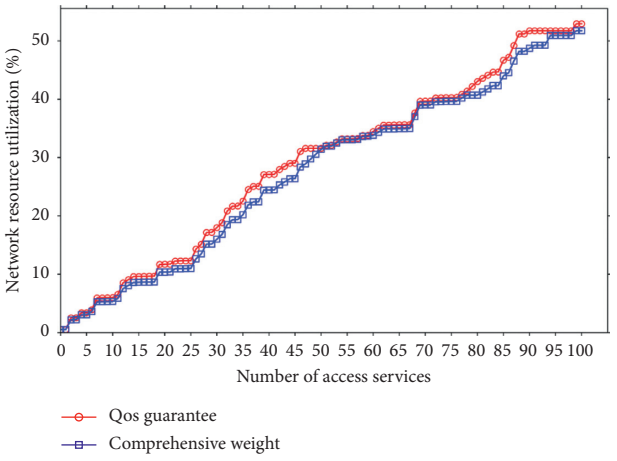


Figure 6: Comparison of network resource utilization between the two algorithms.

5. Conclusion

Considering different business requirements and network performances, this paper proposes an access network selection algorithm based on a QoS guarantee for terrestrial-satellite networks in remote areas. According to the characteristics of the network attributes, utility functions are designed to normalize the attributes. A sorting method is used to sort the candidate networks and complete the access network selection from the best candidate networks. By establishing the association between network attributes and business requirements, this algorithm can accurately reflect the roles of business requirements in candidate networks and realize the accurate evaluation of candidate network performances. Through a comparison with the access network selection algorithm based on the comprehensive weight, the proposed algorithm can reduce the blocking rate while improving the load balance and resource utilization. It not only guarantees the demand of the business but also avoids the waste of network resources. Therefore, it can provide highly reliable network access for the existing monitoring terminals under the scenario of limited network resources.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Equipment Development Department under grant 41412010702, the Major R&D Platform Project of New R&D Institutions in Zhongshan City under grant 2017F1FC0001, the Special Funds for Innovation in Scientific Research Program of Zhongshan under grant 181129112748101, the Special Fund for Science and Technology of Guangdong Province under grant 2019SDR002, and the Key Project of Earth Observation and Navigation under grant 2017YFB0504100.

References

[1] H. Wang, Z. Zhuang, and Y. Q. Zhong, "Research of the communication technologies for transmission line monitoring," *Journal of Physics Conference Series*, vol. 1176, no. 6, Article ID 62012, 2019.

[2] M. Marchese, A. Moheddine, and F. Patrone, "IoT and UAV integration in 5G hybrid terrestrial-satellite networks," *Sensors (Basel, Switzerland)*, vol. 19, no. 17, 2019.

[3] X. W. Zhu and Q. Zhu, "MADM and rank similarity-based network selection algorithm in heterogeneous networks," *Applied Mechanics and Materials*, vol. 605, pp. 3342–3346, 2014.

[4] B. Di, H. Zhang, L. Song, Y. Li, and G. Y. Li, "Ultra-dense LEO: integrating terrestrial-satellite networks into 5G and beyond for data offloading," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 47–62, 2019.

[5] L. Yang and D. Guo, "A satellite heterogeneous network selection algorithm combining AHP and TOPSIS," in *Proceedings of the 2012 National Conference on Information Technology and Computer Science*, Beijing, China, December 2012.

[6] Y. Gwon, D. Funato, and A. Takeshita, "Adaptive approach for locally optimized IP handoffs across heterogeneous wireless networks," in *Proceedings of the 4th International Workshop on Mobile and Wireless Communications Network*, pp. 475–479, Stockholm, Sweden, September 2002.

[7] X. Yan, N. Mani, and Y. A. Sekercioglu, "A traveling distance prediction based method to minimize unnecessary handovers from cellular networks to WLANs," *IEEE Communications Letters*, vol. 12, no. 1, pp. 14–16, 2008.

[8] Q. Y. Song and A. Jamalipour, "A network selection mechanism for next generation networks," *IEEE International Conference on Communications*, vol. 12, pp. 1418–1422, 2005.

[9] S. Kunarak and R. Suleesathira, *Predictive RSS with Fuzzy Logic Based Vertical Handoff Decision Scheme for Seamless Ubiquitous Access*, King Mongkut's University of Technology Thonburi Thailand, Bangkok, Thailand, 2011.

[10] H. B. Wu, Y. F. Mao, W. J. Wu, and W. Zhang, "A network selection algorithm based on combined weights," *IOP Conference Series: Materials Science and Engineering*, vol. 466, no. 1, Article ID 12072, 2018.

[11] A. B. Zineb, M. Ayadi, and S. Tabbane, "Fuzzy MADM based vertical handover algorithm for enhancing network performances," in *Proceedings of the 2015 23rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp. 153–159, Split-Bol, Croatia, September 2015.

[12] H. S. Park, S. H. Yoon, T. H. Kim, J. S. Park, and J. Y. Lee, "Vertical handoff procedure and algorithm between IEEE802. 11 WLAN and CDMA cellular network," in *Proceedings of the CDMA International Conference*, pp. 103–112, Seoul, Korea, October 2002.

[13] A. Roveri, C.-F. Chiasserini, M. Femminella et al., "The RAMON module: architecture framework and performance results, Quality of Service in Multiservice IP Networks," in *in Proceedings of the International Workshop on Quality of Service in Multiservice IP Networks*, pp. 471–484, Milano, Italy, February 2003.

[14] I. Kustiawan, C.-Y. Liu, and D. F. Hsu, "Vertical handoff decision using fuzzification and combinatorial fusion," *IEEE Communications Letters*, vol. 21, no. 9, pp. 2089–2092, 2017.

[15] Q. Guo, J. Zhu, and X. H. Xu, "An adaptive multi-criteria vertical handoff decision algorithm for radio heterogeneous network," *IEEE International Conference on Communications*, vol. 4, pp. 2769–2773, 2005.

[16] M. Xiang, Q. Q. Qu, C. Chen, L. Tian, and L. Zeng, "Access selection algorithm of heterogeneous wireless networks for smart distribution grid based on entropy-weight and rough set," *International Journal of Emerging Electric Power Systems*, vol. 18, no. 6, 2017.

[17] I. Chamodrakas and D. Martakos, "A utility-based fuzzy TOPSIS method for energy efficient network selection in heterogeneous wireless networks," *Applied Soft Computing*, vol. 11, no. 4, pp. 3734–3743, 2011.

[18] D. Jiang, L. Huo, Z. Lv, H. Song, and W. Qin, "A joint multi-criteria utility-based network selection approach for vehicle-to-infrastructure networking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 10, pp. 3305–3319, 2018.

[19] R. Trestian, O. Ormond, and G.-M. Muntean, "Enhanced power-friendly access network selection strategy for multimedia delivery over heterogeneous wireless networks," *IEEE Transactions on Broadcasting*, vol. 60, no. 1, pp. 85–101, 2014.

[20] H. W. Yu and J. P. Yin, "Wireless network selection algorithm based on independence of influence factors of QoS," in *Proceedings of the 7th International Conference on Computing Communication and Networking Technologies*, pp. 1–4, Dallas TX USA, July 2016.

[21] L. Tang, S. Ji, and J. Yan, "A heterogeneous network access selection algorithm based on attribute dependence," *Wireless Personal Communications*, vol. 92, no. 3, pp. 1163–1176, 2017.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5577777, 9 pages
https://doi.org/10.1155/2021/5577777

Research Article

Global Existence and Extinction Singularity for a Fast Diffusive Polytropic Filtration Equation with Variable Coefficient

Dengming Liu and Changyu Liu

School of Mathematics and Computational Science, Hunan University of Science and Technology, Xiangtan, Hunan 411201, China

Correspondence should be addressed to Dengming Liu; liudengming@hnust.edu.cn

Received 2 February 2021; Revised 4 March 2021; Accepted 1 April 2021; Published 28 April 2021

Academic Editor: Hou-Sheng Su

In this article, we deal with an inhomogeneous fast diffusive polytropic filtration equation. By using the energy estimate approach, Hardy–Littlewood–Sobolev inequality, and a series of ordinary differential inequalities, we prove the global existence result and obtain the conditions on the occurrence of the extinction phenomenon of the weak solution.

1. Introduction

Our main objectives in this article are to deal with the global existence and the extinction phenomenon of the inhomogeneous fast diffusive polytropic filtration equation:

$$\begin{cases} |x|^{-s}u_t - \operatorname{div}\left(\left|\nabla u^m\right|^{p-2}\nabla u^m\right) = u^q, & (x,t) \in \Omega \times (0,+\infty), \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times (0,+\infty), \\ u(x,0) = u_0(x), & x \in \Omega, \end{cases} \tag{1}$$

where $\Omega \subset \mathbb{R}^N$ $(N > p)$ is a bounded domain with smooth boundary $\partial\Omega$, $x = (x_1, \ldots, x_N) \in \Omega$, $|x| = \sqrt{x_1^2 + \cdots + x_N^2}$, $u_0(x)$ is a nonnegative and bounded function with $u_0^m \in W_0^{1,p}(\Omega)$, and the parameters m, s, p, and q satisfy

$$\begin{aligned} &0 < m \le 1, \\ &0 < m(p-1) < 1, \\ &\frac{1-m}{2} < q \le 1, \\ &0 \le s < \frac{Nq}{q+m}. \end{aligned} \tag{2}$$

Inhomogeneous parabolic problems arise in a wide range of physical contexts (see for instance [1–3] and the references therein, where a more detailed physical background can be found). Problem (1) can be used to describe the compressible fluid flows in a homogeneous isotropic rigid porous medium with $u(x,t)$ being the density of the fluid and $\alpha(x) = |x|^{-s}$ acting as the volumetric moisture content. On the other hand parabolic models like (1), together with differential equation models, stochastic differential equations, and linear systems, are regarded as the powerful tools to solve lots of problems from control engineering, image processing, and other areas (see [4–8]). Because of the degeneracy and the singularity, problem (1) might not have classical solution in general, and hence, we introduce definition of the weak solution as follows.

Definition 1. By a local weak solution to problem (1), we understand a function $u \in \mathfrak{S} \overset{\text{def}}{=} \{u \in C(0,T; L^1(\Omega)), u \in L^{2q}(\Omega \times (0,T)) \cap L^2(\Omega \times (0,T)), \nabla u^m \in L^p(\Omega \times (0,T))\}$ for some $T > 0$, which moreover satisfies the following assumptions:

(i) For any $0 \le \phi \in \widetilde{\mathfrak{S}} \overset{\text{def}}{=} \{\phi \in \mathfrak{S}, \phi|_{\partial\Omega} = 0\}$ and $0 < t_1 < t_2 < T$, one has

$$\int_\Omega |x|^{-s} u_t \phi \mathrm{d}x + \int_\Omega \left|\nabla u^m\right|^{p-2} \nabla u^m \cdot \nabla \phi \mathrm{d}x = \int_\Omega u^q \phi \mathrm{d}x. \tag{3}$$

(ii) $u(x,t) \longrightarrow u_0(x)$ as $t \longrightarrow 0$ with convergence in $L^1(\Omega)$.

In the past few decades, many mathematicians have studied the global existence, blow-up, and extinction phenomena of the following parabolic equation:

$$|x|^{-s} u_t - \operatorname{div}\left(\left|\nabla u^m\right|^{p-2} \nabla u^m\right) = f(u), \quad (x,t) \in \Omega \times (0,+\infty), \tag{4}$$

subject to various assumptions (see [9–23] and the references therein). For the case $s = 0$, the authors in [24–26] concerned with the global existence and blow-up properties of the solutions to problem (4) with $m = 1$ and $p \geq 2$. Yuan et al. [27] considered problem (4) with $m = 1$ and $f(u) = 0$ and showed that the solution of problem (4) vanishes in finite time if and only if $1 < p < 2$. Gu [28] studied problem (4) with $m = 1$ and $f(u) = -u^q$ and claimed that the necessary and sufficient condition on the occurrence of extinction phenomenon is $p \in (1,2)$ or $q \in (0,1)$. Tian and Mu [29] and Jin and Yin [30] studied problem (4) with $m = 1$ and $f(u) = \lambda u^q$ with $\lambda > 0$ and showed that $q = p - 1$ is the critical extinction exponent of the solutions. When $s = 0$ and $f(u) = \lambda u^q$ with $\lambda > 0$, Jin et al. [31] and Zhou and Mu [32] concluded that the critical extinction exponent of the solution to problem (1) is $q = m(p-1)$. Compared with $s = 0$, there are few literatures for the case $s > 0$. By Hardy inequality and potential well method, Tan [33] obtained the global existence and blow-up results of problem (4) with $s = 2$ and $m = 1$. Wang [34] generated the results in [33] to the case $0 < s \leq 2$. Zhou generated the results in [33] to the case $s \geq 2$ and gave the global existence and blow-up results for (4) with $p = 2$ and $m = 1$ in [35, 36], respectively. To the best knowledge of us, there is little work on the global existence and extinction behavior of problem (1). In a recent paper, Deng and Zhou [37] considered the special case $m = 1$ and analysed the effect of the singular potential on the global existence and extinction behavior of the solutions.

In order to state well our results, we first introduce some definitions, fundamental facts, and useful symbols. Since Ω is a bounded domain in $\mathbb{R}^N$, then there is a ball $B(0,R) \subset \mathbb{R}^N$ centered at 0 with radius

$$R = \sup_{x \in \Omega} \sqrt{x_1^2 + \cdots + x_N^2}, \tag{5}$$

such that $\Omega \subseteq B(0,R)$.

We denote the norm of $L^r(\Omega)$ by $\|\cdot\|_r$ and the norm of $W^{1,r}(\Omega)$ by $\|\cdot\|_{W^{1,r}(\Omega)}$; that is, for any $\phi \in L^r(\Omega)$,

$$\|\phi\|_r = \begin{cases} \left(\int_\Omega |\phi(x)|^r \mathrm{d}x\right)^{1/r} & \text{if } 1 \leq r < +\infty, \\ \operatorname{ess\,sup}_{x \in \Omega} |\phi(x)|, & \text{if } r = +\infty, \end{cases} \tag{6}$$

and for any $\phi \in W^{1,r}(\Omega)$, $\|\phi\|_{W^{1,r}(\Omega)} = \sqrt{[r]\|\phi\|_r^r + \|\nabla\phi\|_r^r}$. According to Poincaré's inequality, one can see that $\|\nabla(\cdot)\|_r$ is equivalent to $\|\cdot\|_{W^{1,r}(\Omega)}$ in $W_0^{1,r}(\Omega)$, and hence, we equip $W_0^{1,r}(\Omega)$ with the norm $\|\phi\| \overset{\text{def}}{=} \|\nabla\phi\|_r$.

Let $u(x,t)$ be a weak solution of problem (1). Define an energy functional as the following form:

$$E(u) = \frac{1}{p}\int_\Omega \left|\nabla u^m\right|^p \mathrm{d}x - \frac{m}{m+q}\int_\Omega u^{m+q} \mathrm{d}x. \tag{7}$$

Then, by (3), one can easily show that

$$\frac{\partial E(u)}{\partial t} = -m\int_\Omega u^{m-1}(u_t)^2 \mathrm{d}x \leq 0, \tag{8}$$

which tells us that $E(u)$ is nonincreasing with respect to t.

We state our main results as follows.

Theorem 1. *Suppose that the parameters m, p, q, and s satisfy (2), and the initial data $u_0(x)$ are a nonnegative and bounded function with $u_0^m \in W_0^{1,p}(\Omega)$. Let $u(x,t)$ be a solution of problem (1). Then, the maximal existence time of $u(x,t)$ is $T = +\infty$; that is, $u(x,t)$ is a global solution. Moreover,*

(i) If $0 < \max\{m(p-1), (1-m)/2\} < q \leq 1$ and $0 \leq s \leq \min\{p, Nq/(m+q)\}$ and there is a constant $a > \max\{(m+1-\alpha m)/(m\alpha - mp), -(1/p)\}$ with $\alpha = p(N-s)/(N-p)$ such that

$$\left(\int_\Omega |x|^{-s} u_0^{m(pa+1)+1} \mathrm{d}x\right)^{(q-m(p-1))/(m(pa+1)+1)} \leq \frac{pa+1}{2(a+1)^p} R^{-(s[m(pa+1)+q])/(m(pa+1)+1)} |\Omega|^{(q-1)/(m(pa+1)+1)} \kappa_1^{((N-p)/(N-s))-(mp(a+1)/(m(pa+1)+1))} \kappa_2^{-((N-p)/(N-s))}, \tag{9}$$

where R, κ_1, and κ_2 are given by (5), (12), and Lemma 2, respectively, then the solution $u(x,t)$ of problem (1) vanishes in finite time.

(ii) If $0 < ((1-m)/2) < q \leq m(p-1) < 1$ and

$$\begin{cases} \int_\Omega |x|^{-s} u_0^{m+1} \mathrm{d}x > 0 \text{ and } E(u_0) \leq 0, & \text{when } q = m(p-1), \\ \int_\Omega |x|^{-s} u_0^{m+1} \mathrm{d}x > 0 \text{ and } E(u_0) < 0, & \text{when } q < m(p-1), \end{cases} \tag{10}$$

where

$$E(u_0) = \frac{1}{p}\int_\Omega \left|\nabla u_0^m\right|^p \mathrm{d}x - \frac{m}{m+q}\int_\Omega u_0^{m+q}\mathrm{d}x, \tag{11}$$

then the solution $u(x,t)$ of problem (1) does not possess extinction phenomenon.

The rest of this article is organized as follows. In Section 2, we collect some useful auxiliary lemmas. The last section is mainly focused on the global existence and the conditions on the occurrence of the extinction phenomenon of the solution. By Hardy–Littlewood–Sobolev inequality and some ordinary differential inequalities, the proof of Theorem 1 will be given in Section 3.

2. Preliminaries

In this section, as preliminaries, we collect some well-known results, which play an important role in our proof of Theorem 1.

Lemma 1 (see [37]). *Suppose $N > s$ and $\Omega \subset \mathbb{R}^N$ is a bounded domain. Then, we have*

$$\kappa_1 \overset{\text{def}}{=} \int_\Omega |x|^{-s}\mathrm{d}x \le \int_{B(0,R)} |x|^{-s}\mathrm{d}x = \int_0^R \left[\int_{\partial B(0,r)} |x|^{-s}\mathrm{d}S(x)\right]\mathrm{d}r$$
$$= \omega_N \int_0^R r^{-s} r^{N-1}\mathrm{d}r = \frac{\omega_N}{N-s}R^{N-s} < +\infty, \tag{12}$$

where $B(0,R)$ is the ball in $\mathbb{R}^N$ centered at 0 with radius $R = \sup_{x\in\Omega}\sqrt{x_1^2 + \cdots + x_N^2}$ satisfying $\Omega \subset B(0,R)$ and

$$\omega_N = \frac{N\pi^{N/2}}{\Gamma((N/2)+1)} \tag{13}$$

denotes the surface area of the unit sphere $\partial B(0,1)$, and Γ is the usual Gamma function.

Lemma 2 (see [38]). *Suppose $N \ge 2$, $1 < \mu < N$, $0 \le \vartheta \le \mu$, and $\sigma = \mu(N-\vartheta)/(N-\mu)$. Then, there is a positive constant $\kappa_2 = \kappa_2(\mu, \vartheta, N)$ such that*

$$\int_\Omega \frac{|u(x)|^\sigma}{|x|^\vartheta}\mathrm{d}x \le \kappa_2\left(\int_\Omega |\nabla u|^\mu \mathrm{d}x\right)^{(N-\vartheta)/(N-\mu)} \tag{14}$$

holds for any $u \in W_0^{1,\mu}(\Omega)$, where $\Omega \subset \mathbb{R}^N$ is a bounded domain.

Lemma 3 (see [39]). *Assume θ, δ, and β are positive constants. Let $y(t)$ be a nonnegative absolutely continuous function satisfying*

$$\frac{\mathrm{d}y}{\mathrm{d}t} + \delta y^\theta(t) \ge \beta, \quad t > 0. \tag{15}$$

Then, we have

$$y(t) \ge \min\left\{y(0), \left(\frac{\beta}{\delta}\right)^{1/\theta}\right\}. \tag{16}$$

Lemma 4 (see [40]). *Suppose $0 < k < r \le 1$. Let $y(t)$ be the solution of the ordinary differential inequality:*

$$\begin{cases} \dfrac{\mathrm{d}y}{\mathrm{d}t} + Cy^k \le \gamma y^r, \quad t > 0, \\ \\ y(0) = y_0 > 0, \end{cases} \tag{17}$$

where $C > 0$ and $0 < \gamma < (Cy_0^{k-r}/2)$. Then, there are two positive constants η and ξ such that, for $t \ge 0$,

$$0 \le y(t) \le \xi e^{-\eta t}. \tag{18}$$

3. Proof of Theorem 1

In this section, we will give the proof of the global existence result and the conditions on the occurrence of the extinction phenomenon of the solution $u(x,t)$.

Case 1. If $((1-m)/2) < q < 1$. Taking the test function $\phi = u^m(x,t)$ in (3), and using Hölder's inequality, one has

$$\frac{1}{m+1}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega |x|^{-s}u^{m+1}\mathrm{d}x + \left\|u^m\right\|^p = \int_\Omega u^{q+m}\mathrm{d}x = \int_\Omega |x|^{-(s(q+m)/(1+m))}u^{q+m}|x|^{(s(q+m)/(1+m))}\mathrm{d}x$$
$$\le \left(\int_\Omega |x|^{-s}u^{m+1}\mathrm{d}x\right)^{(m+q)/(m+1)}\left(\int_\Omega |x|^{s(m+q)/(1-q)}\mathrm{d}x\right)^{(1-q)/(m+1)} \tag{19}$$
$$\le \kappa_3\left(\int_\Omega |x|^{-s}u^{m+1}\mathrm{d}x\right)^{(m+q)/(m+1)},$$

which implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega |x|^{-s}u^{m+1}\mathrm{d}x \le \kappa_3 (m+1)\left(\int_\Omega |x|^{-s}u^{m+1}\mathrm{d}x\right)^{(m+q)/(m+1)}, \tag{20}$$

where

$$\kappa_3 = \left(\int_{B(0,R)} |x|^{s(q+m)/(1-q)}\mathrm{d}x\right)^{(1-q)/(1+m)} = \left(\frac{\omega_N (1-q)}{s(q+m)+N(1-q)} R^{(s(q+m)+N(1-q))/(1-q)}\right)^{(1-q)/(1+m)}. \tag{21}$$

Integrating (20) from 0 to t, one gets

$$\int_\Omega |x|^{-s}u^{m+1}\mathrm{d}x \le \left[\kappa_3 (1-q)t + \left(\int_\Omega |x|^{-s}u_0^{m+1}\mathrm{d}x\right)^{(1-q)/(m+1)}\right]^{(m+1)/(1-q)}. \tag{22}$$

From (19) and (22), it follows that

$$\|u\|_{q+m}^{q+m} \le \kappa_3\left[\kappa_3 (1-q)t + \left(\int_\Omega |x|^{-s}u_0^{m+1}\mathrm{d}x\right)^{(1-q)/(m+1)}\right]^{(m+q)/(1-q)}. \tag{23}$$

On the other hand, taking the test function $\phi = (u^m)_t$ in (3), then by using Cauchy's inequality with ε and Hölder's inequality, one can obtain

$$\begin{aligned}
&\frac{4m}{(m+1)^2}\int_\Omega |x|^{-s}\left[\left(u^{(m+1)/2}\right)_t\right]^2\mathrm{d}x + \frac{1}{p}\frac{\mathrm{d}}{\mathrm{d}t}\|u^m\|^p \\
&= \frac{2m}{m+1}\int_\Omega |x|^{s/2}u^{(m+2q-1)/2}|x|^{-(s/2)}\left(u^{(m+1)/2}\right)_t\mathrm{d}x \\
&\le \frac{2m\varepsilon}{m+1}\int_\Omega |x|^{-s}\left[\left(u^{(m+1)/2}\right)_t\right]^2\mathrm{d}x + \frac{m}{2\varepsilon(m+1)}\int_\Omega |x|^{s}u^{m+2q-1}\mathrm{d}x \\
&\le \frac{2m\varepsilon}{m+1}\int_\Omega |x|^{-s}\left[\left(u^{(m+1)/2}\right)_t\right]^2\mathrm{d}x + \frac{m}{2\varepsilon(m+1)}\left(\int_\Omega |x|^{s(q+m)/(1-q)}\mathrm{d}x\right)^{(1-q)/(q+m)}\left(\int_\Omega u^{q+m}\mathrm{d}x\right)^{(m+2q-1)/(q+m)}.
\end{aligned} \tag{24}$$

Let ε be sufficiently small to ensure that $(4m/(m+1)^2) - (2m\varepsilon/(m+1)) \ge 0$, then by (21), (23), and (24), one has

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u^m\|^p \le \frac{mp\kappa_3^2}{2\varepsilon(m+1)}\left[\kappa_3 (1-q)t + \left(\int_\Omega |x|^{-s}u_0^{m+1}\mathrm{d}x\right)^{(1-q)/(m+1)}\right]^{(m+2q-1)/(1-q)}. \tag{25}$$

Integrating (25) from 0 to t yields that

$$\begin{aligned}\left\|u^m\right\|^p &\le \left\|u_0^m\right\|^p + \frac{mp\kappa_3^2}{2\varepsilon(m+1)}\int_0^t\left[\kappa_3(1-q)\tau + \left(\int_\Omega |x|^{-s}u_0^{m+1}\mathrm{d}x\right)^{(1-q)/(m+1)}\right]^{(m+2q-1)/(1-q)}\mathrm{d}\tau \\ &\le \left\|u_0^m\right\|^p - \frac{mp\kappa_3}{2\varepsilon(m+1)(m+q)}\left(\int_\Omega |x|^{-s}u_0^{m+1}\mathrm{d}x\right)^{(m+q)/(m+1)} \\ &\quad + \frac{mp\kappa_3}{2\varepsilon(m+1)(m+q)}\left(\kappa_3(1-q)t + \left(\int_\Omega |x|^{-s}u_0^{m+1}\mathrm{d}x\right)^{(1-q)/(m+1)}\right)^{(m+q)/(1-q)},\end{aligned} \tag{26}$$

which means that the solution $u(x,t)$ of the problem (1) is global.

Case 2. If $q = 1$, taking the test function $\phi = u^m(x,t)$ in (3), then we can see that

$$\begin{aligned}&\frac{1}{m+1}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega |x|^{-s}u^{m+1}\mathrm{d}x + \left\|u^m\right\|^p \\ &= \int_\Omega |x|^{-s}u^{m+1}|x|^s\mathrm{d}x \le R^s\int_\Omega |x|^{-s}u^{m+1}\mathrm{d}x,\end{aligned} \tag{27}$$

which tells us that

$$\begin{aligned}\int_\Omega |x|^{-s}u^{m+1}\mathrm{d}x &\le e^{(m+1)R^st}\int_\Omega |x|^{-s}u_0^{m+1}\mathrm{d}x, \\ \|u\|_{m+1}^{m+1} &\le R^s e^{(m+1)R^st}\int_\Omega |x|^{-s}u_0^{m+1}\mathrm{d}x.\end{aligned} \tag{28}$$

On the other hand, taking the test function $\phi = (u^m)_t$ in (3), then Cauchy's inequality with ε leads to

$$\begin{aligned}&\frac{4m}{(m+1)^2}\int_\Omega |x|^{-s}\left[\left(u^{(m+1)/2}\right)_t\right]^2\mathrm{d}x + \frac{1}{p}\frac{\mathrm{d}}{\mathrm{d}t}\left\|u^m\right\|^p \\ &= \frac{2m}{m+1}\int_\Omega |x|^{s/2}u^{(m+1)/2}|x|^{-(s/2)}\left(u^{(m+1)/2}\right)_t\mathrm{d}x \\ &\le \frac{2m\varepsilon}{m+1}\int_\Omega |x|^{-s}\left[\left(u^{(m+1)/2}\right)_t\right]^2\mathrm{d}x + \frac{m}{2\varepsilon(m+1)}\int_\Omega |x|^s u^{m+1}\mathrm{d}x \\ &\le \frac{2m\varepsilon}{m+1}\int_\Omega |x|^{-s}\left[\left(u^{(m+1)/2}\right)_t\right]^2\mathrm{d}x + \frac{mR^s}{2\varepsilon(m+1)}\int_\Omega u^{m+1}\mathrm{d}x.\end{aligned} \tag{29}$$

Choosing $\varepsilon \in (0, (2/(m+1)))$ to guarantee that $(4m/(m+1)^2) - (2\lambda m\varepsilon/(m+1)) \ge 0$, then by (29), one has

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\|u^m\right\|^p \le \frac{mpR^{2s}}{2\varepsilon(m+1)}e^{(m+1)R^st}\int_\Omega |x|^{-s}u_0^{m+1}\mathrm{d}x, \tag{30}$$

which implies that

$$\left\|u^m\right\|^p \le \left\|u_0^m\right\|^p + \frac{mpR^s}{2\varepsilon(m+1)^2}\left(e^{(m+1)R^st} - 1\right)\int_\Omega |x|^{-s}u_0^{m+1}\mathrm{d}x. \tag{31}$$

Then, the proof of the global existence result is complete.

Now, we take our attention to the extinction singularity of the solution $u(x,t)$ to problem (1). We denote $\alpha = p(N-s)/(N-p)$. Noticing that $0 \le s < p$, we can verify that $\alpha > p$. Let a be a constant satisfying

$$a > \max\left\{\frac{m+1-\alpha m}{m\alpha - mp}, -\frac{1}{p}\right\}. \tag{32}$$

From (32), it follows that

$$0 < \frac{m(pa+1)+1}{\alpha m(a+1)} < 1. \tag{33}$$

Selecting the test function $\phi = u^{m(pa+1)}(x,t)$ in (3), one has

$$\frac{1}{m(pa+1)+1}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega |x|^{-s}u^{m(pa+1)+1}\mathrm{d}x + \frac{pa+1}{(a+1)^p}\int_\Omega \left|\nabla u^{m(a+1)}\right|^p\mathrm{d}x = \int_\Omega u^{m(pa+1)+q}\mathrm{d}x. \tag{34}$$

Making use of Hölder's inequality, one can find that

$$\begin{aligned}\int_\Omega |x|^{-s}u^{m(pa+1)+1}\mathrm{d}x &= \int_\Omega |x|^{-(s[m(pa+1)+1]/\alpha m(a+1))}u^{m(a+1)\cdot((m(pa+1)+1)/m(a+1))}|x|^{-s(1-((m(pa+1)+1)/\alpha m(a+1)))}\mathrm{d}x\\ &\le \left(\int_\Omega |x|^{-s}u^{\alpha m(a+1)}\mathrm{d}x\right)^{(m(pa+1)+1)/\alpha m(a+1)}\left(\int_\Omega |x|^{-s}\mathrm{d}x\right)^{1-((m(pa+1)+1)/\alpha m(a+1))}\\ &= \kappa_1^{1-((m(pa+1)+1)/\alpha m(a+1))}\left(\int_\Omega |x|^{-s}u^{\alpha m(a+1)}\mathrm{d}x\right)^{(m(pa+1)+1)/\alpha m(a+1)},\end{aligned} \tag{35}$$

where κ_1 is the same as that in (12). Since $0<m(p-1)<1$ and $0<m\le 1$, one can deduce that $1<p<1+(1/m)$. This together with the assumption $N>p$ one has $1<p<\min\{N,1+(1/m)\}$. Meanwhile, recalling that $0\le s<\min\{p,(Nq/(q+m))\}$, then it follows from Lemma 2 that

$$\int_\Omega |x|^{-s}u^{\alpha m(a+1)}\mathrm{d}x = \int_\Omega |x|^{-s}u^{m(a+1)\cdot(p(N-s)/(N-p))}\mathrm{d}x \le \kappa_2\left(\int_\Omega \left|\nabla u^{m(a+1)}\right|^p\mathrm{d}x\right)^{(N-s)/(N-p)}, \tag{36}$$

where κ_2 is given in Lemma 2. Combining (35) with (32), one sees

$$\begin{aligned}\int_\Omega |x|^{-s}u^{m(pa+1)+1}\mathrm{d}x &\le \kappa_1^{1-((m(pa+1)+1)/\alpha m(a+1))}\kappa_2^{((m(pa+1)+1)/\alpha m(a+1))}\left(\int_\Omega \left|\nabla u^{m(a+1)}\right|^p\mathrm{d}x\right)^{((N-s)[m(pa+1)+1])/(\alpha m(a+1)(N-p))}\\ &= \kappa_1^{1-((m(pa+1)+1)/\alpha m(a+1))}\kappa_2^{((m(pa+1)+1)/\alpha m(a+1))}\left(\int_\Omega \left|\nabla u^{m(a+1)}\right|^p\mathrm{d}x\right)^{((m(pa+1)+1)/mp(a+1))}.\end{aligned} \tag{37}$$

Exploiting (34) and (37), one can arrive at

$$\begin{aligned}&\frac{1}{m(pa+1)+1}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega |x|^{-s}u^{m(pa+1)+1}\mathrm{d}x\\ &+\frac{pa+1}{(a+1)^p}\kappa_1^{((N-p)/(N-s))-(mp(a+1)/(m(pa+1)+1))}\kappa_2^{-((N-p)/(N-s))}\left(\int_\Omega |x|^{-s}u^{m(pa+1)+1}\mathrm{d}x\right)^{mp(a+1)/(m(pa+1)+1)}\\ &\le \int_\Omega u^{q+m(pa+1)}\mathrm{d}x.\end{aligned} \tag{38}$$

In what follows, for the sake of simplicity, we denote $y(t)=\int_\Omega |x|^{-s}u^{m(pa+1)+1}\mathrm{d}x$, and $C_1=[m(pa+1)+1]((pa+1)/(a+1)^p)\kappa_1^{((N-p)/(N-s))-(mp(a+1)/(m(pa+1)+1))}\kappa_2^{-((N-p)/(N-s))}$.

If $\max\{((1-m)/2),m(p-1)\}<q=1$, then from (38), one can immediately know that

$$\frac{\mathrm{d}y}{\mathrm{d}t}+C_1y^{mp(a+1)/(m(pa+1)+1)}(t)\le C_2y(t), \tag{39}$$

where $C_2=[m(pa+1)+1]R^s$. Remembering that $0<m(p-1)<1$ and (32), one can check that $0<(mp(a+1)/(m(pa+1)+1))<1$. If $2C_2<C_1y^{(m(p-1)-1)/(m(pa+1)+1)}(0)$, then Lemma 4 tells us that there are two positive constants η_1 and ξ_1 satisfying

$$0\le y(t)\le \xi_1e^{-\eta_1t}. \tag{40}$$

Putting $T_0=\max\{0,((m(pa+1)+1)/\eta_1[1-m(p-1)])\ln((2C_2/C_1)\xi_1^{(1-m(p-1))/(m(pa+1)+1)})\}$, then for any $t>T_0$, (40) leads to

$$2C_2y^{(1-m(p-1))/(m(pa+1)+1)}(t)<C_1, \tag{41}$$

which together with (39) yields

$$\frac{\mathrm{d}y}{\mathrm{d}t}+\frac{C_1}{2}y^{mp(a+1)/(m(pa+1)+1)}\le 0,\quad t\ge T_0. \tag{42}$$

Integrating above inequality from T_0 to t leads to

$$y^{(1-m(p-1))/(m(pa+1)+1)}(t) \le y^{(1-m(p-1))/(m(pa+1)+1)}(T_0) - \frac{C_1[1-m(p-1)]}{2[m(pa+1)+1]}(t-T_0), \quad t \ge T_0. \tag{43}$$

The above inequality means that

$$\lim_{t\longrightarrow T_1} y(t) = \lim_{t\longrightarrow T_1} \int_\Omega |x|^{-s} u^{m(pa+1)+1} \mathrm{d}x = 0, \tag{44}$$

where

$$T_1 = T_0 + \frac{2[m(pa+1)+1]}{C_1[1-m(p-1)]} y^{\frac{1-m(p-1)}{m(pa+1)+1}}(T_0). \tag{45}$$

If $\max\{((1-m)/2), m(p-1)\} < q < 1$. In view of Hölder's inequality, one has

$$\begin{aligned} \int_\Omega u^{m(pa+1)+q} \mathrm{d}x &\le |\Omega|^{(1-q)/(m(pa+1)+1)} \left(\int_\Omega u^{m(pa+1)+1} \mathrm{d}x \right)^{(m(pa+1)+q)/(m(pa+1)+1)} \\ &\le |\Omega|^{(1-q)/(m(pa+1)+1)} R^{s[m(pa+1)+q]/(m(pa+1)+1)} \left(\int_\Omega |x|^{-s} u^{m(pa+1)+1} \mathrm{d}x \right)^{(m(pa+1)+q)/(m(pa+1)+1)}. \end{aligned} \tag{46}$$

Combining (38) with (46), one can conclude that

$$\frac{\mathrm{d}y}{\mathrm{d}t} + C_1 y^{mp(a+1)/(m(pa+1)+1)}(t) \le C_3 y^{(m(pa+1)+q)/(m(pa+1)+1)}(t), \tag{47}$$

where

$$C_3 = \lambda[m(pa+1)+1]|\Omega|^{(1-q)/(m(pa+1)+1)} R^{s[m(pa+1)+q]/(m(pa+1)+1)}. \tag{48}$$

Recalling that $0 < m(p-1) < q < 1$ and (32), one can check that

$$0 < \frac{mp(a+1)}{m(pa+1)+1} < \frac{m(pa+1)+q}{m(pa+1)+1} < 1. \tag{49}$$

Then, by (47) and Lemma 4, one knows that there are two positive constants η_2 and ξ_2 satisfying

$$0 \le y(t) \le \xi_2 e^{-\eta_2 t}, \tag{50}$$

provided that $2C_3 y^{(q-m(p-1))/(m(pa+1)+1)}(0) < C_1$. Setting $T_2 = \max\{0, ((m(pa+1)+1)/\eta_2[q-m(p-1)])\ln((2C_3/C_1)\xi_2^{(q-m(p-1))/(m(pa+1)+1)})\}$, then for any $t > T_1$, (50) leads to

$$2C_3 y^{(q-m(p-1))/(m(pa+1)+1)}(t) < C_1, \tag{51}$$

which together with (47) yields

$$\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{C_1}{2} y^{mp(a+1)/(m(pa+1)+1)} \le 0, \quad t \ge T_2. \tag{52}$$

The remainder proof is the same as the previous one in the case $q = 1$, and we omit it here. Up to now, the proof of the extinction phenomenon of the solution $u(x,t)$ to problem (1) is complete.

Now, we begin to prove the nonextinction result. Denoting

$$M(t) = \frac{1}{m+1} \int_\Omega |x|^{-s} u^{m+1} \mathrm{d}x, \tag{53}$$

and taking the derivative of $M(t)$ with respect to t, one has

$$\begin{aligned} M'(t) &= \int_\Omega |x|^{-s} u^m u_t \mathrm{d}x \\ &= \int_\Omega u^{m+q} \mathrm{d}x - \int_\Omega |\nabla u^m|^p \mathrm{d}x \\ &= \frac{q-m(p-1)}{m+q} \int_\Omega u^{m+q} \mathrm{d}x - pE(u). \end{aligned} \tag{54}$$

If $q = m(p-1)$. From (8), one knows that $E(u)$ is nonincreasing with respect to t. Then, for any $t \ge 0$, one has

$$M'(t) \ge -pE(u_0). \tag{55}$$

Integrating, one obtains

$$M(t) \ge M(0) - pE(u_0)t, \quad \text{for } t \ge 0, \tag{56}$$

which tells us that $M(t) > 0$ since $M(0) > 0$ and $E(u_0) \le 0$; that is, the solution $u(x,t)$ of problem (1) does not possess extinction phenomenon.

If $q < m(p-1)$, with the help of Hölder's inequality, one gets

$$\frac{q-m(p-1)}{m+q} \int_\Omega u^{m+q} \mathrm{d}x \ge \frac{q-m(p-1)}{m+q} M^{(m+q)/(m+1)}(t) \left(\int_\Omega |x|^{s(m+q)/(1-q)} \mathrm{d}x \right)^{(1-q)/(m+1)}. \tag{57}$$

Exploiting (54) and (57), one can claim that

$$M'(t) - C_4 M^{(m+q)/(m+1)}(t) \geq -pE(u_0), \tag{58}$$

where

$$0 > C_4 = \frac{q - m(p-1)}{m+q}\left(\int_\Omega |x|^{s(m+q)/(1-q)} \mathrm{d}x\right)^{(1-q)/(m+1)} > \frac{q - m(p-1)}{m+q} R^{s(m+q)/(m+1)} |\Omega|^{(1-q)/(m+1)} > -\infty. \tag{59}$$

Since $M(0) > 0$ and $E(u_0) < 0$, then from (58) and Lemma 3, it follows that

$$M(t) \geq \min\left\{M(0), \left(\frac{pE(u_0)}{C_4}\right)^{(m+1)/(m+q)}\right\} > 0, \tag{60}$$

which means that the solution $u(x,t)$ of problem (1) does not possess extinction phenomenon.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported by the Natural Science Foundation of Hunan Province (Grant no. 2019JJ50160), Scientific Research Fund of Hunan Provincial Education Department (Grant no. 20A174), and Scientific Research Fund of Hunan University of Science and Technology (Grant no. KJ2123).

References

[1] D. M. Liu and C. L. Mu, "Cauchy problem for a doubly degenerate parabolic equation with inhomogeneous source and measure data," *Differential and Intergral Equations*, vol. 27, pp. 1001–1012, 2014.

[2] J. L. Vázquez, *The Porous Medium Equation: Mathematical Theory*, The Clarendon Press, Oxford University Press, Oxford, UK, 2007.

[3] Z. Q. Wu, J. N. Zhao, J. X. Yin, and H. L. Li, *Nonlinear Diffusion Equations*, World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2001.

[4] D. Li, J. Liu, and L. Zheng, "A zero-dimensional valuation ring is 1-Gröbner," *Journal of Algebra*, vol. 484, pp. 334–343, 2017.

[5] Y. Liu and H. Su, "Some necessary and sufficient conditions for containment of second-order multi-agent systems with sampled position data," *Neurocomputing*, vol. 378, pp. 228–237, 2020.

[6] C. J. Xu, H. C. Xu, H. S. Su, and C. Liu, "Disturbance-observer based consensus of linear multi-agent systems with exogenous disturbance under intermittent communication," *Neurocomputing*, vol. 404, pp. 26–33, 2020.

[7] Q. Xin, "Partial differential equation and its applications in image denoising and segmentation (in Chinese)," Ph. D. thesis, Chongqing: College of Mathematics and Statistics, Chongqing University, Chongqing, China, 2011.

[8] C. J. Xu, B. F. Li, and L. Yang, "Semi-global containment of discrete-time high-order multi-agent systems with input saturation via intermittent control," *IET Control Theory and Applications*, vol. 14, pp. 2303–2309, 2020.

[9] Z. Chaouai and A. E. Hachimi, "Qualitative properties of weak solutions for p- Laplacian equations with nonlocal source and gradient absorption," *Bulletin of the Korean Mathematical Society*, vol. 57, pp. 1003–1031, 2020.

[10] Y. Chen, J. Wang, and H. Zhang, "Extinction for a couple of fast diffusion systems with nonlinear sources," *Nonlinear Analysis: Real World Applications*, vol. 14, no. 4, pp. 1931–1937, 2013.

[11] E. DiBenedetto, *Degenerate Parabolic Equations*, Springer, NewYork, NY, USA, 1993.

[12] Z. B. Fang and X. Xu, "Extinction behavior of solutions for the -Laplacian equations with nonlocal sources," *Nonlinear Analysis: Real World Applications*, vol. 13, no. 4, pp. 1780–1789, 2012.

[13] Y. Han and W. Gao, "Extinction for a fast diffusion equation with a nonlinear nonlocal source," *Archiv der Mathematik*, vol. 97, no. 4, pp. 353–363, 2011.

[14] Y. Han, H. Li, and W. Gao, "Extinction and non-extinction of solutions to a P-laplace equation with a nonlocal source and an absorption term," *Mathematical Modelling and Analysis*, vol. 19, no. 2, pp. 169–179, 2014.

[15] R. G. Iagar and P. Laurençot, "Positivity, decay, and extinction for a singular diffusion equation with gradient absorption," *Journal of Functional Analysis*, vol. 262, no. 7, pp. 3186–3239, 2012.

[16] M. Liao and W. Gao, "Blow-up phenomena for a nonlocal p-Laplace equation with Neumann boundary conditions," *Archiv der Mathematik*, vol. 108, no. 3, pp. 313–324, 2017.

[17] D. M. Liu and C. L. Mu, "Extinction for a quasilinear parabolic equation with a nonlinear gradient source," *Taiwanese Journal of Mathematics*, vol. 18, pp. 1329–1343, 2014.

[18] D. M. Liu and C. L. Mu, "Extinction for a quasilinear parabolic equation with a nonlinear gradient source and absorption," *Journal of Applied Analysis and Computation*, vol. 5, pp. 114–137, 2015.

[19] D. Liu and C. Mu, "Critical extinction exponent for a doubly degenerate non-divergent parabolic equation with a gradient source," *Applicable Analysis*, vol. 97, no. 12, pp. 2132–2141, 2018.

[20] D. M. Liu, C. L. Mu, and I. Ahmed, "Blow-up for a semilinear parabolic equation with nonlinear memory and nonlocal nonlinear boundary," *Taiwanese Journal of Mathematics*, vol. 17, pp. 1353–1370, 2013.

[21] D. M. Liu and L. Yang, "Extinction phenomenon and decay estimate for a quasilinear parabolic equation with a nonlinear source," *Advances in Mathematical Physics*, vol. 2021, Article ID 5569043, 7 pages, 2021.

[22] Y. Wang and J. Yin, "Critical extinction exponents for a polytropic filtration equation with absorption and source," *Mathematical Methods in the Applied Sciences*, vol. 36, no. 12, pp. 1591–1597, 2013.

[23] X. Xu and T. Cheng, "Extinction and decay estimates of solutions for a non-Newton polytropic filtration system," *Bulletin of the Malaysian Mathematical Sciences Society*, vol. 43, no. 3, pp. 2399–2415, 2020.

[24] J.-S. Guo, B. Hu, and B. Hu, "Blowup rate estimates for the heat equation with a nonlinear gradient source term," *Discrete*

and Continuous Dynamical Systems - A, vol. 20, no. 4, pp. 927–937, 2008.

[25] P. Quittner and P. Souplet, *Superlinear Parabolic Problems: Blow-Up, Global Existence and Steady States*, Birkhäuser Verlag AG, Basel, Switzerland, 2007.

[26] J. L. Vázquez and P. Souplet, "Stabilization towards a singular steady state with gradient blow-up for a diffusion-convection problem," *Discrete and Continuous Dynamical Systems*, vol. 14, pp. 221–234, 2006.

[27] H. Yuan, X. Xu, W. Gao, S. Lian, and C. Cao, "Extinction and positivity for the evolution p-Laplacian equation with L1 initial value," *Journal of Mathematical Analysis and Applications*, vol. 310, no. 1, pp. 328–337, 2005.

[28] Y. G. Gu, "Necessary and sufficient conditions of extinction of solution on parabolic equations," *Acta Math. Sinica*, vol. 37, pp. 73–79, 1994, in Chinese.

[29] Y. Tian and C. Mu, "Extinction and non-extinction for a *p*-Laplacian equation with nonlinear source," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 69, no. 8, pp. 2422–2431, 2008.

[30] J. X. Yin and C. H. Jin, "Critical extinction and blow-up exponents for fast diffusive *p*-Laplacian with sources," *Mathematical Methods in the Applied Sciences*, vol. 31, pp. 1383–1386, 2007.

[31] C. Jin, J. Yin, and Y. Ke, "Critical extinction and blow-up exponents for fast diffusive polytropic filtration equation with sources," *Proceedings of the Edinburgh Mathematical Society*, vol. 52, no. 2, pp. 419–444, 2009.

[32] J. Zhou and C. Mu, "Critical blow-up and extinction exponents for non-Newton polytropic filtration equation with source," *Bulletin of the Korean Mathematical Society*, vol. 46, no. 6, pp. 1159–1173, 2009.

[33] Z. Tan, "Non-Newton filtration equation with special medium void," *Acta Mathematica Scientia*, vol. 24, no. 1, pp. 118–128, 2004.

[34] Y. Wang, "The existence of global solution and the blowup problem for some p-Laplace heat equations," *Acta Mathematica Scientia*, vol. 27, no. 2, pp. 274–282, 2007.

[35] J. Zhou, "A multi-dimension blow-up problem to a porous medium diffusion equation with special medium void," *Applied Mathematics Letters*, vol. 30, pp. 6–11, 2014.

[36] J. Zhou, "Global existence and blow-up of solutions for a non-Newton polytropic filtration system with special volumetric moisture content," *Computers and Mathematics with Applications*, vol. 71, no. 5, pp. 1163–1172, 2016.

[37] X. Deng and J. Zhou, "Global existence, extinction, and non-extinction of solutions to a fast diffusion p-laplace evolution equation with singular potential," *Journal of Dynamical and Control Systems*, vol. 26, no. 3, pp. 509–523, 2020.

[38] M. Badiale and G. Tarantello, "A sobolev-hardy inequality with applications to a nonlinear elliptic equation arising in astrophysics," *Archive for Rational Mechanics and Analysis*, vol. 163, no. 4, pp. 259–293, 2002.

[39] B. Guo and W. Gao, "Non-extinction of solutions to a fast diffusive *p*-Laplace equation with Neumann boundary conditions," *Journal of Mathematical Analysis and Applications*, vol. 422, no. 2, pp. 1527–1531, 2015.

[40] W. Liu and B. Wu, "A note on extinction for fast diffusivep-Laplacian with sources," *Mathematical Methods in the Applied Sciences*, vol. 31, no. 12, pp. 1383–1386, 2008.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5575815, 12 pages
https://doi.org/10.1155/2021/5575815

Research Article

Cooperative Evolution Mechanism of Unmanned Swarm within the Framework of Public Goods Game

Minggang Yu,[1,2] **Ming He**,[1] **Ziyu Ma**,[1] **Mingguang Zou**,[3] **Lei Wan**,[4] **and Kai Kang**[1]

[1]*Institute of Command and Control Engineering, Army Engineering University of PLA, Nanjing 210007, China*
[2]*Institute of Communication Engineering, Army Engineering University of PLA, Nanjing 210007, China*
[3]*School of Computer of Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China*
[4]*Turbomachinery and Propulsion Technology Institute of Zhejiang Province, Hangzhou, Zhejiang 310000, China*

Correspondence should be addressed to Ming He; 329674406@qq.com

Received 26 January 2021; Revised 20 March 2021; Accepted 28 March 2021; Published 7 April 2021

Academic Editor: Xiao Ling Wang

One of the key advantages of unmanned swarm operation is its autonomous cooperation. When the communication is interrupted or the centralized control manner is lost, the cooperative operation can still be carried out orderly. This work proposed a cooperative evolution mechanism within the framework of multiplayer public goods game to solve the problem of autonomous collaboration of unmanned swarm in case of failure of centralized control. It starts with the requirement analysis of autonomous cooperation in unmanned swarm, and then, the evolutionary game model of multiplayer public goods based on aspiration-driven dynamics is established. On this basis, the average abundance function is constructed by theoretical derivation, and furthermore, the influence of cost, multiplication factor, and aspiration level on the average abundance is simulated. Finally, the evolutionary mechanism of parameter adjustment in swarm cooperation is revealed via case study, and deliberate proposals are suggested to provide a meaningful exploration in the actual control of unmanned swarm cooperation.

1. Introduction

With the continuous advancing of the third wave of artificial intelligence, "group evolutionary intelligence" developed from "single-agent autonomous intelligence" has become one of the important characteristics of the new generation of artificial intelligence. Particularly in the military field, unmanned swarm (unmanned vehicle cluster [1], unmanned boat cluster [2], and unmanned aerial vehicle cluster [3]) operations have received unprecedented attention over the past two years. The US military has listed unmanned swarm operations as a "subversive technology" that can change the rules of war.

There are mainly two kinds of control modes of unmanned swarm: centralized control and autonomous collaboration. Under the premise of good communication, the Command and Control (C2) center can implement centralized control on the swarm. However, in the complex electromagnetic environment of the battlefield, there is a real risk of communication failure [4]. In such a predicament, the centralized control mode fails, and the unmanned swarm must make effective response on the spot according to the external situation and achieve self-management and self-coordination. An issue that has led to considerable interest is how unmanned swarms autonomously and cooperatively complete established military operations. The sketch of autonomous cooperation of unmanned swarm is shown in Figure 1.

Overall planning and reallocation of operation resources (communication, firepower, intelligence, etc.) within the unmanned swarms is required when autonomous collaboration occurs. However, in the process mentioned above, there are often contradictions between individual partiality and swarm needs, which are difficult to reconcile. For example, in the fire strike task, the "rational" unmanned units with intelligence and decision-making ability will choose to "contribute" ammunition to the swarm as little as possible in order to maintain its combat effectiveness, while on the

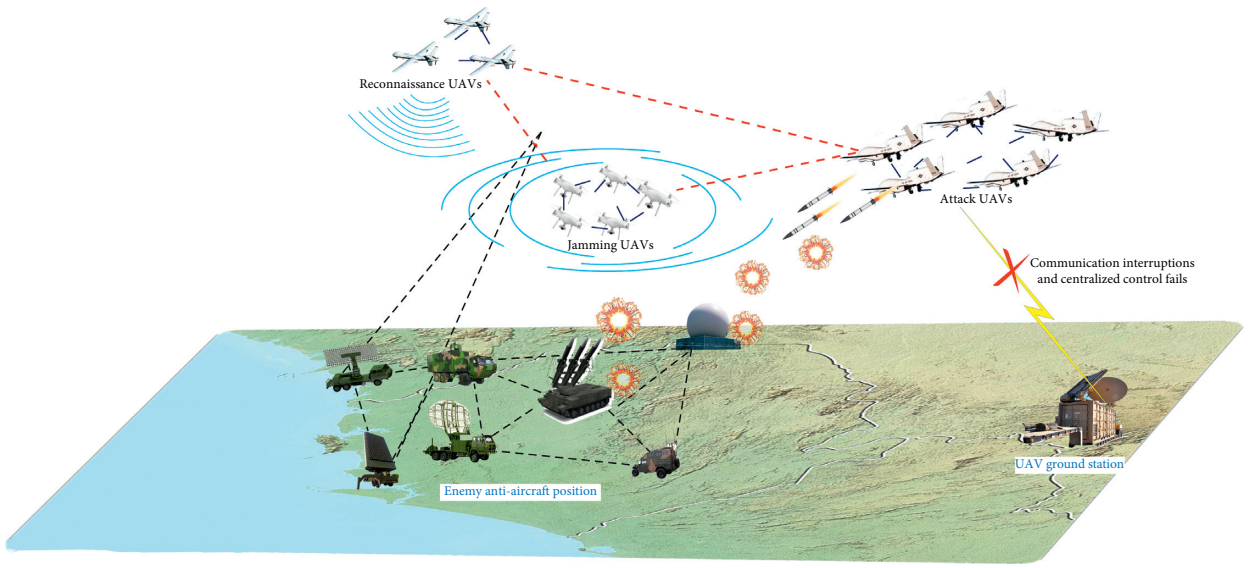


Figure 1: Sketch map of autonomous cooperation of unmanned swarm.

other hand, the more ammunition each unit contributes to the swarm, the higher the survival rate and the greater the combat effectiveness of the whole swarm will be. The contradiction between the two will lead to "tragedy of the commons" [5]; therefore, how to increase the number of units' willing to positively contribute ammunition to the swarm and avoid the tragedy has become a crucial and urgent problem in both technology research and practical application of unmanned swarm.

2. Related Work

Evolutionary game theory [6–9] combines "equilibrium" in economics with "adaptability" in biology to depict the process that individuals adapt to the external environment through learning, imitation, and trial-and-error under boundary rationality and asymmetric information. And the evolutionary game of public goods (PGG) [10] provides a basic theoretical framework for revealing the cooperative evolution mechanism and coping with the tragedy of the commons. PGG reflects that investors (collaborators) and hitchhikers (betrayers) play strategic games over time based on cost, multiplication factor, selection intensity, etc., which makes the proportion of collaborators and betrayers in the population change dynamically, and finally tends to an evolutionarily stable state (ESS). The research focus of PGG is to calculate the mathematical expectation of the proportion of collaborators in a population after multiround game, that is, the average abundance, and then analyze the relationship between average abundance and parameters (cost, multiplication factor, selection intensity, etc.) to achieve the ultimate purpose of manual control.

At present, there are two main research directions to solve the problem of swarm cooperation with evolutionary game theory: one is to study the evolutionary dynamics process and cooperation mechanism of spatial structured population such as complex network based on graph theory [11, 12], and the other is to study the evolutionary stability state of well-mixed population and the dominant condition of cooperation based on Markov stochastic process [13, 14].

For the former, the team of Professor Nowak from Harvard University theoretically deduced the evolution of population in spatial structure such as circle, random graph, and scale-free network and creatively proposed the relationship between the ratio b/c and the network average degree k. They pointed out that the smaller the network connectivity is, the more conducive the cooperation in natural selection is [15]. Then, they used the pair approximation theory to theoretically deduce the cooperation phenomenon on the regular lattice and obtained the boundary conditions for the generation and expansion of cooperation [16]. On the basis of the above achievements, further comparative analysis was made on the differences between homogeneous and heterogeneous networks in promoting cooperative behavior, and simulation results show that weak connection can better promote cooperative behavior on heterogeneous networks [17]. At the same time, other researchers studied the dynamic process of multiparty game on the graph, and simulation results show that spatial structured population can promote the occurrence of cooperation better compared with unstructured population. In the recent two years, the team of Nowak has applied the evolution dynamics of cooperation in spatial structure to social network, analyzed the critical conditions of cooperation behavior in human society [18], initially explored the trade-off between the evolution convergence probability and the evolution convergence time [19], and extended the cooperative evolution on structural population to weighted graph [20]. Other representative studies include literature [21, 22] on the specific model of multiplayer snowdrift game, the relationship curves between the ratio b/c and cooperation level on the well-mixed population and the structured population, respectively, and the significant differences between the homogeneous/heterogeneous network and the unstructured population in the promotion of cooperation are compared.

For the latter, the representative researches are as follows: Wang at Wuhan University obtained the average abundance function of the snowdrift evolutionary game using the stable distribution of Markov chain and simulated the effect of parameters on the average abundance [23]. Based on the research work of Tarnita et al. [24], Du at Peking University obtained the inequality of strategic dominance conditions in two-party evolutionary game through strict mathematical derivation; besides, the simulation results show that the average abundance under weak selection intensity is independent of aspiration level [25].

The aforementioned work on swarm cooperation is of great theoretical and engineering value. We have also conducted an exploration on the cooperation mechanism of unmanned swarm, and the relevant results can be referred to in [26–30]. Nevertheless, there are still two shortcomings in the above achievements in solving the issue of cooperative evolution of unmanned swarms: first, it has not focused on the public goods game; that is, although there are similarities between the snowdrift game and public goods game [31], there are essential differences in the game mechanism; in addition, the cooperative evolution of unmanned swarm is the multiple interaction of combat units; that is, the evolution result is not only related to the strategy selection of single unit, but also depends on the strategy of other units in the swarm, which is characterized by multiplayer games [32, 33]. So far, the academic community has mastered the payoff matrix [34] of the public goods game with multiplayer and has simulated the influence of different selection intensity [35–37] and threshold values [31] on the cooperation level. In particular, in literature [34], the authors derived a general average abundance formula of multiparty games in a finite population under aspiration-driven dynamics, which can be applicable to any multiparty game under the aspiration-driven dynamics of a finite population. However, the average abundance formula for specific public goods game is not mentioned, so the important work of this study is to obtain the average abundance analytic expression of public goods game based on the existing payment matrix.

Furthermore, the generalized evolutionary game model can be simplified as Markov chain + strategy update rules in finite population, so the average abundance function is also closely related to the strategy update rules. By studying strategy update rules in the framework of evolutionary game theory, one can differentiate between imitation processes and aspiration-driven dynamics [38]. In the former case, individuals imitate the strategy of a more successful peer [39]. In the latter case, individuals adjust their strategies based on a comparison between their own payoff and the value they aspire, called the level of aspiration [40]. Unlike the imitation processes of pairwise comparison, aspiration-driven updates do not require additional information about the strategic environment and can thus be interpreted as being more spontaneous [41, 42]. In the complex battlefield environment, the information acquisition is incomplete, asymmetric, which requires the swarm to achieve self-management and self-coordination, and this requirement just coincides with the aspiration-driven dynamics. Moreover, the existing results show that, in both prisoner's dilemma game and public goods game, the dynamic mechanism driven by aspiration can improve the average abundance value and promote cooperation more than the traditional imitation dynamics [43, 44].

Aiming at the cooperative evolution mechanism of the unmanned swarm, we modelled the evolution process based on multiplayer public goods game framework and aspiration-driven update rule and then deduced the average abundance function of the model by analyzing the stable distribution of the Markov chain; on this foundation, we studied the influence of relevant parameters on the average abundance through theoretical analysis and numerical calculation; finally, we studied the effect of parameter adjusting on swarm cooperation via case study and discussed the corresponding solutions and advice to avoid "the tragedy of the commons."

3. Model Hypothesis

In essence, the autonomous collaboration of unmanned swarms is a game process of multiparty and multiround, which focuses on the autonomous allocation of public resources. Therefore, we use multiplayer public goods evolutionary game to model the cooperative evolution of unmanned swarms. The mapping between the concepts of cooperative evolution in unmanned swarms and multiplayer public goods evolutionary game is listed in Table 1.

3.1. Framework of Multiplayer Public Goods Evolutionary Game. It is set that the autonomous cooperation of unmanned units takes place in a well-mixed swarm of size N, and every unit has two alternative strategies, A and B. Every d units interact simultaneously to get their payoffs; i.e., they are in a two-strategy and d-player game. The strategy update procedure is as follows:

(i) Select any focal individual X in the population of N and select $d-1$ individuals from the remaining $N-1$ individuals to form a group of $d\,(d < N)$. A focal individual can be of type A or B, and encounters a group containing $k\,(0 \le k \le d-1)$ other players of type A and $d-k-1$ players of type B.

(ii) The focal individual X plays games with the rest $d-1$ individuals in strategy space $\{A, B\}$. Denote by a_k and b_k the payoffs of strategy A and B a player obtains, respectively, when facing i other A individuals within the group of size d, where i ranges from 0 to $d-1$.

(iii) At the end of each round of the game, the focal individual X evaluates the benefits under different strategy choices and then updates its strategy according to aspiration-driven dynamics. The above process is repeated until the proportion of a certain strategy tends to be stable in the whole population. Obviously, the value of k determines the payoffs—a_k and b_k. All possible payoffs of a focal individual are uniquely defined by the number of A in the group, and the payoff matrix is as follows.

Table 1: Mapping between the concepts of cooperative evolution and multiplayer public goods game.

Cooperative evolution in unmanned swarms	Multiplayer public goods evolutionary game
Unmanned swarms	Mixed homogeneous population
Public operational resources (communication, firepower, etc.)	Public goods
Multiple unmanned units participating in autonomous collaboration	Multiplayer
Single combat unit	Individuals
Single combat unit as research object	Focal individual
Investors (collaborators)	Strategy A
Hitchhikers (betrayers)	Strategy B
Total resources that individuals ultimately obtain after each round of game	Payoffs
Payoff-based strategy transformation between units	Game
Dynamic change of proportion of units with different strategies in multiround game	Evolutionary
The proportion is stable and the game ends after multiround game	Evolutionary stable state

For any group engaging in a one-shot game, we can obtain each member's payoff according to Table 2. When X chooses strategy A, the total contribution by individuals to the swarm is $kc + c$, the total gain of the swarm is $r(kc + c)$ multiplied by the profit coefficient r, and the gain of each individual is $r(kc + c)/d$. As the cost of X is c, the net gain of X is $r(kc + c)/d - c$. When X chooses strategy B, the total contribution by individuals to the swarm is kc, the total gain of the swarm is rkc, and the gain of each individual is rkc/d. As there is no cost in such a case, the net gain of X is rkc/d. Thus, the payoffs for A and B are

$$a_k = \frac{r(kc + c)}{d - c}, \tag{1}$$

$$b_k = \frac{rkc}{d}. \tag{2}$$

3.2. Expected Payoff for Strategies A and B. In a finite well-mixed population of size N, groups of size d are assembled randomly, so the probability of choosing a group that consists of another k players of type A and $d - k - 1$ players of type B is given by a hypergeometric distribution [45]. For example, the probability that an A player is in a group of k other A's is given by $P_A(N, i; d, k) = C_{i-1}^{k} C_{N-i}^{d-k-1} / C_{N-1}^{d-1}$, where i is the number of A players in the population. The symbol C_n^k denotes a combinatorial notation, which is the number of ways to choose a k element subset from an n element set.

The expected payoffs for any A or B in a population of size N, with i players of type A and $N-i$ players of type B, are given by

$$\pi_A(i) = \sum_{k=0}^{d-1} \frac{C_{i-1}^{k} C_{N-i}^{d-k-1}}{C_{N-1}^{d-1}} a_k, \tag{3}$$

$$\pi_B(i) = \sum_{k=0}^{d-1} \frac{C_{i}^{k} C_{N-i-1}^{d-k-1}}{C_{N-1}^{d-1}} b_k. \tag{4}$$

3.3. Aspiration-Driven Dynamics. There are several typical strategy update rules in evolutionary game, such as unconditional imitation [35], Fermi rule [46, 47], etc.

Table 2: Payoff in multiplayer evolutionary game.

	$d-1$	...	k	...	0
A	a_{d-1}	...	a_k	...	a_0
B	b_{d-1}	...	b_k	...	b_0

Aspiration-driven dynamics focuses on comparing the payoff with aspiration level to make new decisions. Players need not see any particular payoffs but their own, which they compare with an aspiration value. The aspiration-driven dynamics coincides with the requirement of self-management and self-coordination of unmanned swarm in the case of incomplete information acquisition in complex battlefield. The level of aspiration, α, is a variable that influences the stochastic strategy updating. The probability of switching strategy is random when individuals' payoffs are close to the level of α, reflecting the basic degree of uncertainty in the population. When payoff exceeds α, strategy switching is unlikely. At high values of α compared with payoff, switching probabilities are high.

To model stochastic aspiration-driven switching (from strategy A to B), we can use the following probability function:

$$P_{A \longrightarrow B} = \frac{1}{1 + e^{\omega(\pi_A(i) - \alpha)}}. \tag{5}$$

The aspiration level, α, provides the benchmark used to evaluate how "greedy" an individual is. Higher aspiration levels mean that individuals aspire to higher payoffs. The intensity of selection, ω, provides a measure of how important individuals deem the impact of the actual game on their update. Let $\Delta = \pi_A - \alpha$; if $\Delta = 0$, then $P_{A \longrightarrow B} = 1/2$, which means that individuals have the same preference for strategies A and B; if $\Delta > 0$ (i.e., the individual payoff π_A is higher than aspiration level α), then $P_{A \longrightarrow B} < 1/2$, which means individuals prefer strategy A; if $\Delta < 0$ (i.e., individual payoff π_A is lower than aspiration level α), then $P_{A \longrightarrow B} > 1/2$, which means individuals prefer strategy B. As for whether an individual updates strategy or keeps strategy unchanged in a certain round of game, it can be further determined by other algorithms, such as roulette algorithm.

In the same way, the probability of the focal individual updating from strategy B to A is

$$P_{B\longrightarrow A} = \frac{1}{1+e^{\omega\left(\pi_B(i)-\alpha\right)}}. \tag{6}$$

In the aspiration-driven dynamics, at each time step, the number of strategy A, i.e., i, can only increase by one, decrease by one, or stay the same. When the number of strategy A increases by one, two subsequent events happen: first, a B strategy individual is selected from the population; then it does not satisfy with the payoff it obtains and switches to the strategy A. A similar process holds for the number of strategy B. Therefore, the probability that the number of A individuals changes at one time step is

$$P(i \longrightarrow i-1) = T_i^- = \frac{i}{N}\frac{1}{1+e^{\omega\left(\pi_A(i)-\alpha\right)}}, \tag{7}$$

$$P(i \longrightarrow i+1) = T_i^+ = \frac{N-i}{N}\frac{1}{1+e^{\omega\left(\pi_B(i)-\alpha\right)}}, \tag{8}$$

$$P(i \longrightarrow i) = 1 - T_i^- - T_i^+. \tag{9}$$

Because there is a stable distribution in the Markov chain without absorbing state, the average abundance function of multiplayer evolutionary game can be derived based on the above state transformation equation.

4. Average Abundance Function

At present, most of the research on average abundance is based on digital simulation, but no strict mathematical expression is given. In this part, we first give the definition of the average abundance of unmanned swarm and then derive its mathematical expression by analyzing the stable distribution of the nonabsorbent Markov chain to support the subsequent simulation analysis in Section 5.

4.1. Average Abundance

Definition 1. (average abundance of unmanned swarm). Set the proportion j/N of unmanned units with strategy A in a swarm as a random variable. Let $\nu(j)$ be the probability distribution of j/N; then the expected value of j/N is defined as the average abundance of unmanned units with strategy A.

Therefore, the definition of average abundance $\langle X_A(j)\rangle$ can be expressed as

$$\langle X_A(j)\rangle = \sum_{j=0}^{N}\frac{j}{N}\nu(j). \tag{10}$$

The key to calculating the average abundance is to determine the probability distribution $\nu(j)$. For Markov chains without absorbing state, $\nu(j)$ is just the stable distribution $\varphi_j\,(j\in[0,N])$, and it satisfies the detailed balance condition [48]: $\varphi_j T_j^+ = \varphi_{j+1}T_{j+1}^-$.

Equation (10) is just a definition formula, which cannot be directly applied to the actual calculation and analysis. Next, we will theoretically deduce the average abundance formula based on the detailed balance condition so as to reveal the quantitative relationship between the average abundance and related parameters (cost, multiplication factor, selection intensity, etc.) and provide a theoretical calculation basis for the subsequent characteristic analysis.

4.2. Function Deduction. It can be derived from the detailed balance condition:

$$\varphi_1 = \frac{T_0^+}{T_1^-}\varphi_0, \tag{11}$$

$$\varphi_2 = \frac{T_1^+}{T_2^-}\varphi_1 = \frac{T_0^+T_1^+}{T_1^-T_2^-}\varphi_0, \tag{12}$$

$$\varphi_3 = \frac{T_2^+}{T_3^-}\varphi_2 = \frac{T_0^+T_1^+T_2^+}{T_1^-T_2^-T_3^-}\varphi_0. \tag{13}$$

Further, we induce and summarize the above formulas; then we get

$$\varphi_j = \frac{\prod_{i=0}^{j-1}T_i^+}{\prod_{i=1}^{j}T_i^-}\varphi_0 = \prod_{i=0}^{j-1}h(i)\varphi_0 \quad (j\geq 1), \tag{14}$$

where $h(i) = T_i^+/T_{i+1}^-$ is the strategy dominant function. If $h(i)>1$, that is, the increasing probability of strategy A is greater than the decreasing probability, it means that strategy A is dominant in the swarm; otherwise, strategy B is dominant.

Since $\sum_{j=0}^{N}\varphi_j = 1$,

$$\sum_{j=0}^{N}\varphi_j = \varphi_0 + \sum_{j=1}^{N}\prod_{i=0}^{j-1}h(i)\varphi_0 = 1. \tag{15}$$

Then, we have $\varphi_0 = 1/1+\sum_{j=1}^{N}\prod_{i=0}^{j-1}h(i)$. Inserting φ_0 into formula (14), we have

$$\varphi_j = \frac{\prod_{i=0}^{j-1}h(i)}{1+\sum_{j=1}^{N}\prod_{i=0}^{j-1}h(i)} \quad (j\geq 1). \tag{16}$$

Inserting (16) into (10), $\langle X_A(j)\rangle$ can be written as

$$\begin{aligned}\langle X_A(j)\rangle &= \frac{1}{N}\sum_{j=1}^{N} j\frac{\prod_{i=0}^{j-1}h(i)}{1+\sum_{j=1}^{N}\prod_{i=0}^{j-1}h(i)}\\ &= \frac{1}{N}\frac{\sum_{j=1}^{N} j\prod_{i=0}^{j-1}h(i)}{1+\sum_{j=1}^{N}\prod_{i=0}^{j-1}h(i)},\end{aligned} \tag{17}$$

$$h(i) = \frac{T_i^+}{T_{i+1}^-} = \frac{(N-i)\left(1+e^{\omega\left(\pi_A(i+1)-\alpha\right)}\right)}{(i+1)\left(1+e^{\omega\left(\pi_B(i)-\alpha\right)}\right)}. \tag{18}$$

Equations (17) and (18) are just a general expression for the average abundance of multiplayer evolutionary games under aspiration-driven dynamics, and the specific application depends on a_k and b_k. Therefore, the combination of equations (1)–(4), (17), and (18) forms the average

abundance function of unmanned swarm under the framework of multiplayer public goods evolutionary game.

5. Evolutionary Game Analysis

On the basis of the average abundance of unmanned swarm obtained above, we will analyze the impact of cost c, multiplication factor r, and aspiration level α on it. Set the basic parameters $N = 100$, $d = 15$, $c = 1$, $r = 1.3$, and $\alpha = 1$, and when calculating the impact of one parameter, others remain unchanged. In addition, in order to highlight the different influence degree of parameters on average abundance under different selection intensities, $\omega = 0, 5, 10, 15, 20$ is selected in each simulation scenario.

5.1. Average Abundance With Respect to Cost. It can be easily proved through mathematical induction from equations (1) and (2) that increasing c will increase a_k and b_k and then increase $\pi_A(i)$ and $\pi_B(i)$, resulting in the decrease of both T_i^+ and T_i^-. Since $h(i) = T_i^+/T_{i+1}^-$, in the case of increasing c, the change of $h(i)$ and X_A are difficult to directly determined. Next, we will give a set of numerical solutions to intuitively observe the interaction within a certain range, so as to reveal the influence of c on X_A through simulation.

Select the interval $c \in [0.9, 1.8]$ and draw the average abundance curve of the strategy A as follows.

As shown in Figure 2, as c increases, X_A will monotonically decrease; when $\omega = 0$, $X_A = 0.5$ (i.e., the proportion of collaborators and betrayers in the swarm is balanced), while $\omega \neq 0$, X_A increases with ω; moreover, with the decrease of ω, the influence degree of c on X_A increases: $\Delta\langle X_A(\omega = 20)\rangle \approx 0.028$, while $\Delta\langle X_A(\omega = 5)\rangle \approx 0.063$.

5.1.1. Conclusion 1. The increasing of cost will decrease the average abundance, especially when selection intensity is small.

5.2. Average Abundance with Respect to Multiplication Factor. Similarly, in the case of r increasing, the change of $h(i)$ and $X_A$cannot be determined only by deduction. Select the interval $r \in [0.9, 1.8]$ and draw the average abundance curve of the strategy A under different selection intensities as shown in Figure 3.

As the multiplication factor increases, X_A will monotonically decrease, which means the phenomenon of "free riding" appears, resulting in the weakening of cooperation and the decline of X_A; moreover, with the decrease of $\omega(\omega \neq 0)$, the influence of r on X_A increases: $\Delta\langle X_A(\omega = 20)\rangle \approx 0.014$, while $\Delta\langle X_A(\omega = 5)\rangle \approx 0.047$.

5.2.1. Conclusion 2. The increasing of multiplication factor will decrease the average abundance, especially when selection intensity is small.

5.3. Average Abundance with respect to Aspiration Level. Select the interval $\alpha \in [0.9, 1.4]$ and draw the average abundance curve of the strategy A under different selection intensities as shown in Figure 4.

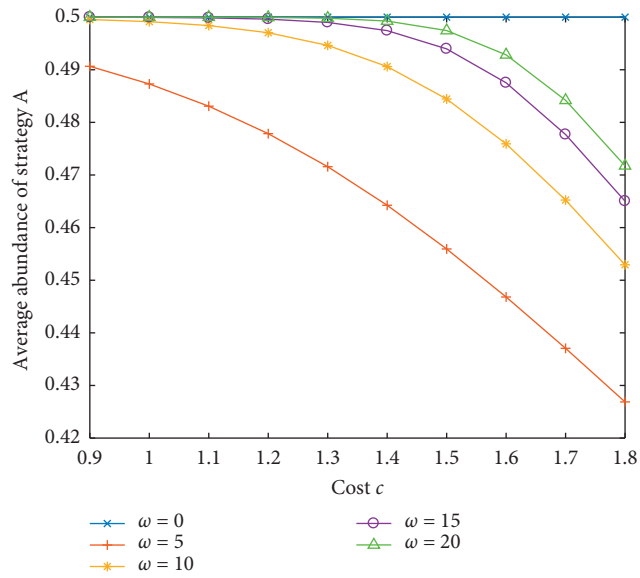


Figure 2: Average abundance with respect to cost.

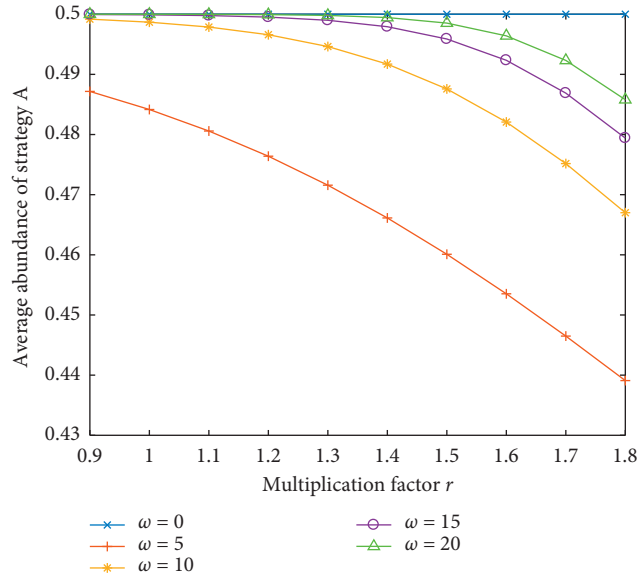


Figure 3: Average abundance with respect to multiplication factor.

As the aspiration level increases, X_A will monotonically increase, which means the rising of aspiration level makes it more difficult for betrayers to satisfy their expectations, and thus more betrayers transfer to cooperators; moreover, with the decrease of $\omega(\omega \neq 0)$, the influence degree of α on X_A increases: $\Delta\langle X_A(\omega = 20)\rangle \approx 0.002$, while $\Delta\langle X_A(\omega = 5)\rangle \approx 0.037$.

5.3.1. Conclusion 3. The increasing of aspiration level will increase the average abundance, especially when selection intensity is small.

According to the simulation results, c, r, and α have an impact on the curve trend of average abundance. When c and r increase, the average abundance decreases monotonically, while, with the increase of α, the average

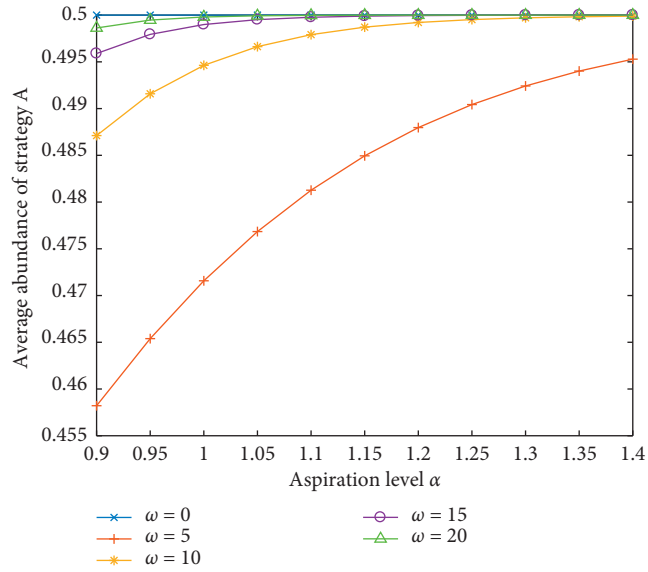


Figure 4: Average abundance with respect to aspiration level.

abundance increases monotonically. The conclusions from the simulation provide a theoretical basis for the regulation of swarms in practical application. Based on the conclusions above, in the following section, a case study is provided to further reveal the cooperative evolution mechanism of the unmanned swarm.

6. Case Study

Fire strike is a typical task in unmanned swarm operation. Limited by the ammunition loading/mounting capacity, when the unmanned swarm carries out the fire strike task in case of failure of centralized control mode, the "rational" unmanned units with intelligence and decision-making ability will strictly control the ammunition launching/delivery quantity with "free riding" mentality, while from the perspective of the whole swarm combat effectiveness, we hope that each unit can provide as much ammunition as possible to ensure the overall strike effectiveness on enemy. The key to coping with this contradiction is how to raise the proportion of cooperators in the swarm through self-regulation and self-coordination.

Consistent with the above section, set $N = 100$, $d = 15$, $c = 1$, $r = 1.3$, and $\alpha = 1$ and draw the basic curve (see Figure 5). Since $X_A < 0.5$, this case is a nondominant case; that is, most units choose strategy B. Therefore, we try to regulate relevant parameters to increase the average abundance of unmanned swarm and promote cooperation.

As in Figure 5, reducing the cost or increasing the aspiration level can raise the proportion of cooperators. However, increasing the multiplication factor will cause the average abundance curve to deviate downward from the basic curve, which is because increasing payoff of cooperators and betrayers by the same margin will make the "free riding" situation more serious. Consequently, we try to separate the multiplication factor of cooperators from that of betrayers, only increase the multiplication factor r_A of cooperators (the multiplication factor r_B of betrayers remains unchanged), and find that the average abundance curve deviates upward from the basic curve.

Furthermore, we simulate the average abundance under different r_A (see Figure 6). When $r_A = 2$, the average abundance is approximately equal to 0.5, which indicates that the proportion of cooperators and betrayers in the swarm is basically balanced. With the further increase of r_A, when $r_A = 2.65$, the average abundance will be greater than 0.5 at $\omega \approx 10$, while when $r_A = 2.9$, the average abundance will be greater than 0.5 at $\omega \approx 5$. Thus, we can reach the following conclusions.

(1) The adjustment on r_A can switch the dominant strategy, making the average abundance of strategy A greater than 0.5.

(2) The lower the ω is, the more stringent requirement for r_A will be, and the higher the ω is, the looser requirement for r_A will be: $X_A(\omega = 5, r_A = 2.9) > 0.5$, while $X_A(\omega = 10, r_A = 2.65) > 0.5$.

In order to investigate the regulation sensitivity of different parameters, we simulate the affecting degree of unit variation of c, α, and r_A on the average abundance. We select the simulation results with $\omega = 0, 5, 15$ to be discussed, as shown in Figures 7(a)–7(c), respectively.

(1) When $\omega = 0$, the average abundance is identically equal to 0.5, and thus the parameter regulation loses its effect (see Figure 7(a)).

(2) When $\omega \neq 0$ and unit variation of parameters (i.e., Δ) is small (note that the threshold of Δ is related to ω: $\Delta \approx 1.70|_{\omega=5}$, $\Delta \approx 1.53|_{\omega=15}$), the change in value of average abundance caused by adjusting α and c is much greater than adjusting r_A (see Figures 7(b) and 7(c)). The regulation of α and c is more sensitive than that of r_A.

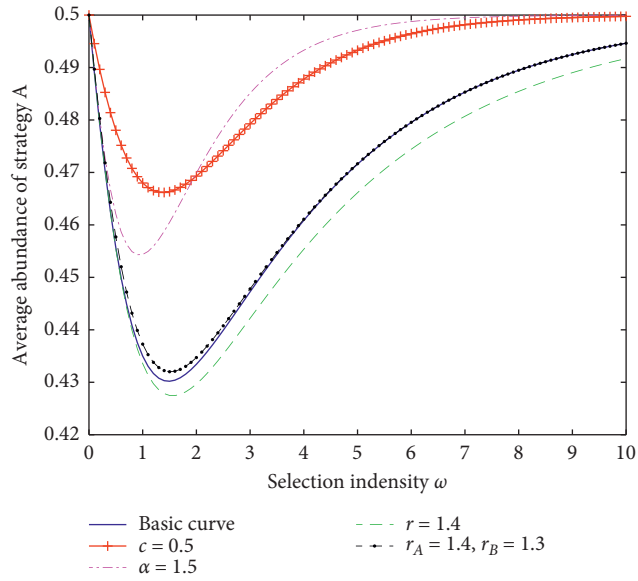


Figure 5: Effects of different factors on average abundance.

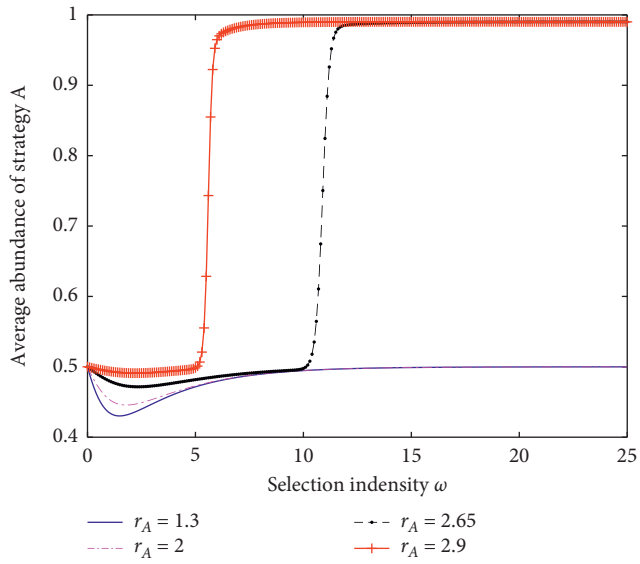


Figure 6: Effects of different r_A on average abundance.

(3) When $\omega \neq 0$ and Δ is large, the regulation effect of r_A is much better than that of α and c. And the larger ω is, the more sensitive r_A is; i.e., a small Δ_{r_A} leads to a large increasing in average abundance: $\Delta\langle X_A(\omega = 5, \Delta_{r_A} = 1.95)\rangle \approx 0.43$, while $\Delta\langle X_A(\omega = 20, \Delta_{r_A} = 1.55)\rangle \approx 0.48$ (see Figures 7(b) and 7(c)).

To improve the average abundance, the ideal measure is to increase the multiplication factor, reduce the cost of cooperators, or both. However, in order to ensure the effectiveness of the operation in the actual battlefield, the cost is difficult to reduce or even increase. Therefore, it is necessary to consider increasing both r_A and c. Figure 8 shows the change of average abundance when r_A and c increase at the same time (c increases by 50%, and r_A increases by 69% and 73%, respectively). Accordingly, as long as r_A increases by more than 73%, not only can the adverse effect of cost increasing on average abundance be offset, but also the cooperation in swarm can be promoted.

Unfortunately, the above regulation can only achieve a limited increase in the average abundance; that is, it cannot make the average abundance greater than 0.5. According to the conclusions from Figures 5 and 6, the conversion of dominant strategies (a large increase in average abundance) depends on a large selection intensity ω and a large unit variation Δ_{r_A}, and thus we further increase r_A under the premise of increasing c by 50% (see Figure 9). According to

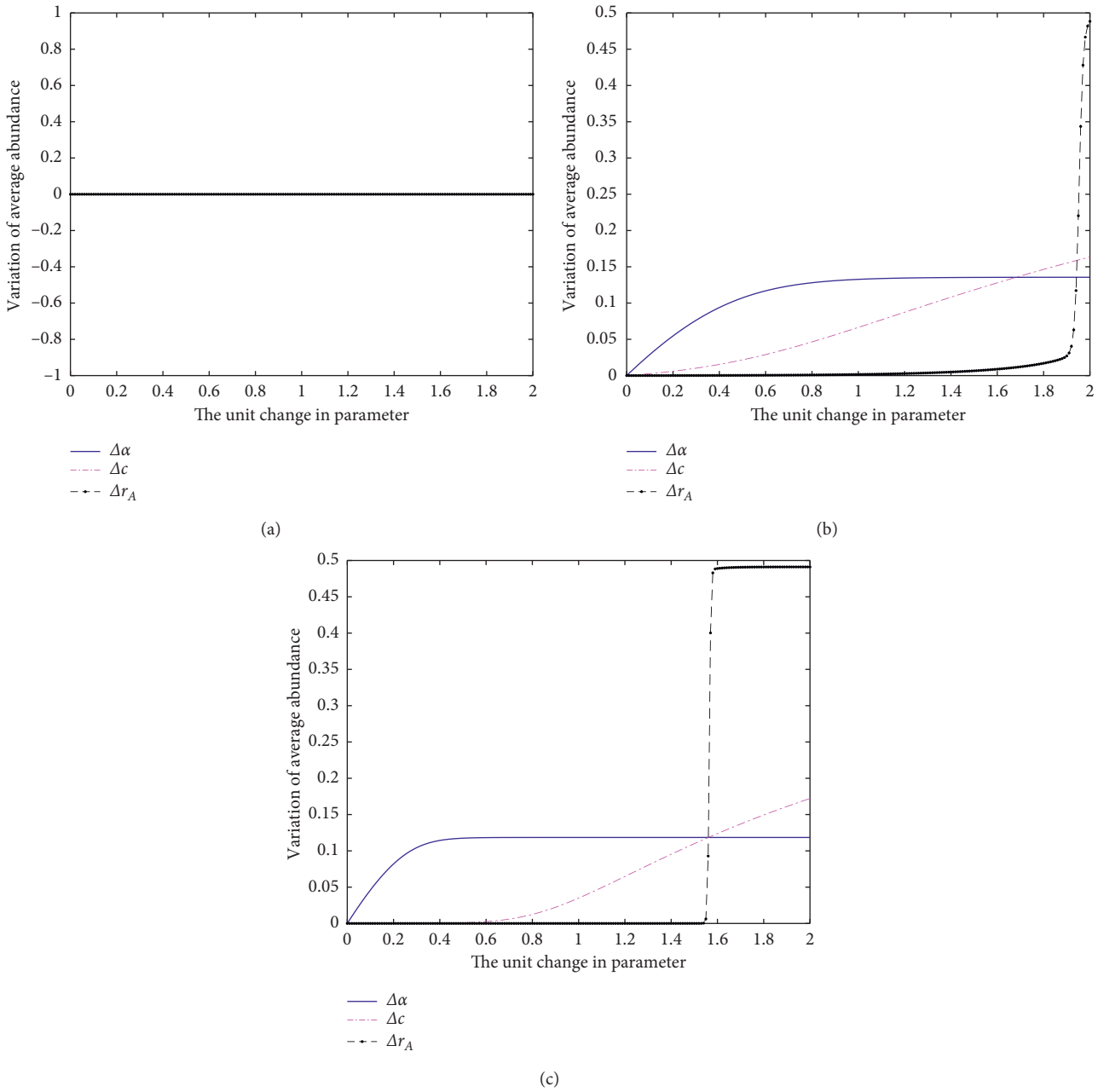


Figure 7: Effects of unit variation in parameters on average abundance. (a) $\omega = 0$. (b) $\omega = 5$. (c) $\omega = 15$.

the results in Figure 9, when $r_A = 2.52$ and $\omega \approx 15$, X_A will be greater than 0.5; when $r_A = 2.65$ and $\omega \approx 5$, X_A will be greater than 0.5.

The increase of r_A means that the hitchhiker will no longer get as much payoff as the cooperator, and the decrease of payoff will directly increase the strategy update probability $P_{B \longrightarrow A}$, so then more units tend to cooperate (more betrayers transfer to cooperators).

According to the above simulation results and conclusions, we can consider the following measures from two dimensions of management and technology in the actual control of unmanned swarm cooperation:

(1) Increase the multiplication factor value r_A of cooperators as much as possible. For example, with the help of advanced management means, for each combat unit in the swarm, its investment (i.e., cost c) in previous operations can be accumulated, and those with higher cumulative investment will be given more supplies (e.g., ammunition) or higher supply priority in the follow-up operations.

(2) Minimize the cost c for each operation. For example, with the help of advanced technology means, improve the reliability and survivability of combat units or the strike accuracy and damage-power of ammunition.

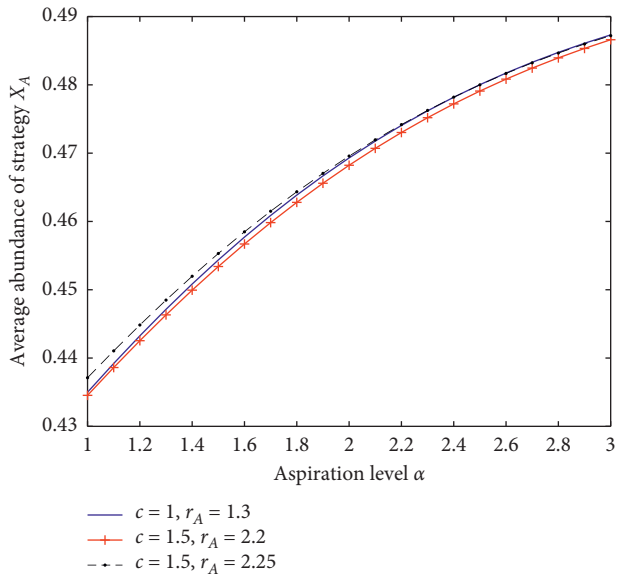


Figure 8: Effect of increase in cost and multiplication factor on average abundance.

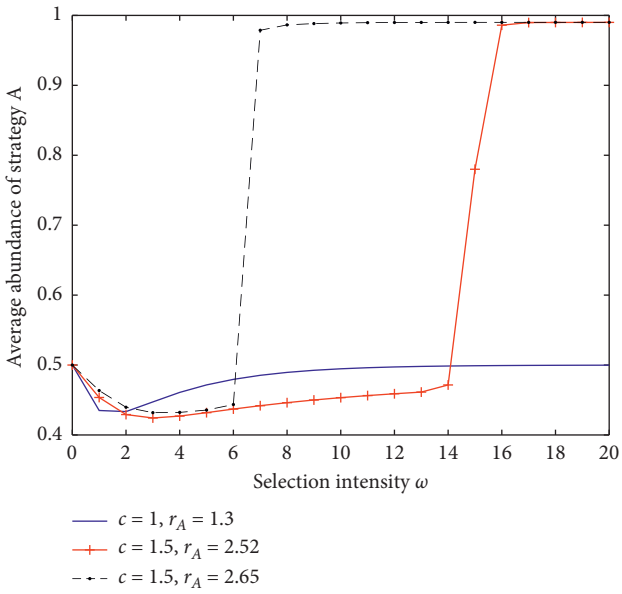


Figure 9: Strategy alternation by increasing cost and multiplication factor.

In addition, since c,r, and α are closely related to specific operation tasks, it is also necessary to discuss specific control measures in combination with operation tasks under the limitation of parameter value range.

7. Conclusion

The advantage of unmanned swarm operation lies in its autonomy that it can continually conduct cooperative operation efficiently in case of combat unit damage or communication failure. This work aims at the autonomous collaboration of the unmanned swarm under the failure of centralized control mode and proposes a cooperative evolution mechanism within the framework of multiplayer public goods evolutionary game. We get the average abundance function by theoretical derivation and then simulate the influence of different parameters (i.e., c, r, and α) on the abundance. The simulation results of unmanned swarm fire attack show that increasing the multiplication factor r_A and reducing the cost c can improve the average abundance of cooperators; furthermore, when the unit variation Δ is large, r_A not only has a high regulation sensitivity, but also can realize the switching of the dominant strategy. Finally, we suggest some proposals to provide an exploration for the transformation from theory to application.

The evolution of cooperation is a fascinating topic that has been studied from different perspectives and theoretical approaches. Our approach by means of multiplayer public goods evolutionary game sheds new light on how to study and analyze the evolution of cooperation in the unmanned swarm. In our work, we assume that the units in the swarm are homogeneous, which indicates a globally consistent α in the process of strategy updating. However, in reality, different units (firepower units, intelligence units, etc.) probably have various requirements for α. Thus, how to get the average abundance and explore the cooperative evolution mechanism when multiple α coexist will be our further work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 71901217); the National Key R&D Program of China (no. 2018YFC0806900); China Postdoctoral Science Foundation funded Project (no. 2018M633757); and the Primary Research and Development Plan of Jiangsu Province under Grant no. BE2019762.

References

[1] P. Xu, G. Dherbomez, E. Hery, A. Abidli, and P. Bonnifait, "System architecture of a driverless electric car in the grand cooperative driving challenge," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 1, pp. 47–59, 2018.

[2] Y. Ma, M. Hu, and X. Yan, "Multi-objective path planning for unmanned surface vehicle with currents effects," *ISA Transactions*, vol. 75, pp. 137–156, 2018.

[3] R. Shakeri, M. A. Al-Garadi, and A. Badawy, "Design challenges of multi-UAV systems in cyber-physical applications: a comprehensive survey, and future directions," *IEEE Communications Surveys & Tutorials*, vol. 99, 2019.

[4] J. Fan, D. Li, and R. Li, "Analysis on MAV/UAV cooperative combat based on complex network," *Defence Technology*, vol. 16, no. 1, pp. 150–157, 2019.

[5] J. Quan, Y. Q. Chu, and X. J. Wang, "Public goods with punishment and the evolution of cooperation," *Systems Engineering—Theory &Practice*, vol. 39, no. 1, pp. 141–149, 2019.

[6] J. M. Smith and G. R. Price, "The logic of animal conflict," *Nature*, vol. 246, no. 5427, pp. 15–18, 1973.

[7] M. A. Nowak and K. Sigmund, "Evolution of indirect reciprocity," *Nature*, vol. 437, no. 6685, pp. 1291–1298, 2005.

[8] E. Lieberman, C. Hauert, and M. A. Nowak, "Evolutionary dynamics on graphs," *Nature*, vol. 433, no. 7023, pp. 312–316, 2005.

[9] M. A. Nowak, "Five rules for the evolution of cooperation," *Science*, vol. 314, no. 5805, pp. 1560–1563, 2006.

[10] I. Wolef, "What are the equilibria in public-good experiments?" *Economics Letters*, vol. 150, pp. 83–85, 2017.

[11] M. A. Nowak, C. E. Tarnita, and T. Antal, "Evolutionary dynamics in structured populations," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365, no. 1537, pp. 19–30, 2010.

[12] J. Pena, B. Wu, and J. Arranz, "Evolutionary games of multiplayer cooperation on graphs," *PLoS Computational Biology*, vol. 12, no. 8, 2016.

[13] Q. Su, A. McAvoy, and L. Wang, "Evolutionary dynamics with stochastic game transitions," 2019, https://arxiv.org/abs/1905.10269.

[14] L. Hindersin, B. Wu, and A. Traulsen, "Computation and simulation of evolutionary game dynamics in finite populations," *Scientific Reports*, vol. 9, no. 1, p. 6946, 2019.

[15] H. Ohtsuki, C. Hauert, E. Lieberman, and M. A. Nowak, "A simple rule for the evolution of cooperation on graphs and social networks," *Nature*, vol. 441, no. 7092, pp. 502–505, 2006.

[16] F. Fu, M. A. Nowak, and C. Hauert, "Invasion and expansion of cooperators in lattice populations: prisoner's dilemma vs. snowdrift games," *Journal of Theoretical Biology*, vol. 266, no. 3, pp. 358–366, 2010.

[17] W. Maciejewski, F. Fu, and C. Hauert, "Evolutionary game dynamics in populations with heterogenous structures," *PLoS Computational Biology*, vol. 10, no. 4, 2014.

[18] B. Fotouhi, N. Momeni, B. Allen, and M. A. Nowak, "Evolution of cooperation on large networks with community structure," *Journal of the Royal Society Interface*, vol. 16, no. 152, Article ID 20180677, 2019.

[19] T. Josef, P. Andreas, and C. Krishnendu, "Population structure determines the tradeoff between fixation probability and fixation time," *Communications Biology*, vol. 2, 2019.

[20] B. Allen, G. Lippner, and M. A. Nowak, "Evolutionary games on isothermal graphs," *Nature Communications*, vol. 10, no. 1, pp. 1–9, 2019.

[21] D. S. Marta, L. P. Flávio, and C. S. Francisco, "Dynamics of N-person snowdrift games in structured populations," *Journal of Theoretical Biology*, vol. 315, pp. 81–86, 2012.

[22] P. P. Li, J. Ke, and Z. Lin, "Cooperative behavior in evolutionary snowdrift games with the unconditional imitation rule on regular lattices," *Physical Review E*, vol. 85, no. 2, p. 21111, 2012.

[23] X. J. Wang and K. Xia, "Extended average abundance function of multi-player snowdrift evolutionary game under aspiration driven rule," *Systems Engineering —Theory &Practice*, vol. 39, no. 5, pp. 1128–1137, 2019.

[24] C. E. Tarnita, H. Ohtsuki, T. Antal, F. Fu, and M. A. Nowak, "Strategy selection in structured populations," *Journal of Theoretical Biology*, vol. 259, no. 3, pp. 570–581, 2009.

[25] J. M. Du, B. Wu, and L. Wang, "Aspiration dynamics in structured population acts as if in a well-mixed one[J]," *Scientific Reports*, vol. 5, p. 8014, 2015.

[26] M. Yu, M. He, and D. Zhang, "An approach to coordinated control of structured unmanned swarm based on evolutionary game," in *Proceedings of the 2020 3rd IEEE International Conference on Unmanned Systems*, Harbin, China, November 2020.

[27] M. Yu, D. Zhang, and K. Kang, "Cooperative evolution mechanism of unmanned cluster based on multi-player public goods evolutionary game," *System Engineering&Electronics*, vol. 42, no. 12, pp. 2787–2794, 2020.

[28] Q. Liu, M. He, and J. Liu, "A mechanism for identifying and suppressing the emergent flocking behaviors of UAV swarms," *Acta Electronica Sinica*, vol. 47, no. 2, pp. 374–381, 2019.

[29] L. Gu, M. He, and G. Chen, "Research on unmanned aerial vehicle swarm system resilience," *System Engineering & Electronics*, vol. 43, no. 1, pp. 156–162, 2020.

[30] Q. Liu and M. He, "Research on the requirements and countermeasures of counter equipment for small UAV swarm at sea," *Military Operations Research and Systems Engineering*, vol. 33, no. 4, pp. 59–65, 2019.

[31] X. Sui, R. Cong, and K. Li, "Evolutionary dynamics of N-person snowdrift game," *Physics Letters A*, vol. 379, no. 45-46, pp. 2922–2934, 2015.

[32] J. Peña, B. Wu, and A. Traulsen, "Ordering structured populations in multiplayer cooperation games," *Journal of the Royal Society Interface*, vol. 13, no. 114, Article ID 20150881, 2016.

[33] B. Wu, A. Traulsen, and C. Gokhale, "Dynamic properties of evolutionary multi-player games in finite populations," *Games*, vol. 4, no. 2, pp. 182–199, 2013.

[34] J. Du, B. Wu, P. M. Altrock, and L. Wang, "Aspiration dynamics of multi-player games in finite populations," *Journal of the Royal Society Interface*, vol. 11, no. 94, Article ID 20140077, 2014.

[35] J. Du, B. Wu, and L. Wang, "Aspiration dynamics and the sustainability of resources in the public goods dilemma," *Physics Letters A*, vol. 380, no. 16, pp. 1432–1436, 2016.

[36] J. M. Du, B. Wu, and L. Wang, "Evolutionary game dynamics of multi-agent cooperation driven by self-learning," in *Proceedings of the 2013 9th Asian Control Conference (ASCC)*, pp. 1–6, IEEE, Istanbul, Turkey, June 2013.

[37] B. Wu, P. M. Altrock, and L. Wang, "Universality of weak selection," *Physical Review E*, vol. 82, no. 4, p. 46106, 2010.

[38] K. Xu, K. Li, R. Cong, and L. Wang, "Cooperation guided by the coexistence of imitation dynamics and aspiration dynamics in structured populations," *EPL (Europhysics Letters)*, vol. 117, no. 4, p. 48002, 2017.

[39] X. J. Wang, C. L. Gu, and S. J. Lv, "Evolutionary game dynamics of combining the Moran and imitation processes," *Chinese Physics B*, vol. 28, no. 2, pp. 1–13, 2019.

[40] Y.-S. Chen, H.-X. Yang, and W.-Z. Guo, "Aspiration-induced dormancy promotes cooperation in the spatial prisoner's dilemma games," *Physica A: Statistical Mechanics and Its Applications*, vol. 469, pp. 625–630, 2017.

[41] T. Platkowski, "Aspiration-based full cooperation in finite systems of players," *Applied Mathematics and Computation*, vol. 251, pp. 46–54, 2015.

[42] X. J. Wang and J. Liu, "Stable coalition structures in cooperative game with externalities," *Systems Engineering—Theory & Practice*, vol. 38, no. 5, pp. 1173–1183, 2018.
[43] X. Liu, M. He, Y. Kang, and Q. Pan, "Fixation of strategies with the Moran and Fermi processes in evolutionary games," *Physica A: Statistical Mechanics and Its Applications*, vol. 484, pp. 336–344, 2017.
[44] X. Liu, M. He, and Y. Kang, "Aspiration promotes cooperation in the prisoner's dilemma game with the imitation rule," *Physical Review E*, vol. 94, no. 1, p. 12124, 2016.
[45] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley Publishing Company, Boston, MA, USA, 2nd edition, 1994.
[46] G. Szabó and C. Tőke, "Evolutionary prisoner's dilemma game on a square lattice," *Physical Review E*, vol. 58, no. 1, pp. 69–73, 1998.
[47] M. A. Nowak and K. Sigmund, "Evolutionary dynamics of biological games," *Science*, vol. 303, no. 5659, pp. 793–799, 2004.
[48] N. G. Van Kampen, *Stochastic Process in Physics and Chemistry*, Elsevier, Amsterdam, Netherlands, 1992.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5585212, 13 pages
https://doi.org/10.1155/2021/5585212

Research Article

Combined Auxiliary Networks and Bird's Eye View Method for Real-Time Multicategory Object Recognition

Zhangpeng Gong,[1] **Luansu Wei**,[2] **Guoye Wang**,[1] **Dongxin Xu**,[1] **and Chang Ge**[1]

[1]*College of Engineering, China Agriculture University, Beijing 100083, China*
[2]*Jiangsu Key Laboratory of Construction Materials, School of Materials Science and Engineering, Southeast University, Nanjing 211189, China*

Correspondence should be addressed to Guoye Wang; wgy1615@126.com

Received 13 January 2021; Revised 16 February 2021; Accepted 28 February 2021; Published 9 March 2021

Academic Editor: Xiao Ling Wang

Object recognition based on LIDAR data is crucial in automotive driving and is the subject of extensive research. However, the lack of accuracy and stability in complex environments obstructs the practical application of real-time recognition algorithms. In this study, we proposed a new real-time network for multicategory object recognition. The manually extracted bird's eye view (BEV) features were adopted to replace the resource-consuming 3D convolutional operation. Besides the subject network, we designed two auxiliary networks to help the network learn the pointwise features and boxwise features, aiming to improve the category and bounding boxes' accuracy. The KITTI dataset was adopted to train and validate the proposed network. Experimental results showed that, for hard mode, the total average precision (AP) of the category reached 97.4%. For an intersection over a union threshold of 0.5 and 0.7, the total AP of regression reached 93.2% and 85.5%; especially, the AP of car's regression reached 95.7% and 92.2%. The proposed network also showed consistent performance in the Apollo dataset with a processing duration of 37 ms. The proposed network exhibits stable and robust object recognition performance in complex environments (multiobject, unordered objects, and multicategory). And it shows sensitivity to occlusion of the LIDAR system and insensitivity to close large objects. The proposed multifunction method simultaneously achieves real-time operation, high accuracy, and stable performance, indicating its great potential value in practical application.

1. Introduction

Autonomous driving is a futuristic technology that will transform mobility industries and ease the burden of driving. Autonomous driving is currently supported by relatively mature planning, decision-making, and algorithm implementation but is mainly hindered by its poor perception. As an efficient and precise remote sensing technique, the LIDAR systems have been widely applied in real-time intelligent systems, such as self-driving vehicles [1, 2]. The data acquired from LIDAR are point clouds, which is a set of points containing coordinates and other feature-related information, such as reflectivity. Detecting objects accurately within a point cloud is crucial and has been a widespread research subject. However, the key challenge is that the raw point cloud data are irregular, unstructured, and unordered. Consequently, specific processing methods that require data with a regular form are not suitable for direct application.

The convolution operation is an efficient approach for extracting deep features [3–7], and it requires a regular grid as the input, which a point cloud does not satisfy. Therefore, the first step is to transform the unstructured point cloud into a regular style. The structured processed data can be graphical [8–12], ordered points [13–17], or voxels [18–22]. In graph-based methods, nodes represent points, and edges represent the relationships between points. The abstract expression is obscure. Although point-based methods can achieve better performance by taking the raw point clouds as their input and predicting bounding boxes based on each point, in general, their inference time cannot meet the demands of a real-time system. Therefore,

they are restricted primarily to offline analysis. Voxels are popular because they have a clear physical structure similar to images. VoxelNet [20] is an example of a classic voxelization method that performs impressively in 3D object recognition tasks. Its strong performance relies heavily on several 3D convolution operations, resulting in a time- and memory-consuming process. To avoid using 3D convolution operations, a structure that replaces 3D voxels with pillars, thereby erasing the vertical dimension, has been reported [21]. This method was also called the bird's eye view method, which led to an improved processing speed, although the performance was unstable due to the lack of vertical information. Due to the lack of color information, the unstable performance is more serious than the image recognition task [22]. Alternatively, it should be a compromise approach to keep the necessary information concisely by using the maximum height, the density of the point set, and the reflectivity of the highest point to express the pillar feature [23].

To achieve higher precision of the bounding boxes, RBG images are fused with LIDAR data [24], thus obtaining a richer expression of the environment. The introduction of camera data means that this method is based on the trigger consistency of two kinds of data [25, 26] and the calibration accuracy of the camera and the LIDAR coordinate system, which may cause robustness problems in practical applications. Inspired by the better performance of point-based methods, an alternative method involves aggregating the voxels into a small number of key points [27], thus combining the advantages of both voxel- and point-based methods. In addition, this study adopted the farthest point sampling (FPS) to sample key points. However, FPS is extremely time-consuming, specifically for a large-scale scene, and the sampling time is not discussed in [27]. Therefore, finding an optimal balance between performance and processing time is still a challenge.

Most researchers use only a single category of data when training networks and assign independent evaluation indexes for the recognition effect of single categories. This method excludes the interference of other types of categories in the result. Furthermore, it causes deviation from the requirement that results in the recognition of multiple categories through one forward propagation in the actual application, which cannot explain the actual effect of the application.

This present study focused on developing a LIDAR-based 3D object recognition method for road scenes. Considering the significant effect of image recognition, we expect to take the advanced image recognition methods to the point cloud recognition task. Hence, the proposed method is a voxel-based recognition method that can simultaneously predict multiple object categories. We evaluated the method based on the 3D localization and bounding box precision, object recognition accuracy, and processing time. Unlike most present methods that heavily rely on 3D convolutional operations, we considered that the bird's eye view (BEV) based method has not yet exhausted its performance potential. Thus, we improved the head network and designed an additional auxiliary network to improve the prediction accuracy. The network was trained and evaluated by the KITTI dataset and its benchmark. The results verify that the new part is beneficial to the network.

The rest of this paper is structured as follows: Section 2 presents the proposed network architecture; Section 3 outlines the implementation of the proposed network and presents the results; Section 4 discusses the specific recognition effects that are not obvious in the evaluation indicators; and Section 5 presents our conclusions.

2. Methods

The proposed network is divided into preprocessing, backbone network, neck network, head network, and auxiliary networks:

(1) The preprocessing stage transforms the unordered point cloud into ordered data.

(2) The backbone and neck networks are used to extract scene features.

(3) The head network transforms the scene features into predicted outputs.

(4) The auxiliary network is set up to help the subject network learn pointwise and boxwise features. It does not participate in the prediction process, so it will not cause an additional computing burden to the network.

2.1. Preprocessing. The method outlined in [23] is referred to. First, the irregular points are transformed into a pillar map according to their location. Besides the three channels mentioned in [23], we add a channel containing the pillar's minimum height, which expresses the difference between the edge and the inside of an object. Therefore, four channels represent the vertical distribution of the points in each pillar: the first channel records the number of points in the pillar; the second and the third record the maximum and minimum vertical coordinates of the points in the pillar; and the fourth records the reflectivity of the highest point in the pillar. Finally, a four-channel bird's eye view (4C-BEV) is obtained as the network input. This method is essentially equivalent to taking the upper cover shell of the spatial point cloud from a top-down perspective. Because of Earth's gravity, very few objects are suspended in the air, and obstacles can usually be clearly distinguished by direct observation of such shells. The channels' values need to be normalized, specifically, the first channel, because the point cloud density increases from far to near. Here, the distance factor K_d is added to make pillars with a similar degree of characteristic expression at different locations:

$$C_{[x,y]}{}^{1} = \frac{\log\left(N_p * K_d + 1\right)}{\log(64)}, \tag{1}$$

where N_p represents the number of points in each pillar and K_d is expressed as

$$K_d = \frac{\sqrt{x^2 + y^2}}{k_e}, \tag{2}$$

where k_e is the coefficient.

As shown in Figure 1, through observation with the naked eye, the objects are visible in the 4C-BEV, indicating that this method can preserve the point cloud's information in a vertical direction while compressing the data efficiently.

2.2. Backbone Network. The 4C-BEV is entirely consistent with the image in terms of data structure. Therefore, many popular backbone networks for image recognition can be used directly, such as ResNets [28], CSPDarknet53 [3], or VGG16 [29]. Besides, there are some differences between the object recognition tasks in image and point clouds. First, multiple scales are not necessary. In the image recognition task, the perspective phenomenon is one of the main factors requiring consideration in the network design. Therefore, the network contains output nodes representing different scales, or several preexisting boxes are predefined to represent different scales. When constructing the feature map using the LIDAR coordinates, as the objects have the same size as the real world, this perspective phenomenon is not encountered. Second, the scales of objects are different. In the image recognition task, in general, the recognition performance varies between large- and small-scale objects when using the same network. Typically, the area of interest appears near the observer, which means the identification accuracy of large targets is more important than others. The image passes through a multilayer network, which significantly reduces its scale and improves the recognition ability regarding large-scale objects. Taking CSPDarknet53 [3] as an example, after an input image was transmitted forward, the scales of the three outputs were reduced by 8, 16, and 32 times, respectively. Using the mentioned encoding approach, the feature vectors at each position can fuse with the features of the broader receptive field, identifying large-scale objects. However, for LIDAR-based tasks, the scene's object is relatively small compared to the scene size, with the large-scale output feature map affecting the recognition accuracy. Third, the orientation of the bounding boxes needs to be predicted. The maximum pooling layers play an essential role in a backbone network because they can prevent overfitting and improve the network's generalization ability; however, they can also enhance the rotation invariance.

The backbone network architecture is more similar to a tiny version of CSPDarknet53. Figure 2 illustrates the modified architecture. We use Conv (k, s, p, c_{out}) to represent a 2D convolutional operator, where c_{out} is the number of output channels; k, s, and p are the kernel size, stride, and padding size, respectively. The "Conv" operation contains a 2D convolutional operator, a group normalization (GN) layer, and an activation function layer sequentially when it acts as a convolutional middle layer. We used several small residual blocks to fuse features of the current layer and the previous layer. Then, we used big residual blocks to fuse shallow features and deep features. The nodes of the backbone network measure $h \times w \times c$, where h and w are the spatial dimensions, and c is the channel dimension. The input is a feature map with a fixed size of $h \times w \times 3$. The backbone network has two outputs: one with a fixed size of $h/2 \times w/2 \times 128$, while the other has a fixed size of $h/4 \times w/4 \times 512$.

2.3. Neck Network. The role of the neck network is to perform further feature extraction and connect the backbone to the head. Figure 3 shows the architecture of our neck network. The residual blocks are retained to aid further feature extraction. Upsampling operations are used to unify the scale of the feature map. Although there is no perspective phenomenon, different categories of objects have different sizes, and multiscale features play a positive role in the network.

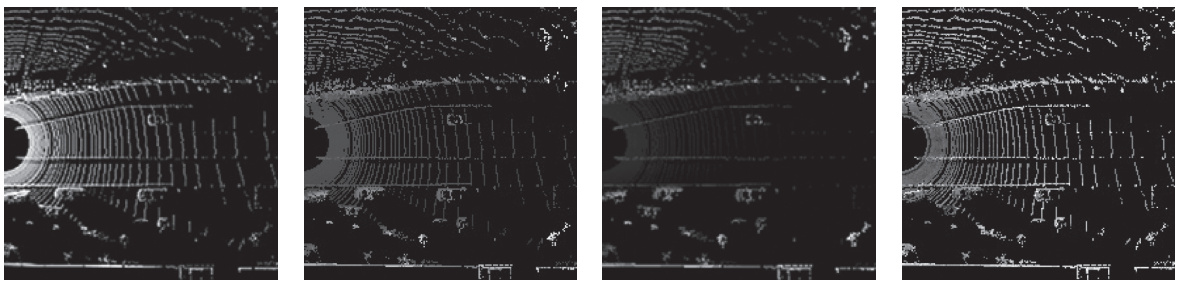
2.4. Head Network. The head network is custom-designed for our specific 3D object recognition task and divided into three parts. The first part is used for confidence prediction, with the sigmoid function used to limit the result range to [0, 1]:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}. \tag{3}$$

Two channels are assigned to each category, representing the regression confidence of this category based on horizontal and vertical anchors.

The second part is used for predicting bounding boxes. The spatial position and physical dimension are predicted in this part. As there is no perspective effect, it is reasonable that bounding box regression based on the standard reference value should arise. Therefore, we predefined an anchor map as the standard reference value in which each position has $2 \times N_c$ anchors, where N_c is the number of predicted categories. In general, the orientation and border predictions are conducted simultaneously [20, 21, 23]. This method cannot express the close relationship between the two ends of the interval. Inevitably, they produce the greatest divergence, which is incorrect. To keep the prediction of orientation continuous, we adopt an anchor-free [30] and anchor-based [20, 21, 23] combined method. Six channels are assigned to represent the regression parameters (except for orientation) of the two anchors at each position. Furthermore, the sine and cosine values are used to represent the orientation indirectly.

In most studies, there is little discussion on multicategory object prediction. By default, when predicting multicategory objects, the regression parameters for all categories are given, which leads to low information utilization (only $1/N_c$ information is useful). Thus, the convergence efficiency is greatly affected. In this study, it is designed to give only a set of border predictions at each position. The box center's category is determined according to the ground truth. The other positions' categories are determined by the overlap between the standard anchor and the ground truth bounding box.



(a) (b) (c) (d)

Figure 1: Four-channel bird's eye view: (a) the first channel, (b) the second channel, (c) the third channel, and (d) the fourth channel.

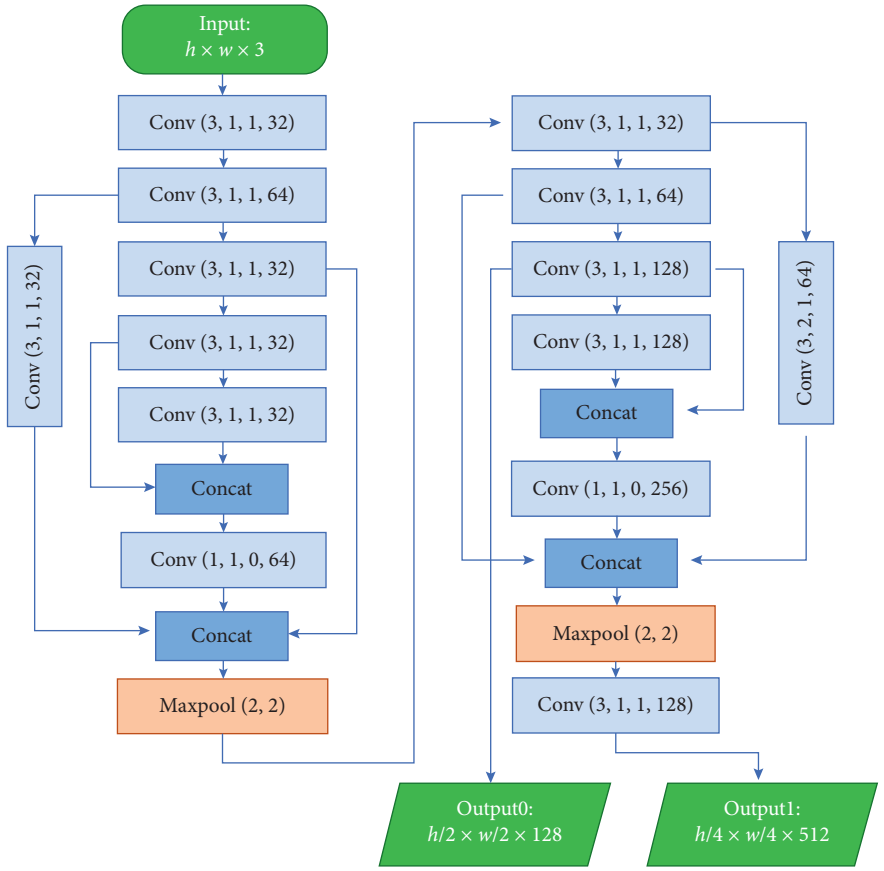


Figure 2: Architecture of the backbone network. The green blocks are input or output nodes (size labeled inside). The shallow blue blocks are the Conv operating set, and the deep blue blocks are the fusion process. The orange blocks are the maximum pooling layers. The values in parentheses indicate the filter size and stride.

For convenience, it is assumed that the category is determined, and there are two anchors for each position. The ground truth of the bounding box regression value R_{gt} of one anchor at each location can be expressed as follows:

$$R_{gt} = \left[\Delta x, \Delta y, \Delta z, \Delta h, \Delta w, \Delta l, \ \sin\theta_{gt}, \ \cos\theta_{gt}\right]^T, \tag{4}$$

$$\theta_{gt} = \arctan\frac{\sin\theta_{gt}}{\cos\theta_{gt}}, \tag{5}$$

$$\mathbf{A} = \left[x_a, y_a, z_a, h_a, w_a, l_a, \theta_a\right]^T, \tag{6}$$

$$\Delta x, \Delta y, \Delta z = \frac{x_{gt} - x_a}{d_a}, \frac{y_{gt} - y_a}{d_a}, \frac{z_{gt} - z_a}{h_a}, \tag{7}$$

$$d_a = \sqrt{w_a^2 + l_a^2}, \tag{8}$$

$$\Delta h, \Delta w, \Delta l = \log\left(\frac{h_{gt}}{h_a}\right), \ \log\left(\frac{w_{gt}}{w_a}\right), \ \log\left(\frac{l_{gt}}{l_a}\right), \tag{9}$$

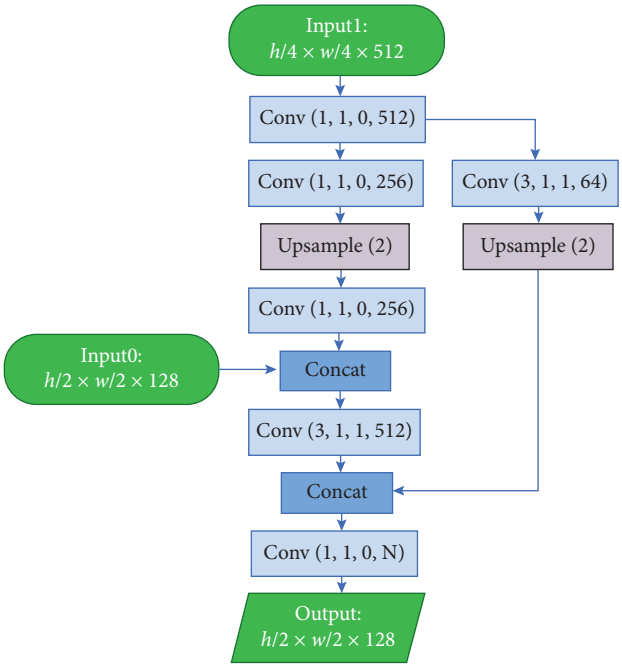


FIGURE 3: Architecture of the neck network. The gray blocks represent upsampling operations (the values in parentheses indicate the magnification rate).

where **A** denotes the parameter of one anchor in each position.

The third part is used for category prediction, for which N_c channels are assigned. The softmax function is used to transform the result to N_c probabilities, whose range is limited to [0, 1]:

$$\operatorname{softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=0}^{N_c} e^{x_j}}. \tag{10}$$

2.5. Auxiliary Network. Because of the ability to obtain more detailed pointwise characteristics, point-based methods usually achieve higher accuracy than voxel-based methods. To enhance our method's accuracy, the pointwise feature was introduced to the network. Referenced by the SA-SSD [31], the pointwise feature learning network was set as an auxiliary network that only works during training, does not play a role in predicting, avoiding additional computational overhead caused by the additional feature extracting. The penultimate layer of the neck network was set as the former feature extraction layer of the auxiliary network, which is ultimately a voxelwise category prediction network. The auxiliary network is elaborated in Figure 4. The accuracy of border regression is highly dependent on the accuracy of category prediction. Therefore, the primary task is to improve the accuracy of object category prediction by the increased category information of the point cloud. Unlike the category prediction part in the head network, which only focuses on the category prediction of the bounding boxes' center voxels, the auxiliary network focuses on the category prediction of the voxels around the bounding box center. Since each voxel contains only one highest point, voxel features are equivalent to pointwise features. We randomly extract no more than 1000 internal points and no more than 250 external points of bounding boxes to save memory space. We recreate the voxel category label, depending on whether its highest point is within the bounding box. The whole operation is similar to an additional "droop-out" process, which improves the generalization performance of the network. The second task of the auxiliary network is to enhance the accuracy of bounding box regression. In this step, we randomly sample no more than 50 highest points within each bounding box and calculate the inverse distance to the bounding box center as its weight. The weighted average and maximum pointwise features among all points in the bounding box region are combined to express a boxwise feature:

$$f_b = \left[\frac{\sum_{j=1}^{50} w_j(p_j) f_j}{\sum_{j=1}^{50} w_j(p_j)}, \max(f_0, \cdots, f_j) \text{ where } f_j \text{ is sampled}\right], \tag{11}$$

$$w_j(p_j) = \begin{cases} \dfrac{1}{1 + \|p_j - p_c\|_2}, & \text{if } p_j \text{ is sampled,} \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

2.6. Loss. The loss contains the central part and the auxiliary part. The central part contains confidence loss, regression loss, and category loss. The auxiliary part contains point category loss and box regression loss. We adopted the smooth$_{L1}$ function [5] to calculate the bounding box regression loss:

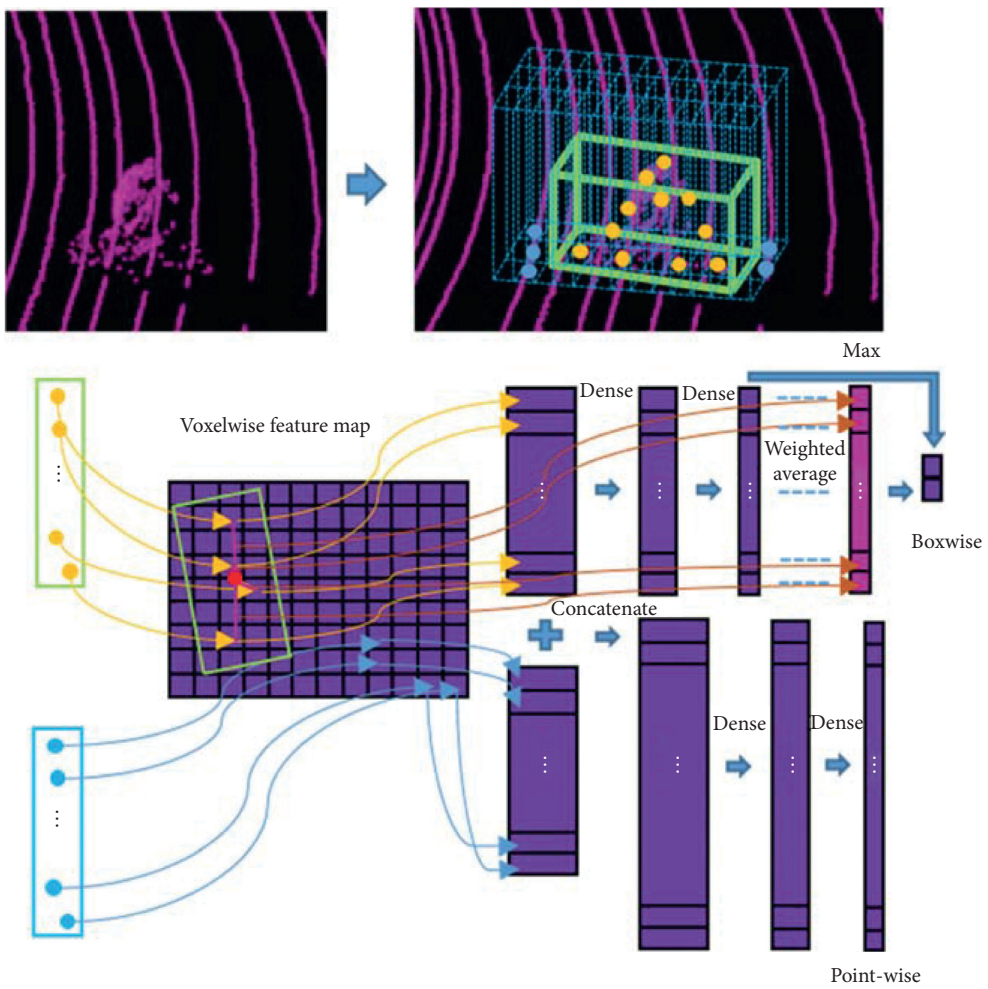


Figure 4: The architecture of the auxiliary network. The blue dash in the scene is the voxel boundary. The green box is the ground truth bounding box. The yellow points and blue points are the voxels' highest points, which are included and not included in the bounding boxes, respectively. The red point is the center of the ground truth bounding box. The boxwise feature is then followed by two fully connected layers (the operation is called "Dense" customarily), generating bounding box regression values similar to R_{gt} (see equation (4)).

$$\mathrm{SMOOTH}_{L1}\left(\mathbf{X}_p, \mathbf{X}_{gt}, a\right) = \frac{\sum_n \mathrm{smooth}_{L1}\left(\mathbf{X}_p - \mathbf{X}_{gt}, a\right)}{n}, \quad (13)$$

$$\mathrm{smooth}_{L1}(x, a) = \begin{cases} 0.5ax^2, & |x| < \frac{1}{a^2}, \\ |x| - \frac{0.5}{a}, & |x| \geq \frac{1}{a^2}. \end{cases} \quad (14)$$

Smooth_{L1} function has stable convergence characteristics in the case of large deviation and adequate convergence in small variation. The predictions of category and confidence are converted into the probability value prediction within the interval of [0, 1]. The cross-entropy function was applicable to calculate their losses:

$$\mathrm{CE}\left(\mathbf{X}_p, \mathbf{X}_{gt}\right) = -\frac{\sum_n \log\left(\mathbf{X}_p\right) * \mathbf{X}_{gt}}{n}, \quad (15)$$

where $\mathbf{X}_p$, $\mathbf{X}_{gt}$ are the predicted value and ground truth value, respectively.

The ground truth values of the category are labeled as a one-hot form. The focus loss [32] can solve the problem that when the proportion of positive and negative samples is unbalanced, the negative ones are submerged in the positive ones. Although the positive and negative labeled data distributions are incredibly uneven, the ratio of positive and negative samples is given. To avoid the focus loss affecting the rate of convergence, we do not adopt focus loss.

Not all losses in each position are calculated in a feature map. Some grids that are far from the center of the object are inaccurate and can be neglected. The positive confidence label is vital because it can be used as a mask to filter out untrusted data not to be included in the loss calculation. In Section 2.4, an anchor map was established. We excluded the angle parameters and determined the confidence by calculating the intersection over union (IoU) between the ground truth bounding box and the anchors in the map. Because the confidence feature map does not rely on the vertical direction position, the projection plane in the vertical direction of the ground truth bounding box and anchors are used when calculating the IoU:

The final loss L is defined as

$$L = L_{\text{conf}} + L_{\text{reg}} + L_{\text{cls}} + L_{\text{point}} + L_{\text{box}}. \tag{16}$$

Among the final loss, the confidence loss is expressed as

$$L_{\text{cof}} = \frac{\sum \mathbf{P}_{gt} * \text{CE}\left(\mathbf{P}_p, \mathbf{P}_{gt}\right)}{\sum \mathbf{P}_{gt}} + \frac{\sum \mathbf{N}_{gt} * \text{CE}\left(1 - \mathbf{P}_p, \mathbf{N}_{gt}\right)}{\sum \mathbf{N}_{gt}}, \tag{17}$$

where $\mathbf{P}_p$ represents the positive confidence prediction, $\mathbf{P}_{gt}$ denotes the positive corresponding ground truth, and $\mathbf{N}_{gt}$ denotes the negative corresponding ground truth.

The regression loss is expressed as

$$L_{\text{reg}} = \frac{\sum \mathbf{P}_{gt} * \text{SMOOTH}_{L1}\left(\mathbf{R}_p, \mathbf{R}_{gt}, a\right)}{\sum \mathbf{P}_{gt}}, \tag{18}$$

where $\mathbf{R}_p$ is the bounding box regression prediction and $\mathbf{R}_{gt}$ is the corresponding ground truth.

The category loss is expressed as

$$L_{\text{cls}} = \frac{\sum \mathbf{M}_{gt} * \text{CE}\left(\mathbf{C}_p - \mathbf{C}_{gt}\right)}{\sum \mathbf{M}_{gt}}, \tag{19}$$

where $\mathbf{M}_{gt}$ is the maximum last channel value of $\mathbf{P}_{gt}$.

The auxiliary parts of the loss are defined as

$$L_{\text{box}} + L_{\text{point}} = \frac{\sum_{n_b}\left(\sum \text{SMOOTH}_{L1}\left(\mathbf{B}_p, \mathbf{B}_{gt}, a\right) * \mathbf{W}_b / \sum \mathbf{W}_b\right)}{n_b} + \frac{\sum_{n_c} \text{CE}\left(\mathbf{C}^p{}_p, \mathbf{C}^p{}_{gt}\right)}{n_c}, \tag{20}$$

where $\mathbf{W}_b$ denotes the weights calculated by equation (12). $\mathbf{B}_p$, $\mathbf{C}^p{}_p$ denote the boxwise regression feature and category prediction of the sample point; and $\mathbf{B}_{gt}$, and $\mathbf{C}^p{}_{gt}$ denote the corresponding ground truth, respectively.

2.7. Dataset. Most 3D object recognition networks are trained using the KITTI dataset [33]. The KITTI dataset contains 7481 frames, among which we selected 2000 frames as the verification set and the remaining 5481 frames as the training set. We were interested in cars, trucks, vans, pedestrians, and cyclists among the object categories. Besides, trucks and vans were merged into one class. In this study, the Apollo dataset [34] was also adopted.

In contrast to the KITTI dataset, the Apollo dataset contains continuous frame data. When a vehicle turns, the surrounding objects show disordered orientations and spatial positions, which are more complicated than those in the nonturning state. This representative disordered data frequently occurs in continuous frame data. The Apollo dataset contains 16 scenes. Each scene contains 2–5 sections of continuous frame data collected at a frequency of 2 Hz, lasting for 1 min. We take a section of each scenario as the verification set and the rest as the training set. The final training set consisted of 3,943 frames, while the validation set consisted of 1,650 frames. We were interested in four types of labeled data: small vehicles, big vehicles, pedestrians, and riders (i.e., motorcyclists and bicyclists), which were labeled using the abbreviations "VEH, TRU, PED, CYC," respectively, during the visualization step. The data augmentation technique [34] was adopted during the training process.

2.8. Details. In this study, points inside the range covered from 41.6 m in front, 20.8 m left and right, and 2 m above and below the LIDAR coordinate system were used to construct the BEV feature map. The resolution of the grid was set as 0.2 m. Therefore, the BEV feature map was divided into 208×208 grids.

A minibatch gradient descent was conducted with a batch size of 1. We placed all batch normalization layers in the network with the group normalization layers because of the small batch size. Each training set involved in the training was defined as an epoch. Epochs were set as 100 and the first 75 epochs had a learning rate of 0.001, while the remaining epochs had a learning rate of 0.0001. All algorithms were run on a workstation with a Core i7 CPU, 8G RAM, an NVIDIA 1080Ti GPU, and the open-source deep learning framework TensorFlow. Nonmaximum suppression was deployed to filter out excess bounding boxes, with the IoU threshold set as 0.1.

2.9. Evaluation Indicators. It is assumed that n_p objects are predicted with n_{gt} objects labeled in the ground truth. First, the prediction object needs to be paired with the ground truth object by calculating the IoU. Considering the predicted objects as the benchmark and matching all predicted objects with the labeled objects using the maximum IoU, the calculated result is denoted as the precision. Similarly, considering the labeled objects as the benchmark and matching all labeled objects with the predicted objects using the maximum IoU, the calculated result is denoted as the recall. As the recall rises, the precision drops. Using recall as the horizontal axis and precision as the vertical axis, the area surrounded by the plotted recall-precision curve and the coordinate axis is the average precision (AP), which is widely adopted to evaluate the performance of the network. We set the parameters consistent with the KITTI benchmark, where the IoU threshold of cars is 0.7, and pedestrians and cyclists are 0.5.

3. Results and Discussion

3.1. Loss Curves. Validation was performed using one batch of data in the validation set per 100 iterations. During the training process, the loss function fluctuated violently with the weight decline of 0.99, representing the changing trend of the loss. The evolution of the loss curves throughout the training process is shown in Figure 5. By the time the

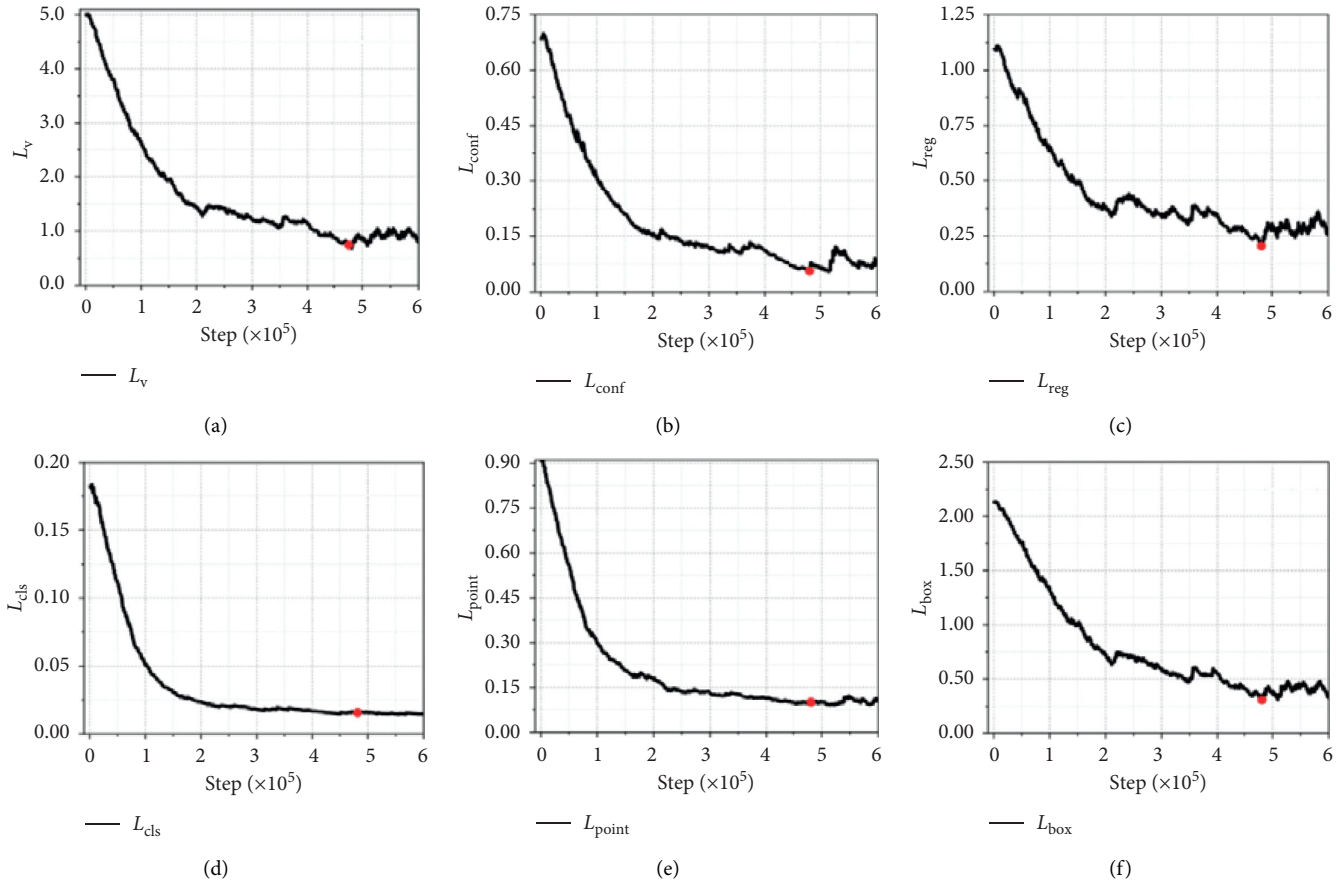


Figure 5: Loss curves for the training process: (a) validation loss L_v, (b) confidence loss L_{conf}, (c) regression loss L_{reg}, (d) category loss L_{cls}, (e) point loss L_{point}, and (f) box loss L_{box}.

iteration reaches 500,000, the network has converged. Our trained model was marked as red points in the figures.

3.2. Speed. The acquisition frequencies commonly used by LIDAR are 5, 10, 15, and 20 Hz. The entire recognition process is divided into preprocessing and inference. The mean preprocessing and inference process duration was approximately 5.7 and 31.2 ms, respectively. The mean recognition process duration was approximately 37 ms, which meets the real-time requirement.

3.3. Accuracy. The recall precision curves (trained by the KITTI dataset) are given in Figure 6. The APs of the regression and category are listed in Table 1 (trained by the KITTI dataset). The data we mainly focus on is marked in bold. The AP for cars was 92.5, which is relatively high, considering that the regression is based on the anchor determined by the category prediction. The trucks, which are not considered in the KITTI benchmark, are included in the identification. Due to the uneven distribution of training data, the AP of other categories is slightly lower. The total AP (0.5), the total AP (0.7), and the total AP (categories) can reach 93.2, 85.5, and 97.4, respectively. In Apollo dataset, the labeled bounding boxes are the objects' visible parts, which vary from the actual physical size, resulting in low indicators. As for apparent objects, using the Apollo dataset shows similar performance to the KITTI dataset, which is described in Section 3.5.

3.4. Comparison. The main contribution of this paper lies in the anchor-free and anchor combined prediction method and auxiliary networks specially designed for the bird's eye view network. The contrast effect is shown in Table 2. Due to the use of the data augmentation technique, the training results can be slightly different. Indicators within ±1% were regarded as the same performance. Our proposed method significantly improves the prediction results of pedestrians and cyclists, and the data we mainly focus on is marked in bold.

3.5. Scene Analysis. The method described in this paper was designed for a complex environment (multiobject, unordered objects, multicategory). In this section, some typical scenes are selected to analyze the network's performance concerning object recognition. The continuous frame data from the Apollo dataset provides sufficient verification of the stability of the recognition effect.

Figure 7 shows three frames of a congested traffic scene. This scene features many vehicles in a dense array, precisely what the proposed network has been designed to identify.

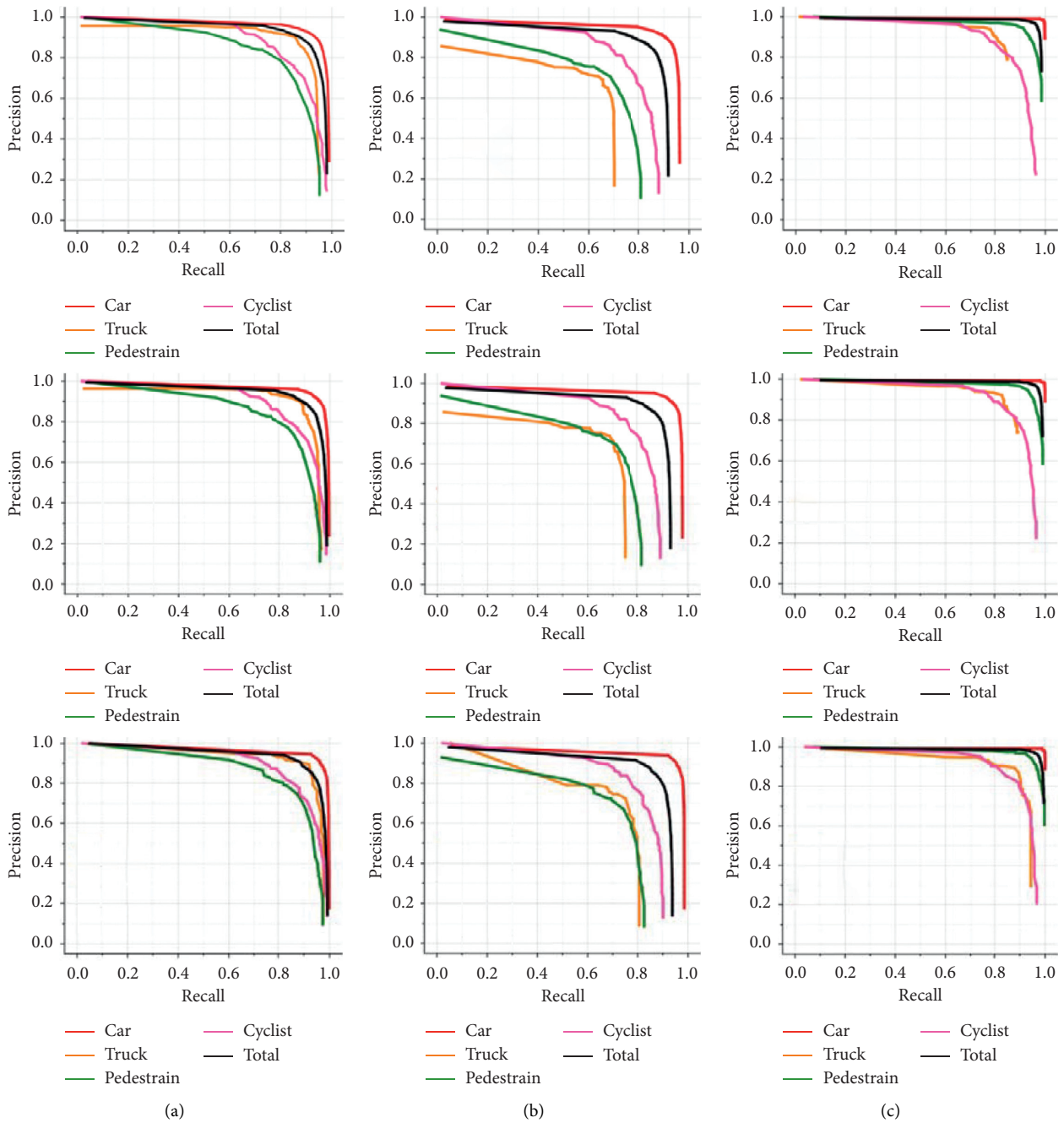


Figure 6: Recall-precision curves (trained by the KITTI dataset): (H, M, E) the hard, moderate, and easy modes separately; (a–c) the recall-precision curve for bounding box regression (IoU > 0.5), the recall-precision curve for bounding box regression (IoU > 0.7), and the recall-precision curve for category separately.

The recognition result corresponding to the turning scene is shown in Figure 8. Most objects are recognized accurately, and the performance remains stable. Despite the large size of the rotation scene, the direction of the objects was also predicted accurately.

Figure 9 depicts a scene containing several objects belonging to different categories, including a vehicle, truck, and cyclist. Each object was recognized consistently and accurately across the consecutive frames.

The recognition visualization of the KITTI dataset is shown in Figure 10, and a series of typical complex scenarios are selected, including multiobject, unordered objects, and multicategory.

3.6. Discussion. The orientation of the bounding box is expressed by sine and cosine values indirectly in this paper. Compared with the directly predicted method, when the

Table 1: Evaluation indicators trained/tested by the KITTI dataset.

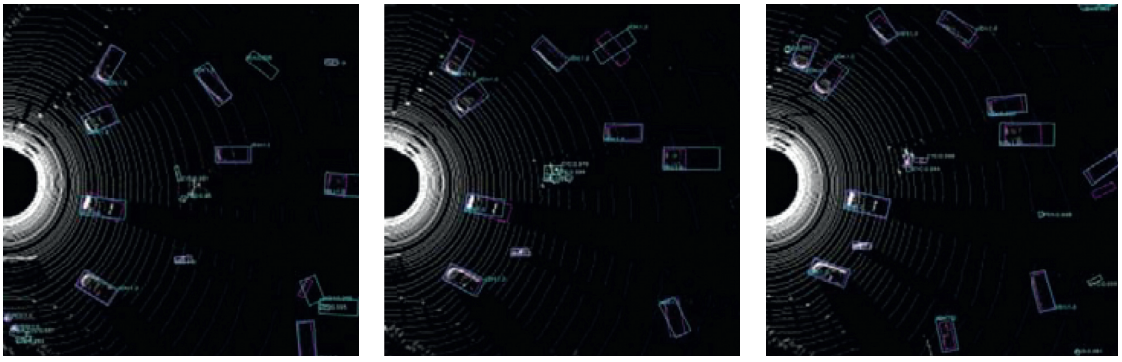
Category	AP (regression) IoU > 0.5			AP (regression) IoU > 0.7			AP (category)		
Mode	Hard	Moderate	Easy	Hard	Moderate	Easy	Hard	Moderate	Easy
Car	95.7	96.8	96.6	**92.2**	**94.1**	94.3	**99.1**	**99.2**	99.1
Truck	88.5	90.4	92.7	54.5	59.7	68.3	**82.5**	**85.9**	89.6
Pedestrian	**83.3**	**84.6**	86.3	64.3	65.4	66.4	**95.9**	**96.9**	97.8
Cyclist	**88.3**	**89.9**	90.2	78.3	80.2	81.2	**89.1**	**90.4**	91.1
Total	**93.2**	**94.1**	94.2	**85.5**	**87.0**	87.1	**97.4**	**97.8**	98.2

Table 2: Results of comparison trained/tested by the KITTI dataset.

Category	AP (regression) IoU > 0.5	AP (regression) IoU > 0.7	AP (category)	L_{point}	L_{box}	$\sin\theta/\cos\theta$
Vehicle	90.0	77.4	97.7			
Pedestrian	66.5	42.2	88.6			
Cyclist	68.2	32.5	73.2			
Vehicle	95.6	93.0	99.2			✓
Pedestrian	77.6	61.8	96.0			✓
Cyclist	85.1	68.3	83.1			✓
Vehicle	94.2	91.2	99.3	✓		✓
Pedestrian	81.4	63.2	95.3	✓		✓
Cyclist	86.7	79.2	87.8	✓		✓
Vehicle	95.7	92.2	99.1	✓	✓	✓
Pedestrian	**83.3**	64.3	95.9	✓	✓	✓
Cyclist	**88.3**	78.3	**89.1**	✓	✓	✓

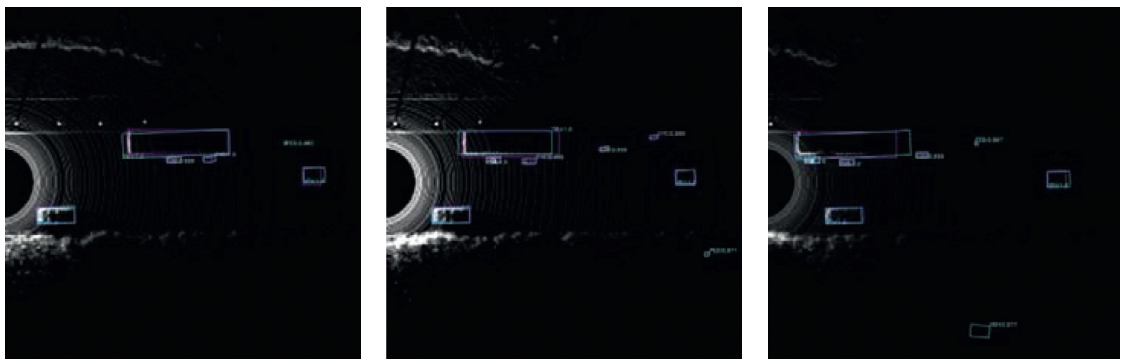
(a) (b) (c)

Figure 7: Scenes of congested traffic: (a–c) Three frames from a single scene.



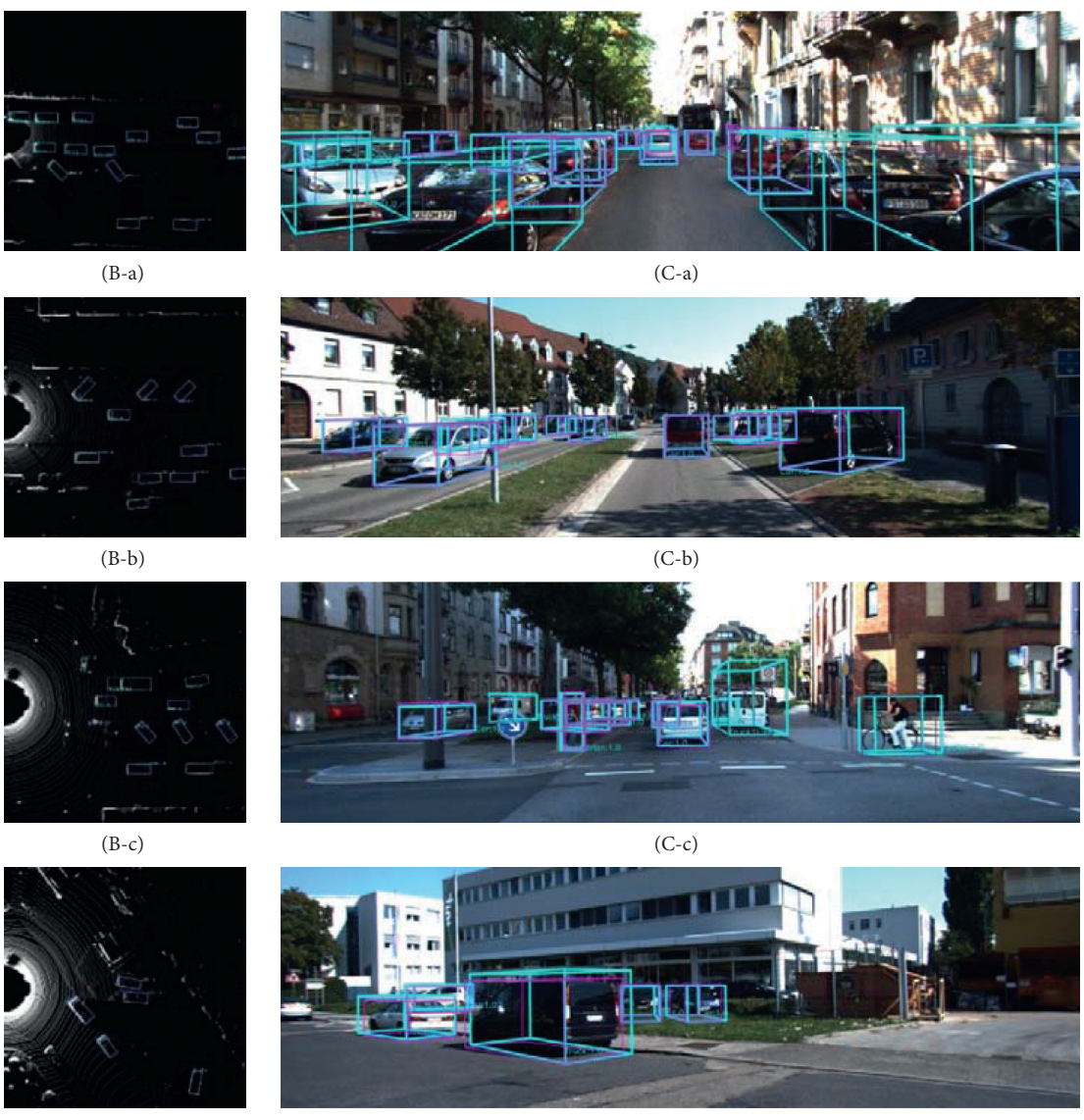
(a) (b) (c)

Figure 8: Three consecutive frames from a turning scene, (a), (b), and (c) are the recognition results for three consecutive frames, with an acquisition interval of 0.5 s.



(a) (b) (c)

Figure 9: Set of consecutive frames from a multicategory scene: (a–c) the recognition results for three consecutive frames, with an acquisition interval of 0.5 s.



(B-a) (C-a)

(B-b) (C-b)

(B-c) (C-c)

(B-d) (C-d)

Figure 10: Recognition visualization of the KITTI dataset. (B, C) The bird's eye view and corresponding camera view separately; (a, b) multiobject scenes; (c) multicategory scene; (d) unordered objects scene.

predicted value is not accurate, it will not deviate wildly. It is the advantage of continuous prediction, which makes the network robust. However, because of an additional prediction dependence in the calculation, the accuracy will worsen if only the prediction results with high quality are compared. Similarly, in the proposed multicategory object recognition network, the bounding box regression is also highly dependent on the category prediction, making the regression not achieve the highest precision, but makes the prediction more robust.

Our approach is not very sensitive to large objects that are very close to the observer. It is the reason that the AP of cars for easy mode was slightly worse than that for the moderate mode. Expanding the receptive field can alleviate such problems, but it will increase the depth of the network. Our experimental results showed that deeper networks increase the inferencing time but have no significant effect on accuracy, which is very different from image recognition tasks.

The indicators in this paper are much higher than those on the KITTI ranking list. The main reason is that the scope of perception selected in this paper is smaller than the standard. Our range covers 41.6 m (70.4 m in standard) in front, 20.8 m (40 m in standard) left and right, and 2 m above and below the LIDAR coordinate system, which has met our application's commands.

4. Conclusions

The main aim of this study was to design a LIDAR-based object recognition method for autonomous vehicle systems. Thus, we proposed a new multifunctional network that operates in real-time with high accuracy and stable performance. As several recognition methods achieve considerable performance differences in different datasets, the Apollo dataset was also adopted besides the KITTI dataset in this study, making the validation results more consistent with actual application scenes. Hence, the proposed recognition method has a high practical value. The key findings of this study are outlined below:

(1) The proposed network realizes the accurate recognition of multiple types of objects in real time.

(2) To tackle the inaccurate category prediction, an auxiliary network was designed to help network auxiliary learning pointwise features. It is not limited to the object's center point category, making the prediction result more robust.

(3) To tackle the inaccurate bounding box prediction, firstly, the validity of the indirect expression of orientation angle by sine and cosine values is verified. Besides, another auxiliary network was designed to help network auxiliary learning box-wise features.

(4) The proposed network delivers a stable and robust object recognition performance in complex environments (multiobject, unordered objects, and multicategory), reflecting its high practical value.

(5) The proposed network's performance is impacted negatively when the LIDAR system is obscured and is not sensitive to large objects that are very close to the observer. Further research is necessary to address this weakness of the network.

In this study, we have considered several possible problems in practical application scenarios. Although our proposed method needs to be further improved, it has demonstrated a very high practical application potential. Based on the phenomenon that most current methods rely heavily on the 3D sparse convolutional operation [35], our research's stable performance showed that artificial bird's eye view features can do the same thing as three-dimensional convolution.

Data Availability

The data used to support the findings of this study are available upon request to the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China, under Grant no. 51775548.

References

[1] W. Liu, J. Sun, W. Li, T. Hu, and P. Wang, "Deep learning on point clouds and its application: a survey," *Sensors*, vol. 19, no. 4188, pp. 1–22, 2019.

[2] M. Li and Y. Chen, "Registration of laser scanning point clouds: a review," *Sensors*, vol. 18, no. 1641, pp. 1–25, 2018.

[3] C. Wang, H. Liao, Y. Wu et al., "CSPNet: a new backbone that can enhance learning capability of CNN," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 390-391, Seattle, WA, USA, June 2020.

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[6] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2018, https://arxiv.org/abs/1804.02767.

[7] A. Bochkovskiy, C. Wang, and H. Liao, "YOLOv4: optimal speed and accuracy of object detection," 2020, https://arxiv.org/abs/2004.10934.

[8] Y. Wang, Y. Sun, Z. Liu et al., "Dynamic graph CNN for learning on point clouds," *ACM Transactions on Graphics*, vol. 38, no. 5, pp. 1–12, 2019.

[9] P. Veličković, A. Casanova, P. Liò et al., "Graph attention networks," in *Proceedings of the International Conference on Learning Representations*, pp. 1–12, Vancouver, BC, Canada, May 2018.

[10] C. Wang, B. Samari, and K. Siddiqi, "Local spectral graph convolution for point set feature learning," in *Proceedings of the European Conference on Computer Vision*, pp. 52–66, Munich, Germany, September 2018.

[11] M. Simonovsky and N. Komodakis, "Dynamic edge-conditioned filters in convolutional neural networks on graphs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3693–3702, Honolulu, HI, USA, July 2017.

[12] L. Yi, H. Su, X. Guo, and L. Guibas, "SyncSpecCNN: synchronized spectral CNN for 3D shape segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6584–6592, Honolulu, HI, USA, July 2017.

[13] K. Roman and V. Lempitsky, "Escape from cells: deep KD-networks for the recognition of 3D point cloud models," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 863–872, Venice, Italy, October 2017.

[14] Y. Li, R. Bu, M. Sun et al., "PointCNN: convolution on x-transformed points," 2018, https://arxiv.org/pdf/1801.07791.pdf.

[15] J. Pamplona, C. Madrigal, and A. Escalera, "PointNet evaluation for on-road object detection using a multi-resolution conditioning," in *Proceedings of the Iberoamerican Congress on Pattern Recognition*, pp. 513–520, Madrid, Spain, November 2018.

[16] C. Qi, H. Su, K. Mo, and L. Guibas, "PointNet: deep learning on point sets for 3D classification and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 77–85, Honolulu, HI, USA, July 2017.

[17] C. Qi, L. Yi, H. Su, and L. Guibas, "PointNet++: deep hierarchical feature learning on point sets in a metric space," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 5100–5109, Long Beach, CA, USA, December 2017.

[18] Z. Wu, S. Song, A. Khosla et al., "3D ShapeNets: a deep representation for volumetric shapes," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1912–1920, Boston, MA, USA, June 2015.

[19] J. Wu, C. Zhang, T. Xue et al., "Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling," 2016, https://arxiv.org/abs/1610.07584.

[20] Y. Zhou and O. Tuzel, "Voxelnet: end-to-end learning for point cloud based 3d object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4490–4499, Salt Lake City, UT, USA, June 2018.

[21] A. Lang, S. Vora, H. Caesar et al., "Pointpillars: fast encoders for object detection from point clouds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12697–12705, Long Beach, CA, USA, June 2019.

[22] C. Sun, W. Zhan, J. She et al., "Object detection from the video taken by drone via convolutional neural networks," *Mathematical Problems in Engineering*, vol. 2020, Article ID 4013647, 10 pages, 2020.

[23] W. Ali, S. Abdelkarim, M. Zidan, M. Zahran, and A. E. Sallab, "YOLO3D: end-to-end real-time 3D oriented object bounding box detection from LiDAR point cloud," *Lecture Notes in Computer Science*, vol. 11131, pp. 716–728, 2019.

[24] P. Cao, H. Chen, Y. Zhang, and G. Wang, "Multi-view frustum pointnet for object detection in autonomous driving," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 3896–3899, Taipei, Taiwan, September 2019.

[25] X. Wang and H. Su, "Completely model-free RL-based consensus of continuous-time multi-agent systems," *Applied Mathematics and Computation*, vol. 382, Article ID 125312, 2020.

[26] X. Wang, H. Su, X. Wang, and G. Chen, "Fully distributed event-triggered semiglobal consensus of multi-agent systems with input saturation," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5055–5064, 2017.

[27] S. Shi, C. Guo, L. Jiang, and Z. Wang, "PV-RCNN: point-voxel feature set abstraction for 3D object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10526–10535, Seattle, WA, USA, June 2020.

[28] K. He, X. Zhang, S. Ren et al., "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference on Learning Representations*, pp. 1–14, San Diego, CA, USA, May 2015.

[30] B. Yang, W. Luo, and R. Urtasun, "PIXOR: real-time 3D object detection from point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7652–7660, Salt Lake City, UT, USA, June 2018.

[31] C. He, H. Zeng, J. Huang et al., "Structure aware single-stage 3D object detection from point cloud," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11873–11882, Seattle, WA, USA, June 2020.

[32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[33] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: the KITTI dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[34] P. Wang, X. Huang, X. Cheng et al., "The ApolloScape open dataset for autonomous driving and its application," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2702–2719, 2019.

[35] Y. Yan, Y. Mao, and B. Li, "Second: sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, pp. 1–17, 2018.