AISChain: Blockchain-Based AIS Data Platform With Dynamic Bloom Filter Tree

Yongshuai Duan, Junqin Huang, Jiale Lei, Linghe Kong, *Senior Member, IEEE*, Yibin Lv, Zhiliang Lin, Guihai Chen, and Muhammad Khurram Khan, *Senior Member, IEEE*

***Abstract*— Since 2002, hundreds of thousands of vessels have equipped the Automatic Identification System (AIS), which continuously broadcasts its identity and location information for vessel collision avoidance. To utilize these scattered AIS data for further analysis, there are multiple AIS data platforms collecting AIS data from vessels around the world through their satellites and land-based stations. Thus, users can obtain AIS data of vessels from these platforms without dedicated devices. However, existing platforms work in silos, and AIS data is distributed across different platforms, resulting in reduced data availability. In addition, AIS is vulnerable to jamming and spoofing attacks, which can undermine the authenticity of AIS data. In this paper, we propose AISChain, a secure and fast blockchain-based AIS data platform. AISChain adopts consortium blockchain, which only permits those authorized parties (i.e., AIS data providers) to participate in the consensus protocol, and is compatible with current commodity AIS hardware. Since the whole system is co-maintained by multiple authorized parties, AISChain can integrate AIS data resources in a secure way. For avoiding repeated recording of AIS data on the chain, we design the Dynamic Bloom Filter Tree (DBFT) to realize efficient duplication detection in the transaction verification phase. We also propose the dual signature scheme to clarify the AIS data ownership. Moreover, we leverage the geographical location-based blockchain sharding approach to further improve the scalability of AISChain. We implement a prototype of AISChain, and conduct extensive experiments to evaluate the performance of AISChain. Evaluation results show that the search time of DBFT is negligible (4.3 ms) with an extreme low error ratio (0.4%). Meanwhile, AISChain can achieve more than 730 tx/s throughput even when nodes scale to 36. To the best of our knowledge, AISChain is the first work to apply the blockchain technology to secure the AIS data platform.**

***Index Terms*— Automatic identification system, blockchain, bloom filter, security.**

Manuscript received 10 December 2021; revised 14 May 2022 and 13 June 2022; accepted 22 June 2022. Date of publication 13 July 2022; date of current version 8 February 2023. This work was supported in part by NSFC under Grant 62141220, Grant 61972253, Grant U1908212, Grant 72061127001, Grant 62172276, and Grant 61972254; in part by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning; and in part by the Open Research Projects of Zhejiang Laboratory under Grant 2022NL0AB01. The Associate Editor for this article was A. K. Bashir. *(Corresponding author: Linghe Kong.)*

Yongshuai Duan, Junqin Huang, Jiale Lei, Linghe Kong, Zhiliang Lin, and Guihai Chen are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: dys1998@sjtu.edu.cn; junqin.huang@sjtu.edu.cn; radiumscripttang@sjtu.edu.cn; linghe.kong@sjtu.edu.cn; linzhiliang@sjtu.edu.cn; chen-gh@sjtu.edu.cn).

Yibin Lv is with China Shipping Industry Company, Shanghai 200135, China (e-mail: lv.yibin@coscoshipping.com).

Muhammad Khurram Khan is with the Center of Excellence in Information Assurance, King Saud University, Riyadh 4545, Saudi Arabia (e-mail: mkhurram@ksu.edu.sa).

Digital Object Identifier 10.1109/TITS.2022.3188691

I. Introduction

THE Automatic Identification System (AIS) is essentially a Very High Frequency (VHF) communication system. An AIS equipment generally includes microprocessor, VHF transceiver, and GPS receiver. Vessels equipped with AIS broadcast their location and identity information continuously and autonomously to surrounding vessels (ranging 20∼40 nautical miles) for avoiding vessel collision. To enhance the safety of vessel navigation, the International Maritime Organization (IMO) [1] demands all registered vessels to install AIS equipment since 2002.

Besides vessel collision avoidance, AIS has been widely used in many critical scenarios, such as aids to navigation (AtoN), search and rescue (SAR) [2]. To make AIS more powerful in these scenarios, it is essential to have a global view of AIS data of different vessels. Thus, there are dozens of AIS data platforms gathering scattered AIS data from various vessels for further resource integration and data analysis, such as MarineC [3] and Sailwx [4], allowing users to obtain historical or real-time AIS data from all over the world without dedicated devices.

However, existing AIS data platforms are facing severe problems. **On the one hand, the unreliability of AIS data sources results in a large amount of invalid or even malicious AIS data collected by AIS data platforms.** In recent years, many studies [5]–[8] have pointed out that AIS has serious security concerns, such as spoofing attack, hijacking attack, and even availability disruption. For example, Balduzz *et al.* [9] spoofed the AIS data platforms by uploading forged AIS data with a software-based AIS transmitter, as shown in Fig. 1. In this way, they also succeeded in lying about a vessel collision accident, and made the collision alert be falsely triggered. From August 2020 to July 2021 alone, researchers discovered more than 100 instances of AIS spoofing attacks that misreported NATO warships' positions. In one case, the destroyer USS Roosevelt was misreported to be sailing four miles inside Russian waters, which is no doubt a territorial violation if it is real [10]. **On the other hand, AIS data platforms are working in silos due to their conflicts of interest, which makes AIS data scattered across different platforms and reduces data availability.** Tu *et al.* [11] analyzed AIS data from all mainstream platforms. They found that the quality of AIS data collected by platforms is not stable. A significant proportion of data is of low integrity or even lost. He *et al.* [12]

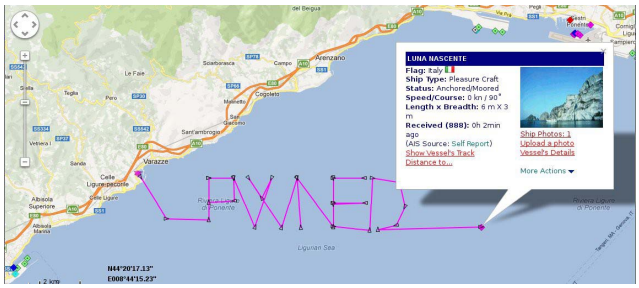

Fig. 1. A fictitious vessel spelling out PWNED in the Mediterranean [9].

designed a visual analytic approach to support evaluation and exploration of AIS data quality, with which they analyzed abnormal data, missing data, and dirty data in their AIS data sets.

There are mainly two research directions for enhancing the availability of AIS data: authenticating data source and improving collected data quality. The most current approach to ensuring the reliability of data sources is to use cryptographic algorithms. For example, Goudossis *et al.* [13] integrated asymmetric and symmetric key cryptography into AIS for identity authentication, in order to resist spoofing attacks. Some commercial AIS products also use symmetric key cryptography [14]. However, the symmetric key between two parties has to be negotiated beforehand, which brings a large communication overhead. More importantly, encrypting AIS data hinders vessels without keys to obtain data, which violates the original design intention of AIS. Another research direction is to improve the quality of collected AIS data. For example, DeAIS [15] is a novel methodological approach for modeling, analyzing, and detecting falsification of AIS data. Besides, Zhao *et al.* [16] introduced a pre-processing algorithm to smooth vessel trajectories based on AIS data. However, these methods are still not a good solution to the root of the above problems. Thus, we come up with an intuitive idea that *if we can integrate AIS data resources from different platforms, we can largely improve data availability and integrity.* But it is hard to push this idea forward because of platforms' conflicts of interest.

We propose AISChain, a secure and fast blockchain-based AIS data platform. To guarantee the authenticity of data source, AISChain adopts consortium blockchain, where only authorized parties can participate, and only qualified AIS data can be recorded in the chain. To get rid of redundant AIS data from multiple data providers, Dynamic Bloom Filter Tree (DBFT) is integrated into AISChain to realize duplication detection. To protect different data providers' interests, we design the dual signature scheme for data source traceability and data ownership confirmation. Moreover, we apply a geographical location-based sharding approach to make AISChain more scalable. We implement a prototype of AISChain and conduct a series of experiments to evaluate the efficiency and robustness of AISChain. Experimental results show that the search time of DBFT is negligible (4.3 ms) with an extremely low error ratio (0.4%). Meanwhile, AISChain can achieve more than 730 tx/s throughputs even when nodes scale to 36. Our main contributions are summarized as follows:

- We propose AISChain, a blockchain-based AIS data platform, with a decentralized consensus strategy while achieving high performance, which is compatible with current commodity AIS hardware. To the best of our knowledge, AISChain is the first work to apply blockchain technology to secure AIS data.
- We design the Dynamic Bloom Filter Tree (DBFT) to achieve efficient duplication detection in the context of massive incoming AIS data. And we utilize the dual signature scheme to trace the AIS data source and confirm the data ownership.
- We implement a prototype of AISChain, and conduct experiments on macro performance (e.g., throughput, latency, robustness) and micro performance (e.g., the efficiency of DBFT) of AISChain. Experimental results demonstrate the efficiency and security of AISChain.

II. Related Work

There is a lot of work on improving security and availability of AIS data. We will present related engineering and research in this area from three perspectives: platform, security, and availability.

A. AIS Data Platform

There have been many AIS data platforms. ExactEarth [17] and Orbcomm [18] mainly provide AIS data collected through low earth-orbit satellites. Some platforms collect AIS data from terrestrial-based AIS stations, such as Marine Traffic [19], Vessel Tracker [20], MarineC [3], and Sailwx [4]. There are some other commercial AIS data platforms, like VT Explorer [21], FleetMon [22], and HiFleet [23]. China Maritime Safety Administration (MSA) also provides an AIS information service platform [24]. Although there are so many AIS data platforms, none of them can provide complete and accurate real-time AIS data sets all over the world.

B. AIS Data Security

Many researchers try to tackle the problem of unreliability of AIS data source, which could mask or favor illegal actions, thereby leading to disturbance of monitoring systems and maritime risks. For example, Hall *et al.* [5] introduce a competent central authority to authenticate the identity of AIS data provider by issuing certificates. Asymmetric key cryptography is a popular way to realize the authentication and verification of AIS data [13], [25], [26]. Aziz *et al.* [27] propose SecureAIS, a key agreement scheme that allows any pair of vessels to agree on a shared session key, while this scheme costs a large communication overhead. They all need one central authority to issue certificates. Some other studies are aimed at ensuring the safety of AIS. For example, Bonetto and Pilosu *et al.* [28], [29] exploit white space communication to increase AIS security. Auth-AIS [30] leverages TESLA and Bloom Filter techniques to build a backward-compatible authentication framework to secure AIS broadcast messages.

C. AIS Data Availability

Since the availability of collected AIS data in current platforms is poor [11], some studies try to improve the availability of AIS data. For example, Zhao *et al.* [16] propose a error pre-processing algorithm of AIS tracks including physical integrity, spatial logical integrity and time accuracy. Ray *et al.* [15] propose DeAIS to model, analyze and detect the falsification of AIS data in real-time. There are also some methods combining AIS with other equipments or techniques to validate AIS data, such as space-based AIS signals [31], radar information [32]. Mao *et al.* [33] construct a standard AIS database for maritime trajectory prediction and data mining for research area. However, these methods still cannot satisfy the requirement of data availability due to the low integrity and high missing rate of collected AIS data.

AIS data provided by current siloed AIS platforms are not sufficient for scientific research and precision applications. The improvement of AIS from local perspective cannot fundamentally guarantee the overall security and availability of AIS data in generation, collection, and sharing. Therefore, we need a new methodology to address these issues.

III. Preliminary

We will introduce some basic knowledge in advance, which includes Automatic Identification System (AIS), consortium blockchain, and Bloom filter.

A. Automatic Identification System

Since 2002, Automatic Identification System (AIS) is demanded to install in class A vessels, which include international voyaging vessels with a gross tonnage of over 300, and all passenger vessels [1]. Since 2008, AIS is also set up on satellites to complement land-based AIS stations [34]. Though AIS is originally developed for collision avoidance, it is also a critical technology used for maritime assistant. For example, the aids-to-navigation standard is used for non-vessel information broadcast, such as buoys, oil platforms, aircraft, or autonomous vehicles [35].

The coverage of AIS signals is limited. AIS signal is transmitted in the air at 161.975MHz and 162.025MHz radio frequency (VHF). As shown in Fig. 2, AIS signal can spread 40 nautical miles near shore, and only 20 nautical miles in the ocean [11]. Thus, AIS data platforms usually collect data through land-based stations and vessel traffic services (VTS) operated by port authorities. Individuals, such as vessel's captain, can upload AIS data by mobile applications or e-mail as well. For vessels more than 40 nautical miles from the coast, only low-Earth orbit satellites can be used to obtain real-time AIS data.

A piece of AIS data contains three types of information: static information, dynamic information, and voyage-related information [1]. Static and voyage-related information is transmitted every 2∼12 seconds and dynamic information is transmitted every 6 minutes. Static information includes Maritime Mobile Service Identity (MMSI), a unique 9-digit number assigned to an AIS equipment. In dynamic information, the vessel's real-time GPS position and its corresponding timestamp in Universal Time Coordinated (UTC) are both recorded.

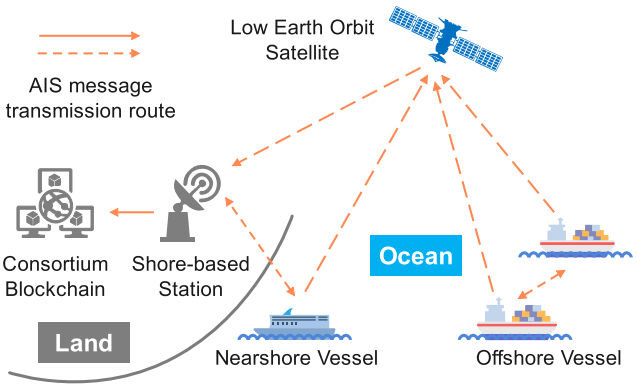

Fig. 2. Generation, transmission and collection of AIS data in AISChain.

B. Consortium Blockchain

Consortium blockchain is a permissioned blockchain, where only authorized users can participate. Consortium blockchain usually adopts Byzantine consensus protocols, such as PBFT, and HotStuff, which can achieve much higher throughput and lower power consumption than lottery-like consensus protocols adopted in the public blockchain. Therefore, consortium blockchain is more suitable as the underlying infrastructure of the AIS data platform. There have been some popular open-source consortium blockchain frameworks, such as Hyperledger Fabric, FISCO BCOS, and Corda. Besides, consortium blockchain has the same advantages as the public blockchain, which are also inherited by AISChain:

- **Immutable.** Since the blocks in the chain are connected first and last by hash pointers, once one block is tampered with, the hash pointer will be broken. Users can easily verify the correctness of a block by checking the corresponding hash pointer. Thus, we consider data recorded in the blockchain is immutable.
- **Traceable.** Users record data in the blockchain by sending transactions, which are signed by users' private keys. Combining the immutable property of the blockchain, each piece of data recorded in the blockchain can be traced. So, we say the data stored in the blockchain is traceable.
- **Decentralized.** consortium blockchain is jointly maintained by multiple authorized nodes, instead of central authorities. All results that are recorded on the blockchain are multi-party consensus, so the whole system is organized in a decentralized manner.

C. Bloom Filter

Bloom filter is a technique to determine whether there is a certain element in a data set efficiently. It takes up very little memory space and is capable of performing efficient inserts and queries at the same time. The essence of Bloom filter is an ingenious probabilistic data structure, which consists of an n-bit vector V and a set of hash functions $H = \{Hash_i \mid i \in \{1, 2, \ldots, k\}\}$. We assume that all hash functions in H can unbiasedly map one element to an integer from 0 to $n-1$.

All bits in V are set to 0 initially. When an element m is added, it will be separately hashed with k hash functions in H, and we can get k integers falling in the range of $[0, n-1]$. These k integers denote the chosen indexes in V, and then we set the bits of chosen indexes to 1. If we want to check whether one certain element m is in a data set, we can hash m with hash functions in H to get k integers. If there exists one bit of the chosen index is 0, it indicates that m is not in the data set. Otherwise, m is considered to exist in the data set with a very high probability.

It is noted that the judgment result of Bloom filter is probabilistic, the error probability ϵ is influenced by the bit vector's length n, the number of hash functions k, and the size of the data set. We will further analyze the error probability ϵ in section VII.

IV. Problem Statement

Existing AIS data platforms suffer from unreliable data sources, poor data availability, and low data quality [11]. To solve these problems, we aim to unite different AIS data providers to govern a blockchain-based AIS data platform together in a decentralized consensus way. By leveraging the authentication mechanism of consortium blockchain, we can guarantee the reliability of AIS data sources without central authorities. Besides, integrating AIS data resources together can largely increase the sampling resolution and data coverage, thereby improving the data quality. However, there are three main challenges that need to be addressed when introducing blockchain into the AIS data platform.

Challenge 1: How to deal with a large amount of incoming AIS data in real-time? Depending on the vessel type and sailing status, vessels are supposed to broadcast their AIS data from every few seconds to every few minutes. According to MarineTraffic [36], over 550,000 vessels install AIS equipment, and more than 800 million pieces of AIS data are recorded monthly now, which means on average, 20,000 pieces of AIS data are recorded every minute. When integrating AIS data from various providers, the total amount of data will be greater. Besides, the generation and acquisition rate of AIS data is not stable, so the volume of incoming AIS data may increase sharply in a certain region or at a certain time.

Challenge 2: How to ensure their AIS data ownership when various providers are integrated? AIS data is generated by hundreds of thousands of AIS equipment installed in vessels, collected by land-based and space-based AIS stations, and stored by providers from all kinds of organizations. For one thing, AIS has no authentication methods, so everyone can generate arbitrary AIS data and send it to data platforms [9], who are not able to verify the authenticity of the collected AIS data. For another, AIS data providers are not willing to share AIS data with other providers, because they do not trust each other and can not clearly claim the data ownership, which could cause conflicts of interest.

Challenge 3: How to detect duplication of AIS data efficiently on a huge data set? AIS data duplication has already been a problem for existing AIS data platforms [11]. There is considerable overlap in AIS data held by providers, once AIS data from various providers is integrated, data duplication is going to be a common case. Besides, AIS data is collected by land-based stations, satellites, or vessels themselves, and the data uploading delay of different routes may vary from seconds to hours or even days. Therefore, avoiding duplication is not trivial. Each piece of AIS data has to be detected for duplication before being recorded, and each detection has to traverse all existing AIS data. Considering the huge scale of AIS data, duplication detection will inevitably bring great computing overhead, and severely hinders the integration of AIS data and the union of AIS data providers.

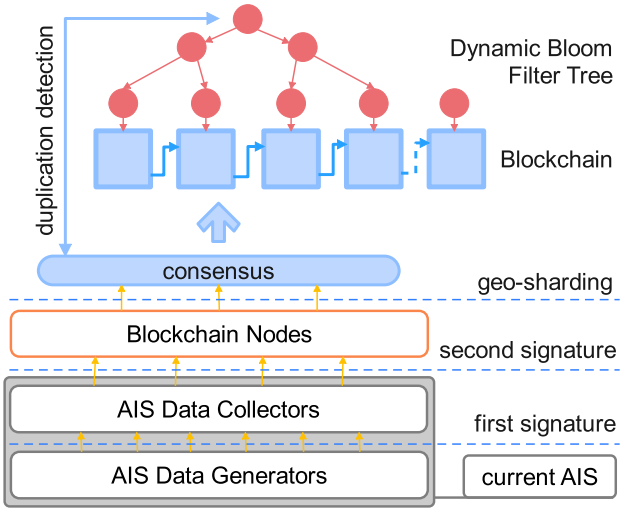

Fig. 3. System architecture and workflow of AISChain.

V. AISChain: A Secure and Fast Blockchain-Based AIS Data Platform

In this section, we firstly introduce the architecture and workflow of AISChain, and then describe three key designs of AISChain in detail, i.e., geo-sharding, dual signature, and Dynamic Bloom Filter Tree (DBFT).

A. Overview

AISChain is a secure and fast AIS data platform, which is built on consortium blockchain. Following the designed blockchain protocol, AIS data will be generated, signed, collected, merged and stored in a decentralized manner.

Fig. 3 shows the system architecture and workflow of AISChain. There are three types of roles in the AIS data platform: AIS data generator, AIS data collector and blockchain node. Among them, AIS data generator and AIS data collector are inherent roles of current AIS, which provide hardware foundation for AISChain. AIS data generators mainly refer to vessels, and also include offshore exploration drilling rigs, helicopters, lighthouses and so on [35]. AIS data collectors are mainly composed of satellites and land-based stations. It is worth noting that some AIS data may not be collected by either satellites or land-based stations. This part of AIS data may be uploaded to the AIS data platform by vessels themselves after docking. Therefore, AIS data generators can also serve as AIS data collectors under certain circumstances. The blockchain nodes are responsible for recording the AIS data

transmitted from AIS data collectors, and jointly maintaining a decentralized AIS database, i.e., a blockchain.

Here we take a piece of AIS data generated by a cargo vessel in voyage as an example to illustrate the workflow of AISChain. As mentioned above, the cargo vessel acts as the AIS data generator. It will automatically register an authenticated key pair in consortium blockchain before leaving the factory. When broadcasting the AIS data, the cargo vessel signs the first signature on the AIS data with its private key, and appends the first signature along with its public key to the end of the AIS data. Subsequently, we assume that the signed AIS data is collected by a satellite, which acts as the AIS data collector. The satellite also has an authenticated key pair registered in the blockchain. It signs the second signature on the received AIS data, and appends the second signature along with its public key to the end of the message. After that, the satellite broadcasts this dual signed AIS data to blockchain nodes as soon as possible. And then, blockchain nodes receive a copy of the dual signed AIS data, and perform the following four operations to check the legality of the received AIS data:

Firstly, blockchain nodes check the geographical location information in the AIS data, and determine whether this AIS data should be handled by the current blockchain shard. If not, blockchain nodes will discard this AIS data directly. In section V-C, more details about geographical location-based sharding will be introduced.

Secondly, blockchain nodes verify the authenticity of the AIS data, through checking if the dual signatures in the AIS data are signed by a registered AIS data generator and a registered collector, separately. AIS data signed by unauthorized AIS data generators and collectors will be discarded.

Thirdly, blockchain nodes verify the integrity of the AIS data by checking the correctness of dual signatures, which are further introduced in section V-B.

Lastly, blockchain nodes perform duplication detection through Dynamic Bloom Filter Tree (DBFT). Each node maintains a Bloom filter tree built on its local blockchain data. The detailed design of DBFT is introduced in section V-D. If no duplication is detected, the node adds the AIS data to the pending block as a transaction, and updates the block's Merkle Tree and the Bloom filter's bitmap. Otherwise, the AIS data will be discarded.

Inside the pending block, the pointers to AIS data are sorted by timestamp, and the sorted pointers are kept as a pointer list in the block header, as shown in Fig. 4. Once the pending block is ready, the leader node will propose it as a new block. All nodes run the Byzantine protocol [37] to reach consensus on the proposed new block, and append to the blockchain. After that, DBFT will also be updated for future duplication detection. To this end, blockchain nodes have successfully recorded the AIS data from the cargo vessel.

B. Dual Signature Scheme

Different from existing AIS platforms, AIS data collectors (e.g., satellites, land-based stations) in AISChain may belong to different organizations. In order to guarantee interests of all participants and organizations, the dual signature scheme is designed to confirm the data ownership of every piece of AIS data.

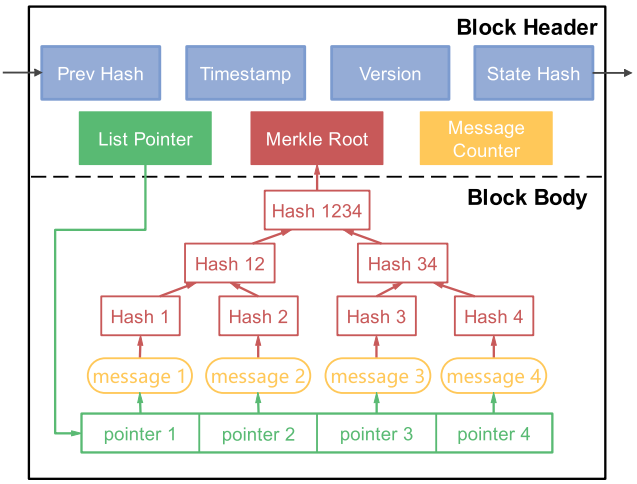

Fig. 4. Block structure of AISChain.

Both AIS data generators and collectors own key pairs registered in the blockchain. They sign the AIS data with their private keys in the generation and collection stage, respectively. The signatures signed by the collectors and generators are combined as the dual signatures of AIS data. With the dual signatures, AISChain can not only recognize the contribution of the data collectors, but also trace the transmission path of each piece of AIS data.

The AIS data generator has a authenticated key pair $(k_g^{public}, k_g^{private})$, which is registered in consortium blockchain beforehand. To sign the AIS data m_g, the AIS data generator firstly hashes m_g to generate a message digest. Then the generator uses its own private key $k_{private}^{g}$ to sign the message digest, and gets a signature s_g for message m_g:

$$s_g = Sign(k_g^{private}, Hash(m_g)). \tag{1}$$

The AIS data generator appends the digital signature s_g and its public key k_g^{public} to the AIS data m_g to obtain a signed message m_c, and then broadcasts $m_c = \{m_g, s_g, k_g^{public}\}$. After the signed AIS data m_c is collected by the AIS data collector, who also has a certified key pair $(k_c^{public}, k_c^{private})$, it also signs the AIS data m_c and gets a digital signature s_c:

$$s_c = Sign(k_c^{private}, Hash(m_c)). \tag{2}$$

The collector also appends the digital signature s_c and its public key k_c^{public} to the AIS data m_c to obtain a dual signed message $m_d = \{m_g, s_g, k_g^{public}, s_c, k_c^{public}\}$. With (1) (2), the AIS data is dual signed.

To verify the integrity of received AIS data, we only need to verify the signatures s_g, s_c as follows:

$$Hash(m_g) = Verify(k_g^{public}, s_g), \tag{3}$$

$$Hash(m_c) = Verify(k_c^{public}, s_c). \tag{4}$$

If (3) and (4) hold, the digital signatures s_g, s_c are valid and the dual signed message m_d is unmodified. And the authenticity of received AIS data holds only if the public keys k_g^{public}, k_c^{public} are registered in consortium blockchain.

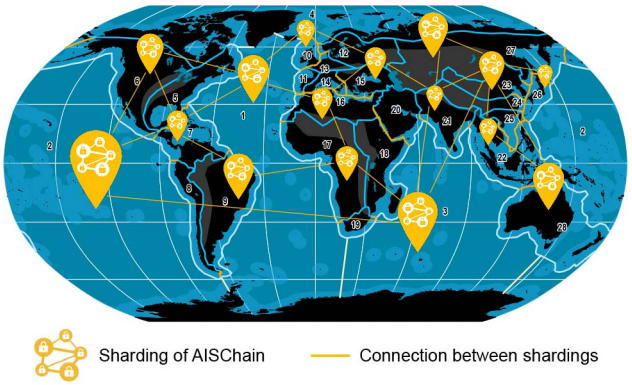

Fig. 5. Blockchain sharding based on geography of maritime ranges.

However, there may be multiple collectors claiming ownership of the same piece of AIS data. Consider a situation that when two satellites both capture the same piece of AIS data, how to determine the ownership of data? To avoid such situation, we stipulate that for a certain piece of AIS data, only the collector who broadcasts it first obtain the ownership. Once the blockchain node receives a piece of new AIS data, it will discard duplicated AIS data received afterward. Since AISChain is decentralized, different nodes may have different views of received data. Thus, we leverage the decentralized consensus to achieve the final consistency of views. Such mechanism, on the one hand, encourages collectors to collect and transmit AIS data as soon as possible for competing the ownership. On the other hand, it encourages data collectors to collect more extensive data to avoid duplication.

C. Geo-Sharding Approach

In this section, we propose the geographical location-based sharding approach, short for geo-sharding, which divides the whole blockchain into multiple shards according to geographical location, to alleviate the pressure of massive AIS data bringing to AISChain.

Although many blockchain-based data sharing systems [38] have been proposed, AISChain is quite different from them in specific designs. The most fundamental difference between them is that the content stored in blockchain of the AIS data platform is pure AIS data, while the content stored in traditional blockchain is the transactions between users. Transactions may occur between different shards, which makes sharding of traditional blockchain has a high risk of inconsistency. On the contrary, AIS data is composed of AIS messages from independent vessels, which needs no interaction cross-sharding. Therefore, the geo-sharding approach does not affect the functionality of AISChain. Instead, it reduces the amount of data that needs to be processed by each shard, so that AIS data can be processed in different shards in parallel.

Another feature of AISChain is that AIS data is closely related to region, both in geographical and political sense. For example, United States Coast Guard is using AIS data for maritime domain awareness [39]. In the context of consortium blockchain, blockchain nodes manage the AIS data in their own region, which is consistent with geographical characteristics (i.e., GPS information) of the AIS data. As a consequence, the area division of geo-sharding requires comprehensive consideration of various geographic factors.

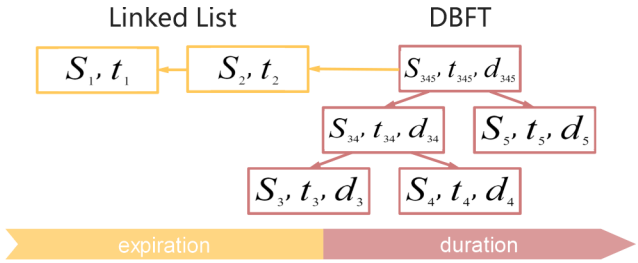

Fig. 6. Structure of DBFT and its dynamic adjustment mechanism.

AISChain sharding scheme refers to Jean-Paul Rodrigue's methodology [40], where global maritime regions are defined as 28 maritime ranges representing functional commercial entities. Maritime ranges are shown in Fig. 5. We divide the blockchain shards according to maritime ranges, and each maritime range represents a shard. AIS data is classified into corresponding shards according to vessel's position marked in AIS data itself. Each blockchain shard only handles those AIS data collected within its maritime ranges.

D. Dynamic Bloom Filter Tree (DBFT)

Query operation of is a key function to realize for blockchain. To accelerate blockchain query process, Wang *et al.* proposed vChain [41], which is a novel verifiable query processing framework and ensures query integrity. And Zhang *et al.* [42] take the first step towards studying authenticated range queries in the hybrid-storage blockchain. But these methods cannot achieve real-time and large number of query operations. In order to realize efficient query operation for enormous AIS data, we propose Dynamic Bloom Filter Tree (DBFT) for duplication detection. DBFT is a data structure which is closely integrated with blockchain. It can be dynamically updated along with the block generation. And DBFT can realizes real-time duplication detection with a small computing overhead.

From the right part of Fig. 6, we can observe that DBFT is a binary tree. Each node of the tree contains a bit vector V, timestamp t, and node depth d. Expired nodes are removed and stored in the linked list, as shown in the left part of Fig. 6.

If node p is a leaf node, and it corresponds to block B_p of blockchain, the value of bit vector V_p, timestamp t_p, and depth d_p can be calculated as follows:

$$V_p[\text{index}] = \begin{cases} 1, & \text{if } \exists Hash_i(id_j \,||\, t_j) = \text{index}, \\ & i \in \{1, 2, \ldots, k\}, \ \ j \in \{1, 2, \ldots, l\} \\ 0, & \text{otherwise}, \end{cases} \quad (5)$$

$$t_p = \max\{t_j \mid j \in \{1, 2, \ldots, l\}\}, \quad (6)$$

$$d_p = 1. \quad (7)$$

Here we assume that block B_p has l pieces of AIS data. For AIS data $m_j (j \in \{1, 2, \ldots, l\})$ in Block B_p, id_j represents MMSI, which is a nine-digit decimal number, and t_j represents UTC timestamp, which is a 32-bit binary number. A piece of AIS data can be uniquely determined by MMSI and timestamp. Therefore, the concatenation of MMSI and

timestamp is set as the primary key of a piece of AIS data. H is a set of unbiased hash functions. V_p in (5) is calculated with the method of Bloom filter, as described in section III-C. The primary key of AIS data is taken as the input element of Bloom filter. V_p is the bit vector of the block B_p, every piece of AIS data in B_p is added to the filter. In (6), the algorithm uses the latest timestamp in block B_p as the timestamp t_p of the node p. In (7), we set the depth of a leaf node as 1.

For a non-leaf node q, the value of bit vector V_q, timestamp t_q, and depth d_q are merged from its two child nodes, i.e., V_{2q} and V_{2q+1}.

$$V_q = V_{2q} \mid V_{2q+1}, \tag{8}$$

$$t_q = \max\{t_{2q},\ t_{2q+1}\}, \tag{9}$$

$$d_q = \min\{d_{2q},\ d_{2q+1}\} + 1. \tag{10}$$

The bit vector V_q is calculated from V_{2q} and V_{2q+1} of its two child nodes with the bitwise OR operation. The timestamp t_q is determined by the larger one between t_{2q} and t_{2q+1}.

DBFT provides three operations to meet the need of AIS-Chain: `INSERT`, `SEARCH`, `DELETE`. The minimum operating unit is a block. When a new block is generated, a corresponding leaf node is inserted to DBFT. Before AIS data is added to a block, they go through the DFBT duplication detection, which needs search operation. Actually, deletion is not common for Bloom Filter, but it is a key operation for DBFT to ensure that DBFT does not degenerate during the continuous insertion. For AISChain with increasing blocks, DBFT is bound to grow larger. However, the length of bit vectors in DBFT is constant, which causes the number of 1 in bit vector of nodes to increase monotonically. In this case, the search operation will have to traverse more and more nodes along with the increasing number of blocks.

Algorithm 1 shows the logic of three operations in DBFT. The algorithm is implemented in a recursive way.

- `INSERT`. We always insert a new leaf node to the shallowest node of the DBFT. In the steady state, the frequency of delete operation is the same as insert operation. Therefore, the size of DBFT remains roughly the same, and DBFT does not need extra operations to keep itself strictly balance.
- `SEARCH`. The algorithm search the primary key from top to down, until it is found in a leaf node. There is a probability of false positive, which can be eliminated by further searching inside the block. AIS data in block is sorted by timestamp. Therefore, when a leaf node of DBFT returns a positive result, we can further search inside the corresponding block is a binary search way.
- `DELETE`. Duration T of DBFT is predefined, which determines how long a node stays in DBFT. Nodes only containing expired AIS data are deleted from DBFT. The delete operation in Algorithm 1 traverses the entire DBFT.

When querying an element with Bloom filter, if an element exists, Bloom filter always gives the right result of 'true'. If the element not exists, although Bloom filter has a high probability of giving the right result of 'false', it still has a small probability of giving the wrong result of 'true', which is called False Positive (FP). Since Bloom filter is the basic

Algorithm 1 `INSERT`, `SEARCH`, and `DELETE` of DBFT

```
1: DBFTNode root                          ▷ root of DBFT
2: string pk          ▷ primary key of a piece of AIS data
3: size_t t                               ▷ current UTC time
4: size_t T                               ▷ duration of DBFT
5:
6: Operation INSERT (root, newNode)
7:     if root is leaf node:
8:         root.left = root
9:         root.right = newNode
10:        root = merge (root.left, root.right)
11:        return
12:    if root.left.depth < root.right.depth:
13:        INSERT (root.left, newNode)
14:    else:
15:        INSERT (root.right, newNode)
16:    return
17:
18: Operation SEARCH (root, pk)
19:    if pk not satisfy root.bitArray:
20:        return false
21:    if root is leaf node:
22:        return true
23:    return SEARCH (root.left, pk)
24:           or SEARCH (root.right, pk)
25:
26: Operation DELETE (root, t)
27:    if root.timestamp < t − T:
28:        delete root
29:        return
30:    if root is not leaf node:
31:        DELETE (root.left, t)
32:        DELETE (root.right, t)
33:    return
```

data structure of DBFT, when recursive to leaf node of DBFT during duplication detection, there is also a possibility of FP as well.

Actually, FP ratio can be calculated by theoretical method. For a node in DBFT, if n is the number of bits in bit vector, k is the number of independent hash functions, and x elements have been inserted to the node, it is easy to infer that the probability of a false positive is:

$$\varepsilon_0 = (1 - (1 - \frac{1}{n})^{kx})^k \approx (1 - e^{-\frac{kx}{n}})^k. \tag{11}$$

Since AIS data in the leaf nodes of DBFT is randomly distributed, when DBFT with y leaf nodes searches a message, the occurrence of false positives for j leaf node is independent and equally distributed. Based on this principle, we can infer that the probability of a false positive for the entire DBFT is:

$$\varepsilon = 1 - (1 - \varepsilon_0)^y. \tag{12}$$

We analyze the complexity of above three operations to prove their efficiency. The number of blocks generated in duration T is stable. Then, the number N of DBFT's leaf nodes, which equals to the number of blocks, is stable as

well. Therefore, time complexity of INSERT is $O(\log N)$, time complexity of SEARCH and DELETE is $O(N)$. With further analysis, the probability of false positive in a single DBFT node is rather low, and the search operation will hardly go down to both child nodes at the same time in the recursive process, which makes the time complexity of SEARCH close to $O(\log N)$. In addition, because DELETE can be integrated into SEARCH and INSERT, which delete node that are expired as well as both of its child nodes when encountering one. In this case, DBFT only need to update parent nodes of the deleted one, then time complexity of delete is also $O(\log N)$.

VI. Analysis

For a multi-party organizing AIS data platform, efficient data collection methods and reliable cooperation mechanisms are necessary. On this basis, to make AISChain work better, many other aspects need to be considered. In this section, we analyze how AISChain responds to various challenges in three aspects, i.e., AIS data security, data ownership confirmation, and scalability of AISChain.

A. AIS Data Security

Compared with the existing AIS data platforms, AISChain has a significant improvement in security. Most attacking methods mentioned in [9] cannot threaten AISChain.

There are mainly six AIS spoofing scenarios: ship, aids-to-navigation (AtoN), search and rescue (SAR), collision, distress beacons, and weather forecasting. All of them launch attacks through crafting and broadcasting non-existent AIS data. Besides, AIS hijacking attacks alter any information in existing AIS data, which is also called man-in-the-middle attack. The enforced signature of AISChain can avoid all the above attack methods. Attackers without authorized certificates cannot forge a signature admitted by others. Vessels with authorized certificates have no motivation to forge or revise their own data. If they did, they will be identified and accused through their signatures on AIS data.

However, AISChain cannot prevent availability disruption attacks [9] from the physical layer, such as slot starvation, frequency hopping, timing attack, and GPS jamming, which are difficult to handle these attacks from the software layer.

B. Data Ownership Confirmation

The Dual signature scheme of AISChain gives credit to AIS data collectors. The second signature confirms the data ownership of each piece of AIS data and is stored in AISChain permanently. Based on data ownership confirmation, collectors can obtain benefits they deserve through data trading or other methods. AISChain also motivates participants by introducing competition. Since AIS data cannot be kept repeatedly in AISChain, collectors are forced to collect AIS data with shorter delay and broadcast them to AISChain nodes as soon as possible, so as to claim data right. At the same time, in order to claim more AIS data, collectors will dedicate to dig extra AIS data not yet been collected by others.

However, preemptive broadcast attack aims to claim others' AIS data. A malicious collector monitors and intercepts AIS data broadcast by other nodes, signs its second signature with its own private key, and broadcasts the modified AIS data. In fact, there is a way to circumvent this attack. Collector broadcast AIS data in three steps. Firstly, AIS broadcasts a claim message with only second signature of the AIS data to AISChain nodes. Secondly, blockchain nodes that receive claim message reply an acknowledgement message to collector. Thirdly, if the collector receives acknowledgement messages from more than half of the AISChain nodes, it broadcasts the original AIS data. Since malicious collector cannot get access to the plain text of original AIS data before it is claimed, preemptive broadcast attack is not possible to perform.

C. Scalability of AISChain

The maritime industry is in the ascendant. In recent years, many countries have enacted legislation to enforce vessels to install AIS equipment. It is foreseeable that as the number of vessels equipped with AIS equipment increases, the load of AISChain will also show an upward trend. Against this backdrop, AISChain reserves space for scalability. In AISChain, consensus speed among AISChain nodes and speed of duplication detection for AIS data are two performance bottlenecks.

The first bottleneck can be solved by the geo-sharding approach. At present, AISChain has 28 shards according to the division of maritime ranges [40]. If the load of a shard exceeds its capacity, the shard can be further divided into two small shards. At the same time, the original blockchain corresponding to this shard is forked to two new chains, corresponding to two new blocks. In this way, the load borne by each new shard returns to tolerable range.

For the second bottleneck, paralleled duplication detection can improve the efficiency significantly. The main idea of paralleled duplication detection is to maintain a query pool. if every piece of AIS data in query pool is unique, then they can be detected for duplication at the same time.

VII. Experiment

We built a prototype of AISChain and tested its throughput and latency in this section. In addition, we also tested and compared the efficiency of DBFT in different scenarios.

A. Prototype Implementation

To simulate the scenario of multiple members participating in the blockchain, we write a prototype of AISChain and open source it on Github.[1] In the experiment, the AISChain prototype is composed of 36 hosts in a local area network. Among them, 8 hosts equip Intel Xeon(R) CPU E5-1620 v3 @3.5GHz, and other 28 hosts equip Intel Xeon(R) CPU W-2123 @3.6GHz. All hosts run on Ubuntu 18.04 LTS.

The AISChain prototype is established with FISCO BCOS.[2] We revise its block structure to introduce DBFT pointers to AIS data in blocks. We use Elliptic Curve Digital Signature

[1] https://github.com/1570005763/AISChain

[2] https://github.com/FISCO-BCOS/FISCO-BCOS

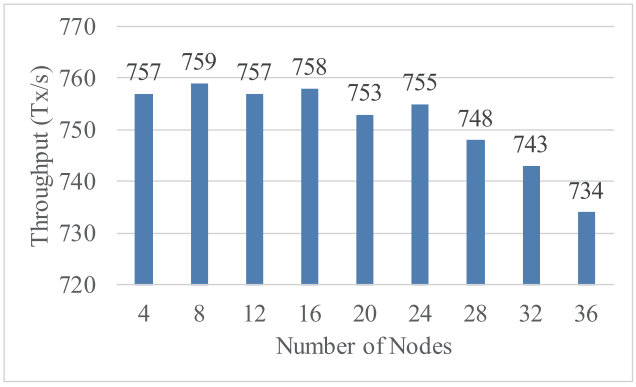

Fig. 7. Throughput of AISChain with the scale of 4∼36 nodes.

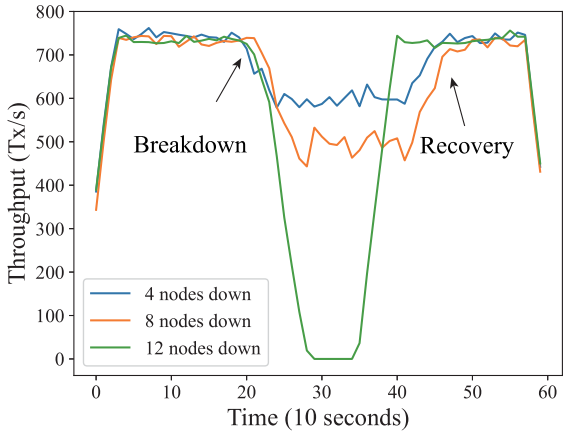

Fig. 8. Throughput of 36-node AISChain with 4/8/12 nodes breakdown.

Algorithm (ECDSA) as our algorithm for dual signature. The length of private key and public key are 256 bits and 512 bits. The size of the hash function set H is 3. And the size of bit vector V is 8 KB. In AISChain, every block contains 1k pieces of AIS data. Besides, we use a public AIS data set [43] as our AIS data source.

B. Throughput and Latency of AISChain

We evaluate the extreme performance of AISChain with the scale of $4, 8, \ldots, 36$ nodes by letting a node multicasts a sufficient amount of AIS data to AISChain. As shown in Fig. 7, the maximum throughput of AISChain is maintained above 700 tx/s. When the scale is less than 24 nodes, the maximum throughput of AISChain remains above 750 tx/s. When the scale exceeds 24 nodes, the overall throughput of AISChain begins to show a downward trend. In fact, as the number of nodes in consortium blockchain increases, the complexity of communication between nodes increases, and the increase in communication overhead reduces the performance of consortium blockchain. Our experimental environment is inside a Local Area Network (LAN), where communication delay between nodes is relatively low. Therefore, when the number of nodes is less than 24, the influence of communication on the throughput of AISCain is not obvious. However, when the number of nodes exceeds 24, the overhead caused by the increase in communication complexity will gradually reduce the maximum throughput of AISChain.

We also evaluate the performance of AISChain when nodes break down (or perform malicious behaviors). At 200 seconds, 4/8/12 nodes are shut down. Then at 400 seconds,

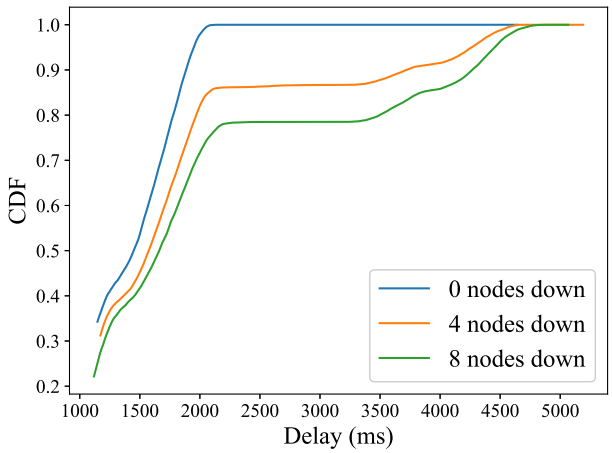

Fig. 9. Latency of 36-node AISChain with 4/8/12 nodes breakdown.

the shut-down node restarted and resumed normal operation. Fig. 8 shows throughput curves of 36-node AISChain with 4/8/12 nodes break down. Throughput is around 600 tx/s when 4 nodes break down and around 500 tx/s when 8 nodes break down. If 12 nodes break down, AISChain's throughput quickly dropped to 0. The reason for this situation is that Byzantine Fault Tolerance requires more than two-thirds of the nodes to work properly. When the number of breakdown nodes reaches one-third (12 in 36), the entire system will go down. What's more, Fig. 9 shows cumulative distribution function of latency. Latency is the time between acceptance and confirmation of a piece of AIS data. In 36-node AISChain, the latency of 99% AIS data is below 2000 ms. When 4 or 8 nodes are shut down, only 75% to 85% AIS data's latency is below 2000 ms, and latency of 99% AIS data is below 4600 ms. In AISChain, every node takes turns serving as the master node. If a node does not respond in its turn for being the master node, other nodes will turn to the next node. Such operations will be performed once in each round. Therefore, when some nodes are down, the overall throughput of AISChain will slightly reduce and remain stable, the throughput of AISChain will quickly recover when the breakdown node is restored as well. Besides, part of AIS data will have a higher latency due to the non-responding of breakdown master nodes.

After the above simulation experiments, we evaluated the maximum throughput and latency of AISChain in the deployed state, as well as the robustness under abnormal conditions. As mentioned in section IV, 20,000 pieces of AIS data are recorded every minute, which means less than 400 pieces of AIS data are recorded every second, which is far less than AISChain's maximum through, even without considering geo-sharding. In conclusion, AISChain can fully meet the current requirements for AIS data recording, and can also ensure the stable recording of AIS data in the future under the background of the rapid growth of AIS data volume.

C. Efficiency of Dynamic Bloom Filter Tree (DBFT)

We set two methods for comparison in experiments about DBFT. One of them is Bloom Filter Tree (BFT), which is DBFT without delete operation. The other is Bloom Filter List (BFL), in which Bloom filters are not constructed to a tree, but a list.

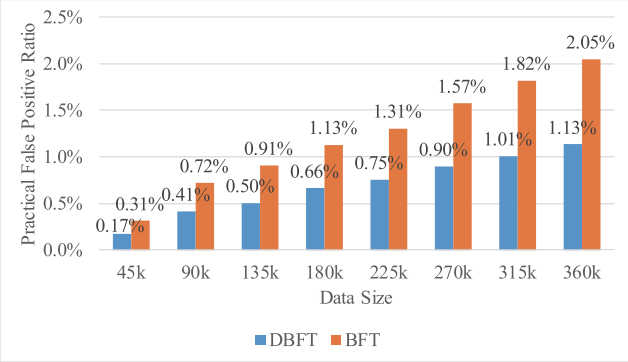

Fig. 10. Practical false positive (FP) ratio under 45k∼360k AIS data size.

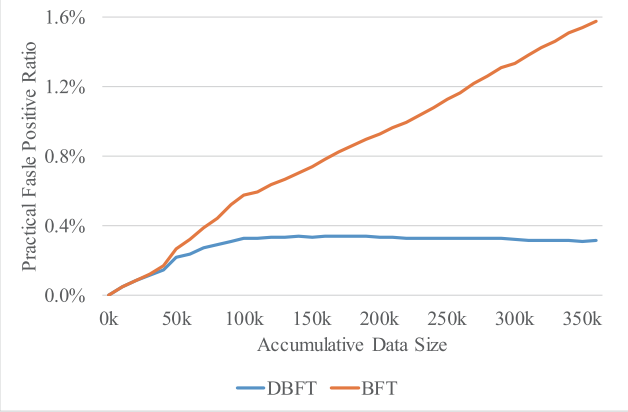

Fig. 11. Practical false positive (FP) ratio under accumulative AIS data size.

We evaluated the FP ratio of DBFT from two perspectives. On the one hand, we compare the FP ratio of DBFT and BFT under different data volumes. As shown in Fig. 10, a total of eight sets of data of different sizes were compared. The volume of AIS data in them started from 45k, with a tolerance of 45k, and increased to 360k in turn. The timestamps of AIS data are almost uniformly distributed chronologically. For each set of experiments, we selected an appropriate duration T so that the amount of AIS data contained in the steady DBFT is one-third of the total amount of AIS data in this set of experiments. For example, in the first group of experiments with a total data volume of 45k pieces of AIS data, DBFT contains an average of 15k pieces of AIS data in a steady-state, which is 15 AISChain blocks, corresponding to 15 leaf nodes in DBFT. We can find that with the increase in the amount of AIS data, the FP ratio of BFT increased from 0.31% to 2.05%, while the FP ratio of DBFT increased from 0.17% to 1.13%. In addition, in each set of experiments, the FP ratio of BFT is near as twice the one of DBFT.

On the other hand, we compare the FP ratio of DBFT and BFT under accumulative data volumes. AIS data is also nearly uniformly distributed in time. And the scale of DBFT is controlled to contain 45k pieces of AIS data. As shown in Fig. 11, the FP ratio of BFT continuously rises in a proportional pattern when the accumulative volume of AIS data changes from 0 to 350k, while for DBFT, the FP ratio rises the same way as BFT when the accumulative volume of AIS data not reaches 45k, the size of the steady DBFT, but stays below 0.4% thereafter. With (11) and (12), theoretical value of the FP ratio for DBFT with a size of 45k pieces of data is 0.43%, which is similar to the FP ratio obtained in our experiments.

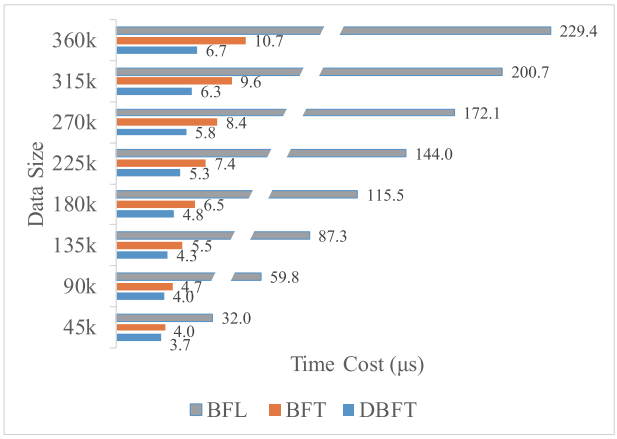

Fig. 12. Average time cost of BFL/BFT/DBFT under different data sizes.

Apart from the FP ratio, search operation's time cost of BFL, BFT, and DBFT is evaluated as well. As shown in Fig. 12, The configuration of this experiment is set to be the same as the experiment for the FP ratio, which is shown in Fig. 10. It can be noticed that, with the growth of AIS data volume, the time costs of DBFT, BFT, and BFL for searching are all increasing. Among them, the average time cost of BFL increases the fastest, which reaches 229.4 microseconds when data size is 360k. In this case, once the amount of AIS data increases, the time cost for search will probably not meet the requirements for duplication detection. On the contrary, both DBFT and BFT can guarantee very low time costs under different data sizes. Especially, the time cost of DBFT is significantly lower. As AIS data contained in Bloom filter of DBFT increases from 15k to 120k, the average time cost for search only increases from 3.7 ms to 6.7 ms.

Last but not least, we explored the impact of DBFT duration T on time cost of search. There is a time between vessels generating AIS data and AISChain recording AIS data. This experiment simulates this process with estimated delay. For each piece of AIS data in the dataset, they will be multicast to AISChain twice. For the first time, there is a delay Δt_1 compared to its timestamp (when it is generated). For the second time, there is a delay Δt_2 compared to the first time. Δt_1 and Δt_2 are independent and identically distributed, and they obey the exponential distribution: $\Delta t_1 \sim E(\lambda)$, $\Delta t_2 \sim E(\lambda)$, which both have an expectation delay λ. As shown in Fig. 13, average collecting delay is represented by the average delay λ, and DBFT duration is represented by duration T. The color in Fig. 13 represents average time cost of searching a piece of AIS data under corresponding delay λ and duration T. Dark red indicates that average time cost is large, while light red indicates that the average time cost is small. As can be observed from the black dotted line, for a certain delay distribution, there is always an optimal value of duration T, where average time cost of searching is low and stable size of DBFT is small. Therefore, by setting the duration T to a reasonable value according to the actual distribution of AIS data delay, DBFT will reach its best performance.

VIII. Conclusion

AISChain is a blockchain-based AIS data platform that realizes the goal of integrating fragmented AIS data providers, and ensures the reliable source of AIS data. To solve the conflicts

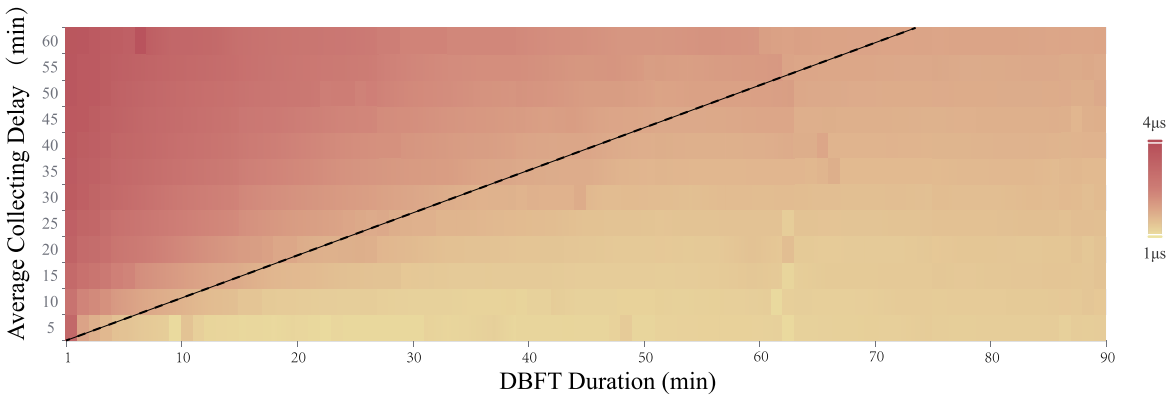

Fig. 13. Time cost of search under different collecting delay and DBFT duration.

of interest of different organizations, the dual signature is designed for the data ownership confirmation. Moreover, DBFT is proposed for efficient AIS data duplication detection. We implement a prototype of AISChain, and the performance of AISChain and DBFT are fully evaluated. AISChain can play a big role in some scenarios. For example, based on the data right confirmation provided by AISChain, AIS data can be freely traded among different parties. With the real-time AIS data provided by AISChain, better real-time vessel tracking can also be achieved, and even vessels can see faster and farther.

REFERENCES

[1] *Guideline for the Installation of a Shipborne Automatic Identifacation System (AIS)*, IMO, London, U.K., 2003.

[2] M. Svanberg, V. Santén, A. Hörteborn, H. Holm, and C. Finnsgård, "AIS in maritime research," *Mar. Policy*, vol. 106, Aug. 2019, Art. no. 103520.

[3] *Marinecadastre*. Accessed: Oct. 17, 2021. [Online]. Available: https://marinecadastre.gov/ais/

[4] *Sailwx*. Accessed: Oct. 17, 2021. [Online]. Available: https://www.sailwx.info/

[5] J. Hall, J. Lee, J. Benin, C. Armstrong, and H. Owen, "IEEE 1609 influenced automatic identification system (AIS)," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.

[6] G. C. Kessler, P. Craiger, and J. C. Haass, "A taxonomy framework for maritime cybersecurity: A demonstration using the automatic identification system," *Int. J. Mar. Navigat. Saf. Sea Transp.*, vol. 12, no. 3, pp. 429–437, 2018.

[7] M. Caprolu, R. D. Pietro, S. Raponi, S. Sciancalepore, and P. Tedeschi, "Vessels cybersecurity: Issues, challenges, and the road ahead," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 90–96, Jun. 2020.

[8] A. Androjna, T. Brcko, I. Pavic, and H. Greidanus, "Assessing cyber challenges of maritime navigation," *J. Mar. Sci. Eng.*, vol. 8, no. 10, p. 776, Oct. 2020.

[9] M. Balduzzi, A. Pasta, and K. Wilhoit, "A security evaluation of AIS automated identification system," in *Proc. 30th Annu. Comput. Secur. Appl. Conf. (ACSAC)*, 2014, pp. 436–445.

[10] *Someone is Faking the Positions of Nato Warships at Sea. It Reeks of Russia*. Accessed: Oct. 24, 2021. [Online]. Available: https://www.popularmechanics.com/military/navy-ships/a37261561/ais-ship-location-data-spoofed/

[11] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1559–1582, May 2018.

[12] W. He, J. Lei, X. Chu, S. Xie, C. Zhong, and Z. Li, "A visual analysis approach to understand and explore quality problems of AIS data," *J. Mar. Sci. Eng.*, vol. 9, no. 2, p. 198, Feb. 2021.

[13] A. Goudossis and S. K. Katsikas, "Towards a secure automatic identification system (AIS)," *J. Mar. Sci. Technol.*, vol. 24, no. 2, pp. 410–423, Jun. 2019.

[14] *R5 Secure W-AIS*. Accessed: Oct. 25, 2021. [Online]. Available: https://www.saab.com/products/r5-supreme-w-ais/

[15] C. Ray, R. Gallen, C. Iphar, A. Napoli, and A. Bouju, "DeAIS project: Detection of AIS spoofing and resulting risks," in *Proc. OCEANS*, 2015, pp. 1–6.

[16] L. Zhao, G. Shi, and J. Yang, "Ship trajectories pre-processing based on AIS data," *J. Navigat.*, vol. 71, no. 5, pp. 1210–1230, Sep. 2018.

[17] *Exactearth*. Accessed: Oct. 31, 2021. [Online]. Available: https://www.exactearth.com/

[18] *Orbcomm*. Accessed: Oct. 31, 2021. [Online]. Available: https://www.orbcomm.com/

[19] *Marinetraffic*. Accessed: Oct. 31, 2021. [Online]. Available: https://www.marinetraffic.com/

[20] *Vesseltracker*. Accessed: Oct. 31, 2021. [Online]. Available: https://www.vesseltracker.com/

[21] *Vt Explorer*. Accessed: Oct. 31, 2021. [Online]. Available: http://www.vtexplorer.com/

[22] *Fleetmon*. Accessed: Oct. 31, 2021. [Online]. Available: https://www.fleetmon.com/

[23] *Hifleet*. Accessed: Oct. 31, 2021. [Online]. Available: https://www.hifleet.com/

[24] *Ais Information Service Platform*. Accessed: Oct. 31, 2021. [Online]. Available: https://ais.msa.gov.cn/

[25] G. C. Kessler, "Protected AIS: A demonstration of capability scheme to provide authentication and message integrity," *Int. J. Mar. Navigat. Saf. Sea Transp.*, vol. 14, no. 2, pp. 279–286, 2020.

[26] R. E. Litts, D. C. Popescu, and O. Popescu, "Authentication protocol for enhanced security of the automatic identification system," in *Proc. IEEE BlackSeaCom*, May 2021, pp. 1–6.

[27] A. Aziz, P. Tedeschi, S. Sciancalepore, and R. Pietro, "Secureais—Securing pairwise vessels communications," in *IEEE CNS*, Jun. 2020, pp. 1–9.

[28] E. Bonetto, D. Brevi, and R. Scopigno, "Exploiting white space communication for increasing GNSS reliability in maritime transportation," in *Proc. IEEE AEIT*, Oct. 2016, pp. 1–6.

[29] L. Pilosu, A. Autolitano, D. Brevi, and R. Scopigno, "Exploring TV white spaces for the mitigation of AIS weaknesses," in *Proc. IEEE Symp. Commun. Veh. Technol. Benelux (SCVT)*, Nov. 2015, pp. 1–6.

[30] S. Sciancalepore, P. Tedeschi, A. Aziz, and R. Di Pietro, "Auth-AIS: Secure, flexible, and backward-compatible authentication of vessels AIS broadcasts," *IEEE Trans. Dependable Secure Comput.*, early access, Mar. 30, 2021, doi: 10.1109/TDSC.2021.3069428.

[31] S. Guo, "Space-based detection of spoofing AIS signals using Doppler frequency," *Proc. SPIE*, vol. 9121, pp. 71–76, May 2014.

[32] F. Katsilieris, P. Braca, and S. Coraluppi, "Detection of malicious AIS position spoofing by exploiting radar information," in *Proc. 16th Int. Conf. Inf. Fusion*, 2013, pp. 1196–1203.

[33] S. Mao, E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "An automatic identification system (AIS) database for maritime trajectory prediction and data mining," in *Proc. ELM*, 2016, pp. 241–257.

[34] G. K. Høye, T. Eriksen, B. J. Meland, and B. T. Narheim, "Space-based AIS for global maritime traffic monitoring," *Acta Astronautica*, vol. 62, nos. 2–3, pp. 240–245, Jan. 2008.

[35] M. Robards *et al.*, "Conservation science and policy applications of the marine vessel automatic identification system (AIS)—A review," *Bull. Mar. Sci.*, vol. 92, no. 1, pp. 75–103, Jan. 2016.

[36] *Marine Traffic: Global Ship Tracking Intelligence*. Accessed: Oct. 25, 2021. [Online]. Available: https://www.marinetraffic.com/en/ais/home/

[37] M. Castro and B. Liskov, "Practical Byzantine fault tolerance and proactive recovery," *ACM Trans. Comput. Syst.*, vol. 20, no. 4, pp. 398–461, 2002.
[38] L. Luu, V. Narayanan, C. Zheng, K. Baweja, S. Gilbert, and P. Saxena, "A secure sharding protocol for open blockchains," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 17–30.
[39] B. Tetreault, "Use of the automatic identification system (AIS) for maritime domain awareness (MDA)," in *Proc. MTS/IEEE OCEANS*, Sep. 2005, pp. 1590–1594.
[40] J.-P. Rodrigue, "The geography of maritime ranges: Interfacing global maritime shipping networks with hinterlands," *GeoJournal*, vol. 87, no. 2, pp. 1231–1244, Oct. 2020.
[41] H. Wang, C. Xu, C. Zhang, and J. Xu, "Vchain: A blockchain system ensuring query integrity," in *Proc. ACM Special Interest Group Manage. Data*, New York, NY, USA, 2020, pp. 2693–2696.
[42] C. Zhang, C. Xu, J. Xu, Y. Tang, and B. Choi, "GEM^2-tree: A gas-efficient structure for authenticated range queries in blockchain," in *Proc. IEEE Int. Conf. Data Eng.*, Apr. 2019, pp. 842–853.
[43] I. Kontopoulos, M. Vodas, G. Spiliopoulos, K. Tserpes, and D. Zissis, "Single ground based AIS receiver vessel tracking dataset," Tech. Rep., Apr. 2020. Accessed: Sep. 6, 2021. [Online]. Available: https://zenodo.org/record/3754481/

Yongshuai Duan received the B.Eng. degree in computer science and technology from Shanghai Jiao Tong University, Shanghai, China, in 2020, where he is currently pursuing the master's degree with the Department of Computer Science and Engineering. His research interests include the Internet of Things, blockchain, and confidential computing.

Junqin Huang received the B.Eng. degree in computer science and technology from the University of Electronic Science and Technology of China, Chengdu, China, in 2018. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include crowdsensing, the Internet of Things, blockchain, and mobile computing.

Jiale Lei received the B.Eng. degree in computer science and technology from the Shanghai University of Finance and Economics, Shanghai, China, in 2020. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai. His research interests include wireless communication, mobile computing, and machine learning.

Linghe Kong (Senior Member, IEEE) received the B.Eng. degree in automation from Xidian University in 2005, the master's degree in telecommunication from Telecom SudParis in 2007, and the Ph.D. degree in computer science from Shanghai Jiao Tong University in 2013. He is currently a Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. Before that, he was a Post-Doctoral Researcher with Columbia University, McGill University, and the Singapore University of Technology and Design. His research interests include the Internet of Things, 5G, blockchain, and mobile computing.

Yibin Lv received the B.S. degree in ship and offshore engineering and the M.E. degree in EMBA from Shanghai Jiao Tong University in 1993 and 2005, respectively. He was the Section Manager of the Technology Department, BOMTA, Shanghai, from 1996 to 1997; and the Manager of the Technology Department and the Enterprises Division of China Shipping (Group) Company, from 1998 to 2015. He is currently the Assistant General Manager with China Shipping Industry Company. He has been in charge of fleet planning, newbuilding, second-hand ship purchasing, ship maintenance, repairing, energy saving, safety, disposal, and shipbuilding and repairing for 23 years. Experiencing nearly all the newbuilding projects of CSG in Japan, South Korea, and China.

Zhiliang Lin received the B.Eng. degree in ocean engineering and the Ph.D. degree in naval architecture and ocean engineering from Shanghai Jiao Tong University in 2005 and 2010, respectively. He was a Visiting Professor with Newcastle University, U.K., in 2013. He is currently an Associate Professor with the Department of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University. His research interests include fluid dynamics, nonlinear mechanics, water waves, numerical simulation, and image recognition.

Guihai Chen received the B.S. degree from Nanjing University in 1984, the M.E. degree from Southeast University in 1987, and the Ph.D. degree from The University of Hong Kong in 1997. He was a Visiting Professor with many universities including the Kyushu Institute of Technology, Japan, in 1998; The University of Queensland, Australia, in 2000; and Wayne State University, USA, from 2001 to 2003. He is currently a Distinguished Professor with Shanghai Jiao Tong University, China. He has a wide range of research interests with a focus on sensor networks, peer-to-peer computing, and high-performance computer architecture and combinatorics.

Muhammad Khurram Khan (Senior Member, IEEE) is currently working as a Professor of cybersecurity with the Center of Excellence in Information Assurance, King Saud University, Saudi Arabia. He is the Founder and the CEO of the "Global Foundation for Cyber Studies and Research," an independent and non-partisan cybersecurity think-tank in Washington, DC, USA. He has published more than 450 papers in the journals and conferences of international repute. He is an inventor of ten US/PCT patents. He has edited ten books/proceedings published by Springer-Verlag, Taylor and Francis, and IEEE. His research areas of interests are cybersecurity, digital authentication, the IoT security, biometrics, multimedia security, cloud computing security, cyber policy, and technological innovation management. He is a fellow of the IET, U.K.; BCS, U.K.; and FTRA, South Korea. He is on the editorial board of several journals, including IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, *IEEE Communications Magazine*, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON CONSUMER ELECTRONICS, *Journal of Network and Computer Applications* (Elsevier), IEEE ACCESS, *IEEE Consumer Electronics Magazine*, *PLOS One*, and *Electronic Commerce Research*. He is the Editor-in-Chief of *Telecommunication Systems* (Springer-Nature) with its recent impact factor of 2.314 (JCR 2021). He is also the Editor-in-Chief of *Cyber Insights Magazine*. His detailed profile can be visited at http://www.professorkhurram.com