Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.062556

ARTICLE

Research on Quantification Mechanism of Data Source Reliability Based on Trust Evaluation

Gaoshang Lu[#], Fa Fu[*,#] and Zixiang Tang

School of Computer Science and Technology, Hainan University, Haikou, 570228, China

*Corresponding Author: Fa Fu. Email: fufa@hainanu.edu.cn

[#]These authors contributed equally to this work

Received: 20 December 2024; Accepted: 10 March 2025; Published: 19 May 2025

ABSTRACT: In the data transaction process within a data asset trading platform, quantifying the trustworthiness of data source nodes is challenging due to their numerous attributes and complex structures. To address this issue, a distributed data source trust assessment management framework, a trust quantification model, and a dynamic adjustment mechanism are proposed. The model integrates the Analytic Hierarchy Process (AHP) and Dempster-Shafer (D-S) evidence theory to determine attribute weights and calculate direct trust values, while the PageRank algorithm is employed to derive indirect trust values. The direct and indirect trust values are then combined to compute the comprehensive trust value of the data source. Furthermore, a dynamic adjustment mechanism is introduced to continuously update the comprehensive trust value based on historical assessment data. By leveraging the collaborative efforts of multiple nodes in the distributed network, the proposed framework enables a comprehensive, dynamic, and objective evaluation of data source trustworthiness. Extensive experimental analyses demonstrate that the trust quantification model effectively handles large-scale data source trust assessments, exhibiting both strong trust differentiation capability and high robustness.

KEYWORDS: Trust evaluation; data source reliability; distributed network; quantification mechanism

1 Introduction

In today's digital era, data has become the cornerstone of the information society, making its reliability and security matters of great concern. However, the credibility of data is inherently tied to its source—that is, the reliability of the data provider [1]. This relationship directly impacts data trustworthiness, and in the era of big data, assessing data source reliability is even more critical. A robust assessment framework not only enhances the efficiency of data exchange and sharing but also mitigates security risks while safeguarding data privacy and integrity. Moreover, within data asset trading platforms, evaluating the reliability of a data source is the first and foremost step before proceeding with operations such as data mining, pricing, and protection. If the data source itself is untrustworthy, any subsequent operations on the data become meaningless. Therefore, to ensure the security and accuracy of data transactions on these platforms, assessing data source credibility has become a key research focus.

However, evaluating the trustworthiness of data sources presents several challenges. The diversity, dynamic nature, and inherent uncertainty of data sources add complexity to the assessment process. Additionally, data source nodes may engage in malicious activities such as data tampering and impersonation [2], leading to unreliable and inaccurate evaluation results. As a result, establishing an effective trust model capable of identifying and filtering out malicious data sources during transactions on data asset trading platforms is a pressing issue in current research.

To date, research on data source trustworthiness primarily falls into the following categories: studies focused on Mobile Ad-hoc Networks, trust perception and interaction models, algorithms and protocols for wireless sensor networks, trust assessment mechanisms for IoT networks, and trust evaluation frameworks based on various policy-driven approaches. Each of these areas will be discussed in detail in the following sections.

1.1 Mobile Ad-hoc Networks

Mobile Self-Organizing Network (MANET) is a multi-hop wireless network with rapid deployment capability, which is characterized by the lack of predefined network infrastructure and centralized network management. Due to the lack of routing infrastructure, MANETs cannot use traditional Internet protocols for routing, name resolution and trust establishment. In such networks, each node functions as both a host and a router and is able to collaboratively communicate with other nodes outside its transmission range through a multi-hop strategy. All nodes execute pre-agreed routing protocols to deliver packets, and the effective execution of these protocols relies on the consistent and well-intentioned behavior of all participating nodes. Thus, MANET exhibits a distributed and self-organized communication model.

Gera et al. emphasized the critical role of trust in secure routing mechanisms in Mobile Ad-hoc Networks (MANETs) and introduced an opinion-based trust assessment model for identifying misbehaving nodes and determining the most reliable data transmission path [3]. However, since the model proposed in the article relies on the continuous monitoring and evaluation of each node's behavior of other nodes, this increases the computational and communication burden on the network. In addition, the model is overly sensitive to network conditions (e.g., node congestion or crosstalk), which may lead to misjudgment and misidentification of well-behaved nodes as poorly behaved nodes. Thus, Hosmani et al. proposed the Robust and Reliable Secure Clustering and Data Transmission (R2SCDT/RRSCDT) protocol for in-vehicle Ad-Hoc networks, which utilizes trust assessment to address the challenges posed by malicious vehicles [4]. The research mainly focuses on detecting and mitigating the impact of malicious nodes on data transmission without sufficiently considering other types of network attacks such as man-in-the-middle attacks or gray hole attacks. While user privacy protection is an important issue during trust assessment and data transmission, the article does not explicitly mention privacy protection measures.

1.2 Trust Perception and Trusted Network Modeling

Trust-aware and trustworthy network modeling is an important research direction in the field of contemporary network security, aiming to cope with increasingly complex network threats by constructing a secure, reliable, controllable and self-recovery network environment. Although traditional security methods can effectively defend against external attacks, it is difficult to effectively identify and defend against potential malicious nodes within the network. Therefore, the trusted network model has emerged, which ensures that data transmission in the network is safe and secure, and node behavior is controllable and manageable through the implementation of trust models, secure routing protocols, and strict management policies. At the same time, the model also has the ability to recover itself in the event of attack or failure, and thus is widely used in many fields such as military, emergency communication and Internet of Things (IoT).

Xu developed a data-driven trust assessment model that relies on sensing sources, with a special focus on the monitoring module used to assess the trustworthiness of sensing nodes [5]. However, due to the high dynamics of wireless sensor networks, nodes may join or leave the network frequently, and these trust models cannot effectively handle such dynamics, and their scalability in large-scale or high-density networks is not fully validated.Xu et al. proposed a cloud-based Trust Perception and Interaction Model (CTAIM) for assessing trust in social interaction data. The model utilizes Bayesian algorithm and MapReduce-based computation for effective trust assessment. It combines local and cloud trust maintenance schemes and aims to provide secure and neutral trustworthiness assessment for social interaction data [6]. Ansheng proposed a Comprehensive Trust-based Trusted Network Assessment Model (ComTrust), which combines subjective evaluation data and objective attribute data to calculate the trustworthiness of nodes in the network, effectively identifies malicious nodes, improves the success rate of the service, and outperforms other trust assessment methods in terms of accuracy and predicted user ratings [7].

1.3 Algorithms and Protocols for Wireless Sensor Networks

As a large-scale wireless network composed of a large number of low-power, self-organized sensor nodes with comprehensive sensing, computing and communication capabilities, wireless sensor networks (WSNs) achieve extensive coverage of the monitoring area and accurate information collection through an efficient multi-hop communication mechanism. The network integrates cutting-edge technologies such as microelectronics, embedded computing, wireless communication and distributed information processing, and shows the potential for wide application and great prospects for development in the fields of environmental monitoring, military reconnaissance, smart home, medical care and disaster relief. Meanwhile, WSNs, as a set of decentralized, energy-constrained but environment-aware small nodes, face multiple challenges such as deployment routing, data collection, energy consumption and security attacks. With the rapid development of wireless communication technology, WSNs are gradually becoming a key component of next-generation applications, which are widely used in many socially important fields, such as industry, weather monitoring, military and healthcare. In this context, the design of algorithms and protocols for wireless sensor networks is particularly important, aiming to optimize node transmission capability, improve communication efficiency, ensure data security, and effectively deal with various challenges.

In recent years, there has been a proliferation of trust assessment methods for WSNs environments. For example, Mythili et al. proposed the Spatial and Energy-Aware Trusted Dynamic Distance Source Routing (SEAT-DSR) algorithm for secure data communication in WSNs, which incorporates a hierarchical trust mechanism based on various attributes of the sensor nodes [8]. Rajasekaran also proposed a location-energy-aware Trusted Distance Source Routing protocol for WSNs, with the goal of improving the network lifetime and introduces a new trust mechanism that considers communication, data, energy and recommendation attributes of sensor nodes [9]. Lv et al. proposed a Trusted Big Data Collection (TBDC) scheme that utilizes wireless sensor networks for data sensing, collection and storage. The trust evaluation model therein includes direct trust, recommended trust, link trust, and backhaul trust, and the trust value is dynamically updated by the ω-FCM algorithm to ensure the authenticity of the collected data [10]. However, in the ω-FCM algorithm, the choice of parameters has a great impact on the performance of the algorithm, and the article does not fully discuss how to choose the optimal parameters. In addition, user interaction and feedback are very important in the trust assessment process, and the article does not consider integrating user input into the trust assessment model. Based on this, Lopez et al. propose a formal distributed network monitoring approach for assessing trust management behavior in collaborative systems. By analyzing network traffic from multiple observation points, it is possible to detect untrustworthy behaviors that may not be detected by traditional single-point monitoring. The aim is to provide rapid feedback on trustworthiness to enhance trust management systems [11]. However, the article uses the NTP protocol to synchronize network traces, but does not discuss possible synchronization problems or the impact of inaccurate synchronization on monitoring and trust assessment. In addition, the methods and tools in the article may need to run for a long period of time to continuously monitor network behavior, but the feasibility and maintenance cost of long-term monitoring needs to be further investigated.

1.4 Trust Evaluation Mechanisms for IoT Networks

The trust assessment mechanism for IoT network is a comprehensive system, the core of which is to ensure that the IoT network can meet the needs of users through the comprehensive assessment of security, reliability and compliance, so as to establish a solid foundation of trust in the hearts of users. The mechanism covers the formulation of assessment standards, the selection of assessment methods, the implementation of regular assessment and continuous improvement, etc., which is of far-reaching significance for promoting IoT technological innovation, safeguarding users' rights and interests, and promoting the healthy development of the whole industry. With the explosive growth of IoT applications, it is often difficult for users to judge the source and reliability of data as they swim in the ocean of data, making trust a key factor in managing these interactions and building a trusted environment. However, due to the diversity of user groups and differences in the understanding of trust, it has become particularly difficult to establish trust relationships between users. Traditional trust management systems calculate trust mainly based on the relationships between end entities and behaviors, but the diversity and complexity of data in IoT environments make this calculation challenging. Therefore, trust assessment mechanisms in IoT networks need to pay more attention to the reliability, timeliness, and other attributes specific to the data to ensure that the trust of the end entities is accurately assessed. Particularly in social IoT, quantifying the trustworthiness of data providers is crucial for generating trustworthy services, which further highlights the urgency and importance of developing appropriate schemes to address such issues.

Building on this foundation, Jayasinghe et al. addressed the challenges of trust assessment in IoT systems by extending the establishment of entity trust to include the assessment of data item trust. Their research focused on extracting, aggregating, evaluating and predicting data trust metrics within their framework [12]. However, data trust assessment models rely too much on historical data and previous user interactions, which limits the ability to quickly assess the trustworthiness of new entities or new types of data. The article also mentions the use of machine learning techniques to determine the weights of trust attributes, which is computationally complex and costly when dealing with large-scale data. In addition to this, Li et al. proposed a novel trust-based service computing scheme in the social Internet of Things (IoT) to enhance trust services based on trusted data [13]. Finally, Lin et al. proposed a granular trust evaluation mechanism (DFTE) for data fusion and transfer learning authorization to assess user trust and ensure the reliability of data sources in IoT environments for the needs of secure IoT networks [14]. Together, these studies emphasize the importance of data source trust assessment in different network environments to improve data security and reliability.

1.5 Trust Evaluation Framework Based on Various Types of Strategies

In addition to the above research on data source trustworthiness, there are also trust assessment management frameworks based on various types of strategies such as node behavioral strategies, process points, and D-S evidence theory, which will be described below.

Aldini et al. defined a process algebra framework for modeling collective adaptive systems that use trust and reputation to manage interactions among nodes, and used a process algebra-based approach to express the global state of the system, defining the system's interaction behavior through parallel combinations and communication rules [15]. However, the trust update mechanism in the article may need further research to ensure timely response to trust changes in a dynamically changing network environment. Feng et al. also proposed a trust evaluation algorithm based on node behavioral policies and D-S evidence theory (NBBTE). The method combines node behavioral strategies with modified evidence theory to assess the trust of nodes through three trust levels (completely distrustful, uncertain and completely trustful). In addition, the article defines concepts and rules related to D-S evidence theory to cope with the subjectivity of trust assessment [16]. Although the article mentions that trust values change over time and behavior, it does not detail how to quickly adapt to rapidly changing environments in the network, and that a compromise between trust assessment and resource consumption may be needed in order to improve the security of the network, but it does not detail a specific trade-off strategy. Belov et al. introduce a methodology called ”Trust in Crowds (TiC), a computational trust assessment system for evaluating trust in rapidly generated open media data. The TiC system includes components for data processing, metadata extraction, feature evaluation, trust network generation, and data storage. It evaluates trustworthiness based on features such as user ratings, corroborating events, topic similarity, and publisher trust, and uses a modified version of the Tidal Trust algorithm for global trustworthiness measurement. The global Tidal Trust (GTT) algorithm of this system efficiently computes the trustworthiness of the nodes in the graph by optimizing caching and parallelization, proving the accuracy and scalability of the evaluation [17]. However, the system currently only uses Factiva as a data source for news articles and Open Source Center as a data source for blog data, which may limit the diversity and comprehensiveness of the data. In addition, although an accuracy assessment was performed, this was conducted by a single subject matter expert and a broader user assessment is needed to validate the conclusions of the system. Gao et al. proposed the Info-Trust mechanism, a multi-criteria, adaptive trustworthiness calculator mechanism for evaluating the sources of social media information. The mechanism integrates identity-based trust, behavior-based trust, relationship-based trust, and feedback-based trust, and uses a combined OWA-WMA algorithm to dynamically assign weights to these factors, and accurately identifies trusted information sources through a comprehensive assessment of trustworthiness [18]. However, the study was mainly conducted based on the Sina Weibo dataset, which limits the generalization ability of the model to adapt to other social media platforms. In addition, although an adaptive weight assignment method is proposed in the paper, it does not fully consider the dynamic characteristics of trust factors over time.

Through the research and analysis of the above studies, it is found that there are still limitations in quantifying the credibility of data sources, especially in complex and ever-changing network environments. Therefore, this article proposes a data source credibility quantification mechanism based on trust evaluation to more accurately evaluate and manage data trust in the network. The mechanism mainly includes a distributed data source trust evaluation management framework, a data source credibility quantification and dynamic adjustment mechanism based on improved D-S evidence theory. The specific content of each section is arranged as follows:

Section 2 introduces the distributed data source trust evaluation management framework, detailing the overall structural design of the framework, the role positioning and core responsibilities of each participant, and their working principles. In addition, this section also provides an in-depth explanation of the collaborative workflow mechanism among all participants within the framework.

Section 3 explains the algorithm for quantifying data source reliability, which consists of three components: direct trust value quantification, indirect trust value quantification, and comprehensive trust value quantification.

Section 4 introduces a dynamic adjustment mechanism for evaluating the trustworthiness of data source nodes.

Section 5 is about experiments and analysis, including four parts: the experimental process of direct, indirect, and comprehensive trust values of data source nodes, experimental result analysis, comparative experiments, and comparative experiment result analysis.

Section 6 first summarizes the work of this article, then explores possible problems, and finally proposes the next step of work planning.

2 Trust Evaluation Management Framework for Distributed Data Source

With the rise of the big data era and the rapid expansion in both the number and variety of data sources, accurately and efficiently assessing data source reliability has become a pressing challenge in the field of data processing. In the research of data source trust evaluation, a key issue lies in constructing a scientific and well-structured quantitative framework that ensures objectivity and rigor while considering multiple influencing factors. Traditional data source trust evaluation methods often rely on centralized trust management mechanisms; however, these approaches struggle with inefficiency and single points of failure when dealing with large-scale, distributed data sources. To address these limitations, a data source reliability quantification mechanism based on distributed trust evaluation is proposed, as illustrated in Fig. 1.

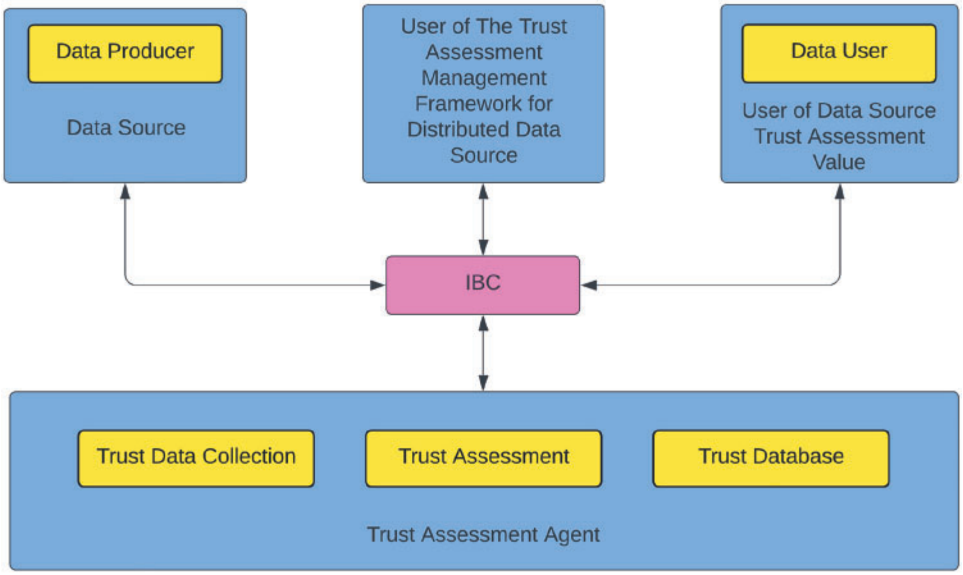
Data Producer
Data Source
User of The Trust Assessment Management Framework for Distributed Data Source
Data User
User of Data Source Trust Assessment Value
IBC
Trust Data Collection
Trust Assessment
Trust Database
Trust Assessment Agent

Figure 1: Trust evaluation management framework for distributed data source

In the distributed data source trust evaluation management framework, the primary participants include: (1) Trust Assessment Agents–These agents consist of data source sub-nodes and user sub-nodes. The trust assessment agent node and all its sub-nodes are interlinked, meaning they share direct trust relationships. The trust assessment agent is primarily responsible for collecting various attribute values necessary for direct trust evaluation of its data source sub-nodes. These attributes include the data source node's arrival rate, performance, security protection capabilities, authentication status, and historical evaluation records. Once these attribute values are gathered, they are analyzed, weighted, and assigned appropriate values to compute the direct trust value of the node, which is then stored in the trust database. Additionally, the trust assessment agent processes trust value requests from users, the data asset trading platform, and other trust assessment agents. If no new historical evaluation data has been generated since the last request, the previously stored direct trust value is retrieved from the database and returned. The requestor then calculates the composite trust value by integrating the acquired direct trust value with the computed indirect trust value. (2) Data Producers–These are the original providers of data, supplying it for either profit or public interest. They typically provide structured data (e.g., tables), semi-structured data (e.g., XML and JSON), and unstructured data (e.g., text, images, audio, and video). A data producer can also function as a data source. (3) Data Sources–These entities sell data on the data asset trading platform for profit and are the primary objects of trust assessment in this framework. A data source may also be a data producer. (4) Framework Users–These individuals or entities are responsible for the design, development, testing, and maintenance of the distributed data source trust evaluation management framework. They can request the direct trust value of any data source node and calculate both indirect and comprehensive trust values. (5) Data Source Trust Assessment Value Users–A subclass of sub-nodes within the trust assessment agent, these participants can request the comprehensive trust value of any data source from the trust assessment agent. This structured framework ensures an efficient and systematic approach to evaluating the trustworthiness of data sources within a distributed environment.

3 Quantification of Data Source Reliability

3.1 Direct Trust

Direct trust is a key indicator in assessing the reliability of data sources, while the weight assignment of attributes serves as the foundation for ensuring the accuracy and credibility of the evaluation results. In this study, the Analytic Hierarchy Process (AHP) is employed to assign weights to the attributes influencing data source trust assessment. Specifically, this research focuses on analyzing five major attributes: data arrival rate (A1), data source performance (A2), data security protection capability (A3), data source authentication (A4), and data source historical evaluation (A5). By applying the AHP method, appropriate weights are assigned to these attributes, ensuring a comprehensive and precise assessment of data source trustworthiness.

$$\lambda_{max} = \sum_{i=1}^{n} \frac{[AW]_i}{nw_i} \tag{1}$$

$$CI = \frac{\lambda_{max} - n}{n-1} \tag{2}$$

$$CR = \frac{CI}{RI} \tag{3}$$

To conduct AHP analysis, firstly, it is necessary to construct a hierarchy of trust assessment attributes of the data source, clarify the hierarchical relationship between each attribute, and invite domain experts to score and construct a judgement matrix a_{ij} in order to determine the relative importance between the attributes. Among them, the value of attribute i relative to attribute j is P (i, j), and the value of attribute j relative to attribute i is P (j, i) = 1/P (i, j). When P (i, j) = 1, it represents that the importance of attribute i is equal to that of attribute j, and P (j, i) = 1; When P (i, j) > 1, it indicates that the importance of attribute i to the decision objective is greater than that of attribute j, while P (j, i) < 1; When P (i, j) < 1, it indicates that the importance of attribute i to the decision objective is less than that of attribute j, and P (j, i) ≥ 1. Specify P_{max} as the maximum value of P, P_{max}(i,j) = 9, Representing that attribute i is absolutely important relative to attribute j; P_{min} is the minimum value of P, P_{min}(i,j) = 1/9, The representative attribute j is absolutely important to attribute i. Second, consistency judgement is performed. Calculate $\lambda_{\max}$ by Eq. (1), calculate consistency index *CI* by Eq. (2), random consistency index *RI* is a fixed value, which is obtained by querying the *RI* table of hierarchical analysis method, and finally calculate the consistency ratio *CR* by Eq. (3). If CR < 0.1, the consistency judgement is passed, and the eigenvectors are the values of each attribute's weights w_i. if CR ≥ 0.1, adjust and improve a_{ij}, so that the CR < 0.1. $\lambda_{\max}$, consistency index *CI* and consistency ratio *CR* are shown in Eqs. (1)–(3), respectively. Finally, the corresponding weights of each attribute are calculated by arithmetic mean, a_{ij} is normalised by column to get a'_{ij}, and the average of row i is calculated to get the weight value of attribute i.

After calculating the weight of each attribute that affects the trustworthiness of the data source, the direct trust value of each data source is calculated. The identification framework Ω is a finite set summary of all potential values of the study object, and each element in Ω is independent of each other, and a power set is constructed based on a subset of Ω, denoted as 2^{Ω}.

Definition 1: The basic probability distribution function m, satisfies: $m{:}2^{\Omega} \rightarrow [0,1]$, $m(\varnothing) = 0$, $\sum\limits_{A\subseteq\Omega} m(A) = 1$, where $m(A)$ is the number of basic probabilities of the hypothesis set A, indicating the level of trust in the hypothesis set A, and $m(A) \neq 1 - m(\neg A)$. According to the weight value of each attribute calculated by AHP hierarchical analysis method, the basic probability distribution function of each attribute's trust in the data source is adjusted according to Eq. (4), and the new basic probability distribution function of the fusion attribute weights is obtained.

$$\begin{cases} U(m_i)m_i(A), & A \neq \Omega \\ 1 - \sum\limits_{B\subseteq\Omega} U(m_i)m_i(B) & A = \Omega \end{cases} \tag{4}$$

Definition 2: Based on the basic probability distribution function, the trust function (also known as the lower bound function) is defined as follows:

$$\begin{cases} Bel : 2^{\Omega} \rightarrow [0,1] \\ Bel(A) = \sum\limits_{B\subseteq A} m(B) \end{cases} \tag{5}$$

where $\mathrm{Bel}(\varnothing) = 0$, $\mathrm{Bel}(\Omega) = 1$, and $Bel(A) + Bel(\neg A) \leq 1$. When the set A consists of only one element, $Bel(A) = m(A)$.

Definition 3: Based on the belief function, the likelihood function (also known as the ceiling function) is defined as follows:

$$\begin{cases} Pl : 2^{\Omega} \rightarrow [0,1] \\ Pl(A) = 1 - Bel(\neg A) \end{cases} \tag{6}$$

where $Pl(\varnothing) = 0$, $Pl(\Omega) = 1$, $Pl(A) + Pl(\neg A) \geq 1$, $Pl(A) \geq Bel(A)$.

Definition 4: Based on the trust function and the likelihood function, Fig. 2 can be obtained as shown in Fig. 2, $[0, \mathrm{Bel}(A)]$ is the support interval of A, i.e., the trust interval of the hypothesis set A is true; $(Bel(A), Pl(A)]$ denotes the interval in which A is both true and false, but based on the available evidence, it cannot be differentiated whether A is true or false; $[0, Pl(A)]$ is the interval in which A is unsuspecting, i.e., the available evidence proves that hypothesis set A is a non-false interval; $(Pl(A), 1]$ denotes the interval in which the hypothesis set A must be false.

Definition 5: Based on the trust function and likelihood function, the class probability function is obtained as follows:

$$f(A) = Bel(A) + \frac{|A|}{|\Omega|}(Pl(A) - Bel(A)) \tag{7}$$

where $|A|$ and $|\Omega|$ denote the number of elements in the hypothesis set A and the identification framework Ω, respectively, and $f(A)$ is the direct trust degree of the hypothesis set A.

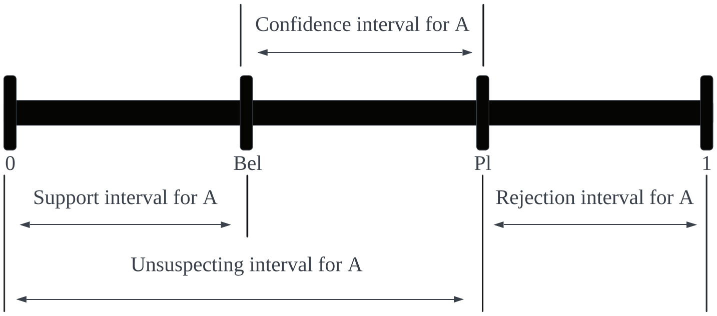

Figure 2: D-S evidence theory confidence interval

When there are two and more groups of basic probability assignment functions, it is necessary to use the Dempster's evidence combination formula to calculate the orthogonal sums of the subsets in the original basic probability assignment function separately, which in turn leads to the basic probability assignment function under multiple sets of evidence. When there are two different basic probability assignment functions m_1 and m_2, the calculation of their orthogonal sums $m = m_1 \oplus m_2$ is shown in Eq. (8).

$$m_1 \oplus m_2 = \frac{1}{K} \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \tag{8}$$

where $m(\varnothing) = 0$, K is the normalisation factor, reflecting the degree of conflict between the evidence, which is calculated as shown in Eq. (9). If there are more than two groups of basic probability distribution functions, after calculating the orthogonal sum m of m_1 and m_2, the orthogonal sum of m and other groups of basic probability distribution functions is calculated in the same way. According to Formulas (4)–(8) and all groups of basic probability distribution function for evidence after the combination of the final fusion of the basic probability distribution function, and then according to Formulas (4)–(7) and the class probability function to calculate the final value of direct trust.

$$K = 1 - \sum_{x \cap y = \varnothing} m_1(x) \cdot m_2(y) = \sum_{x \cap y \neq \varnothing} m_1(x) \cdot m_2(y) \tag{9}$$

3.2 *Indirect Trust*

While direct trust assessments play a crucial role in evaluating data source trustworthiness, indirect trust assessments are equally essential. A data source that receives more incoming links from data asset trading platforms and users is likely to have higher usage frequency and quality. Therefore, this study employs the PageRank algorithm to calculate the indirect trust value of a data source. The PageRank (PR) value represents the indirect trust value and is computed as shown in Eq. (11).

$$PR_{init} = \frac{1}{N} \tag{10}$$

$$PR(v_i) = \frac{1-d}{N} + d \sum_{i=1}^{n} \frac{PR(v_i)}{L(v_i)} \tag{11}$$

where v denotes the node pointing to node u, $PR(v)$ denotes the PR value of node v, and $L(v)$ denotes the number of outgoing chains of node v. The same initial PR value and damping factor for jumping to other nodes are set for each node, the initial PR value is set as in Eq. (10), where N represents the number of nodes in the network, and the damping factor is set to 0.85. During each iteration, each node jumps to other nodes randomly and with equal probability, and after a few iterations, the *PR* value of each node tends to be converged and stabilised, i.e., we get the indirect trust value of each node.

3.3 Comprehensive Trust

In the distributed data source trust assessment management framework, the trust assessment agent, along with data source nodes, data user nodes, and data source trust assessment nodes, collectively form a trust propagation network. In the directed graph of this network, as illustrated in Fig. 3, the arcs represent trust relationships. Specifically, an arc pointing from a tail node to a head node indicates that the tail node directly trusts the head node. Each trust assessment agent node stores the direct trust values of the nodes it directly trusts. When a data source trust assessment node seeks to determine the trust level of data source Fig. 3, but there is no direct trust relationship between them, the trust value must be obtained indirectly through intermediary nodes in the trust propagation network. In such cases, the trust value propagates through the network, with nodes along the propagation path responding and cascading back the trust value in reverse order. In this network, there are two possible trust propagation paths to obtain the trust value of data source 3. The first path is: Data source trust assessment node→Trust assessment agent 1→Trust assessment agent 2→Trust assessment agent 3; The second path is: Data source trust assessment node→Trust assessment agent 4→Trust assessment agent 3. When multiple propagation paths exist, trust values may attenuate with longer paths and strengthen with shorter ones. Therefore, the data source trust assessment node prioritizes the shorter propagation path: Data source trust assessment node→Trust assessment agent 4→Trust assessment agent 3. Subsequently, trust assessment agent 3 responds and cascades back the trustworthiness value of data source 3. Based on the principles of direct trust values, indirect trust values, and the trust propagation network, the formula for calculating the comprehensive trust value is presented in Eq. (12).

$$T_c = T_d(N_1) \otimes T_r(N_1) \cdots T_d(N_n) \otimes T_r(N_n) \tag{12}$$

where N_1 to N_n represent the intermediary nodes between the trust value requesting node and the trust value responding node, $T_d(N_1)$ denotes the direct trust value assigned by the requesting node to the next node in the propagation path, and T_r represents the indirect trust value.

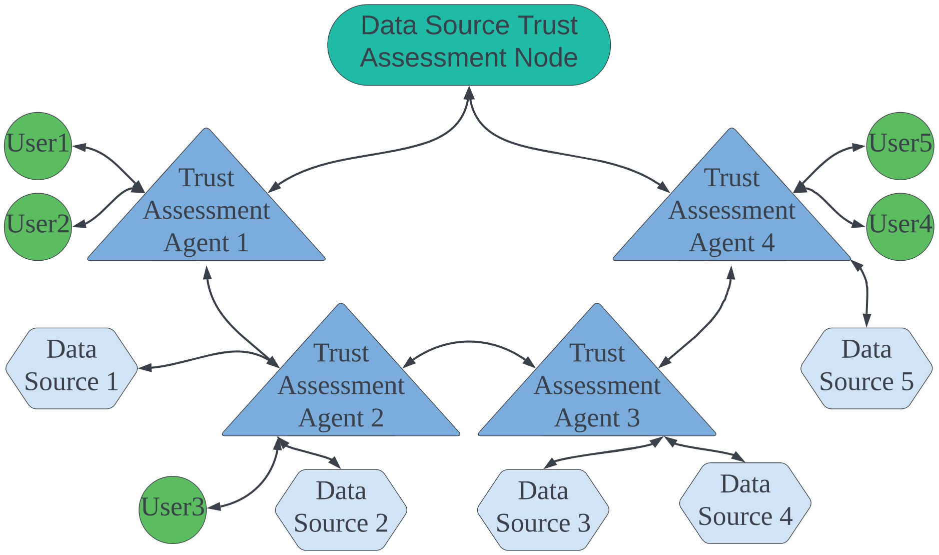

Figure 3: Directional map of the trust transmission network

4 Dynamic Adjustment Mechanism

In the study of data source reliability quantification mechanisms, trust assessment is a continuous and dynamic process. Accurately evaluating data source reliability requires not only considering inherent attributes such as current data quality and data source stability but also dynamically adjusting assessments based on real-time user historical evaluations. This dynamic adjustment mechanism is crucial for accurately reflecting the actual reliability of a data source at any given moment. The flowchart illustrating this mechanism is shown in Fig. 4.

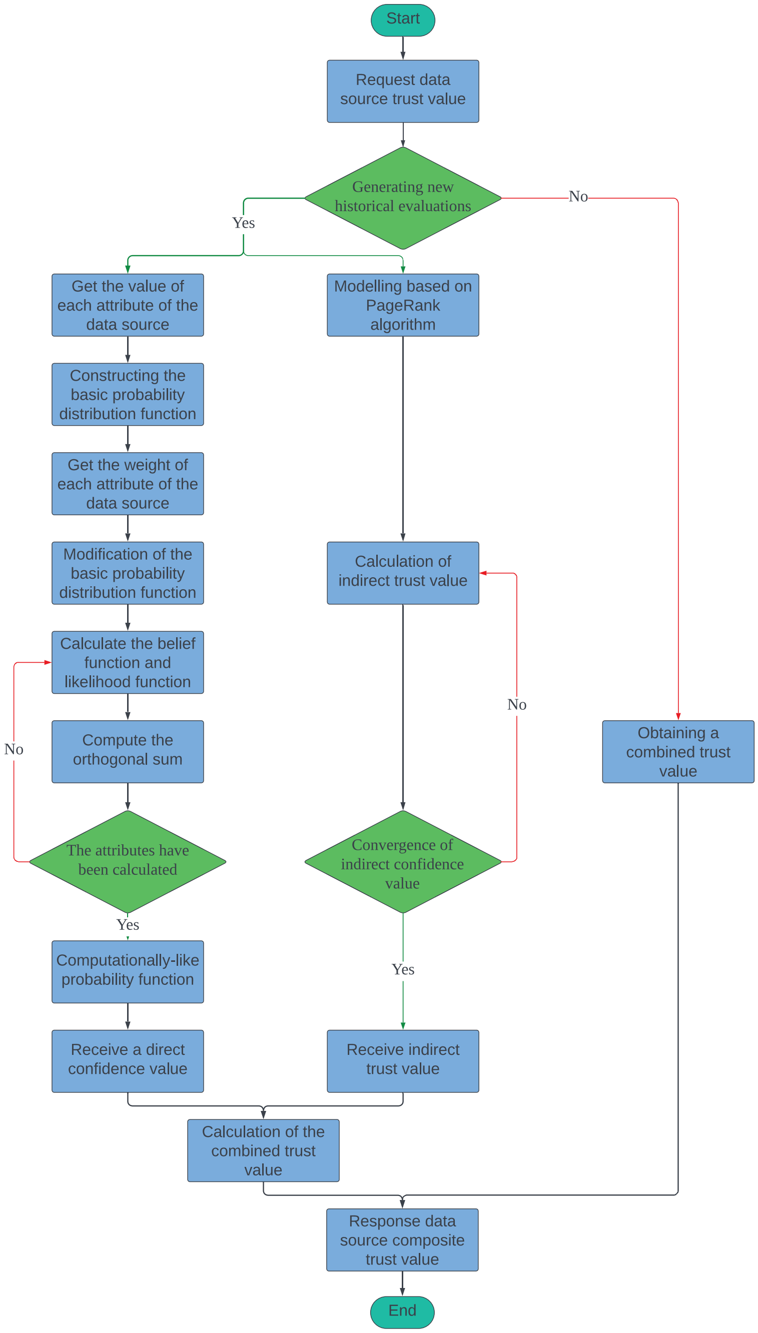
Start
Request data source trust value
Generating new historical evaluations
Yes
No
Get the value of each attribute of the data source
Modelling based on PageRank algorithm
Constructing the basic probability distribution function
Get the weight of each attribute of the data source
Modification of the basic probability distribution function
Calculation of indirect trust value
Calculate the belief function and likelihood function
No
Compute the orthogonal sum
No
Obtaining a combined trust value
The attributes have been calculated
Convergence of indirect confidence value
Yes
Computationally-like probability function
Yes
Receive a direct confidence value
Receive indirect trust value
Calculation of the combined trust value
Response data source composite trust value
End

Figure 4: Flow chart of the dynamic adjustment mechanism

When a user requests the system to retrieve the comprehensive trust value of a data source, the system first determines whether a new historical evaluation has been generated since the user's last request. If a new evaluation exists, the system retrieves the relevant attributes outlined in Section 2 and calculates the comprehensive trust value according to Eqs. (1)–(12). If no new evaluation has been generated, the system directly retrieves the previously stored comprehensive trust value from the trust database and returns it as a response. By incorporating real-time historical evaluations, the system ensures that the most up-to-date assessment of the data source is obtained, enabling dynamic updates to trust evaluations and enhancing both the timeliness and accuracy of the trust assessment.

5 Experiments

To evaluate the effectiveness of the proposed trust model, a big data network comprising 100 nodes was simulated using NetLogo. In this simulation, nodes 1, 11, 21, 31, 41, 51, 61, 71, 81, and 91 were designated as data source nodes, while the remaining 90 nodes served as user nodes. The experiment followed these steps:

- Attribute Weight Quantification: The Analytic Hierarchy Process (AHP) was applied using yaahp 12.12.8367 to determine the weight of each data source attribute.
- Direct Trust Value Calculation: IntelliJ IDEA 2021.3.3 was used to compute the direct trust value of data sources by leveraging an improved D-S evidence theory and multi-condition combination rules.
- Indirect Trust Value Computation: The PageRank algorithm was executed in NetLogo to quantify indirect trust values. After multiple trials, the optimal damping factor was determined to be 0.85, which was then set accordingly.
- Comprehensive Trust Value Calculation: Utilizing IntelliJ IDEA 2021.3.3, the comprehensive trust value algorithm was executed, integrating both direct and indirect trust values to obtain the final trust assessment of each data source.

5.1 Direct Trust Value Calculation

When calculating the weights of each attribute of the data source through AHP, if the data source has no historical evaluation, it means that the data source is a new data source, and the weights of each attribute of the data source are calculated through Eqs. (1)–(3) as shown in Fig. 5; if the data source has historical evaluation, the weights of each attribute of the data source are calculated through Eqs. (1)–(3) as shown in Fig. 6. Where $A1$ represents the data arrival rate, $A2$ represents the performance of the data source, $A3$ represents the data security protection capability, $A4$ represents the authentication of the data source and $A5$ represents the historical evaluation of the data source.

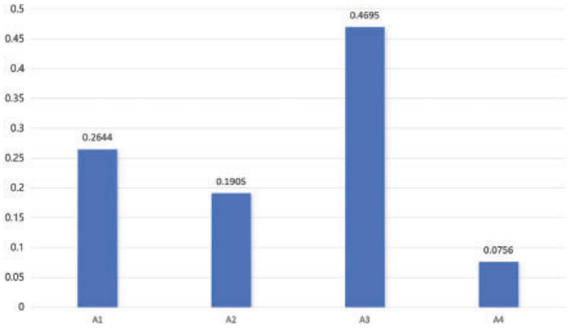

Figure 5: Four attribute weights

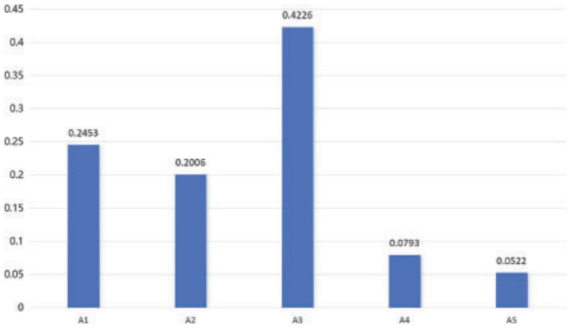

Figure 6: Five attribute weights

Users generate subjective indicators of the trustworthiness of a data source based on various attributes of the data source, as shown in Fig. 7.

Construct the basic probability function of each data source, modify the basic probability function of the data source by combining the attribute weight values through Eqs. (4)–(9) and calculate its direct trust value as shown in Fig. 8. Where $2^{\Omega} = \{\varnothing, \{T\}, \{-T\}, \{T, -T\}\}$, $\{T\}$ stands for trustworthy, $\{-T\}$ stands for untrustworthy, $\{T, -T\}$ stands for uncertainty, and T_d stands for the direct trust value of the data source.

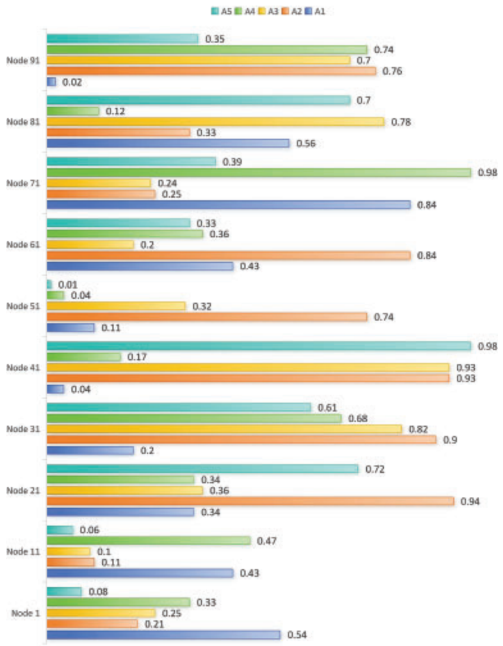

Figure 7: Attribute-based credibility of data sources

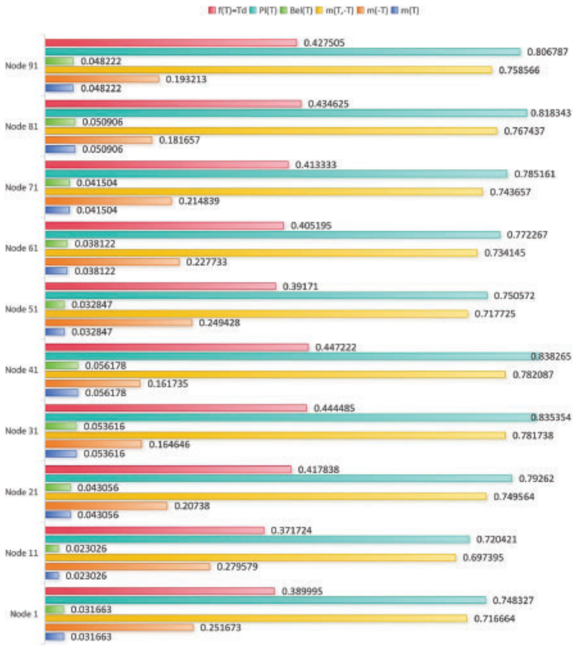

Figure 8: Data source direct trust value

5.2 Indirect Trust Value Calculation

The big data network is simulated with Netlogo, and the PR value of each neighbouring node in the big data source network, i.e., the indirect trust value between a node and its neighbouring nodes, is calculated by the PageRank algorithm and Eqs. (10)–(11), and the PR values between some nodes are shown in Table 1.

Table 1: Data source indirect trust value

Start	PR	End
1	0.095	31
11	0.052	71
21	0.047	61
31	0.039	91
41	0.028	51
51	0.025	21
61	0.023	81
71	0.021	1

(Continued)

Table 1 (continued)

Start	PR	End
81	0.002	11
91	0.049	41

5.3 *Comprehensive Trust Value Calculation*

Through Fig. 8, Table 1 and Eq. (12), the combined trust value of the data source node is calculated, and the combined trust value of some user nodes to the data source node is listed as shown in Table 2, and T_c represents the combined trust value.

Table 2: Data source comprehensive trust value

Start	T_c	End
1	0.058899	91
11	0.021493	71
21	0.029784	11
31	0.016673	91
41	0.051198	11
51	0.029490	61
61	0.009996	81
71	0.050416	31
81	0.000743	11
91	0.032882	51

From the experimental results, it is clear that the nodes with higher values of $m(T)$ and PR have higher combined trust value, i.e., higher trustworthiness and the nodes with lower values of $m(T)$ and PR have lower combined trust value, i.e., lower trustworthiness.

5.4 *Comparative Experimental Analysis*

For clarity, the term "A" refers to the traditional D-S evidence theory algorithm, and "B" represents the data source reliability quantification algorithm proposed in this paper. Since A can only obtain the direct trust value of a data source node, we conducted two sets of comparison experiments.

Experiment 1: In this experiment, the direct trust values were obtained using both A and B, then compared and analyzed, as shown in Fig. 9. From the figure, we observe that for trusted data source nodes 31, 41, and 81, B's direct trust value is significantly higher than A's. For other untrusted nodes, B's indirect trust value is considerably lower than A's. This experiment demonstrates that the data source reliability quantification algorithm proposed in this paper is more effective at distinguishing whether the data source nodes are trustworthy.

Experiment 2: The goal of this experiment is to obtain the comprehensive trust value of each node during data asset transactions. Since A only provides the direct trust value, we combined A with the indirect trust value calculated using the PageRank algorithm proposed in this paper to compute the comprehensive trust value for A. This was then compared with the comprehensive trust value obtained from B, as shown in Fig. 10. As shown in the figure, even with the same indirect trust value, for trusted nodes, B's comprehensive trust value is much higher than A's. Conversely, for untrusted nodes, B's comprehensive trust value is significantly lower than A's. This highlights B's superior ability to distinguish whether a data source is trustworthy. Furthermore, by comparing Figs. 9 and 10, it is evident that for the same node, the ratio of B to A in Fig. 10 is significantly larger than the ratio in Fig. 9. This difference in ratios emphasizes the more pronounced effect of the comprehensive trust value comparison, which is attributable to the characteristic that "trust is hard to earn but easy to lose." This set of experiments further confirms that the data source reliability quantification algorithm proposed in this paper has a more significant ability to assess the comprehensive trust value of data source nodes, thereby validating the superiority of the proposed algorithm.

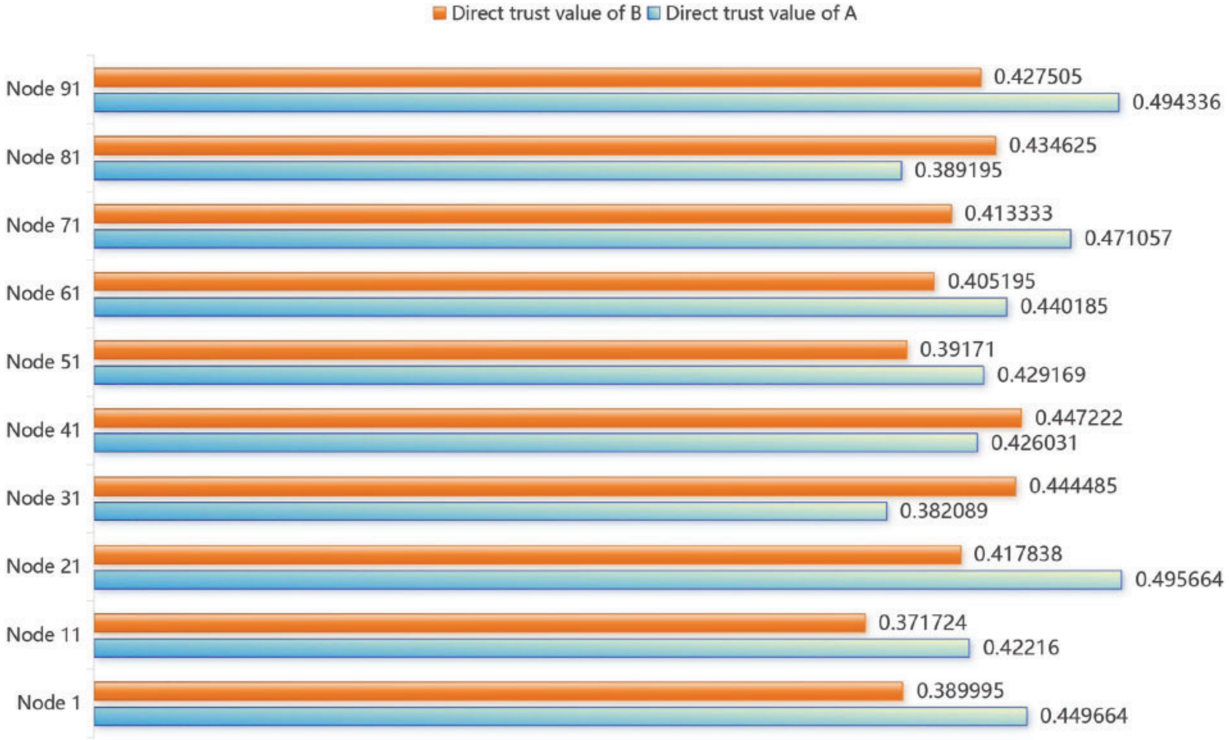

Figure 9: Comparison of direct trust values for data sources

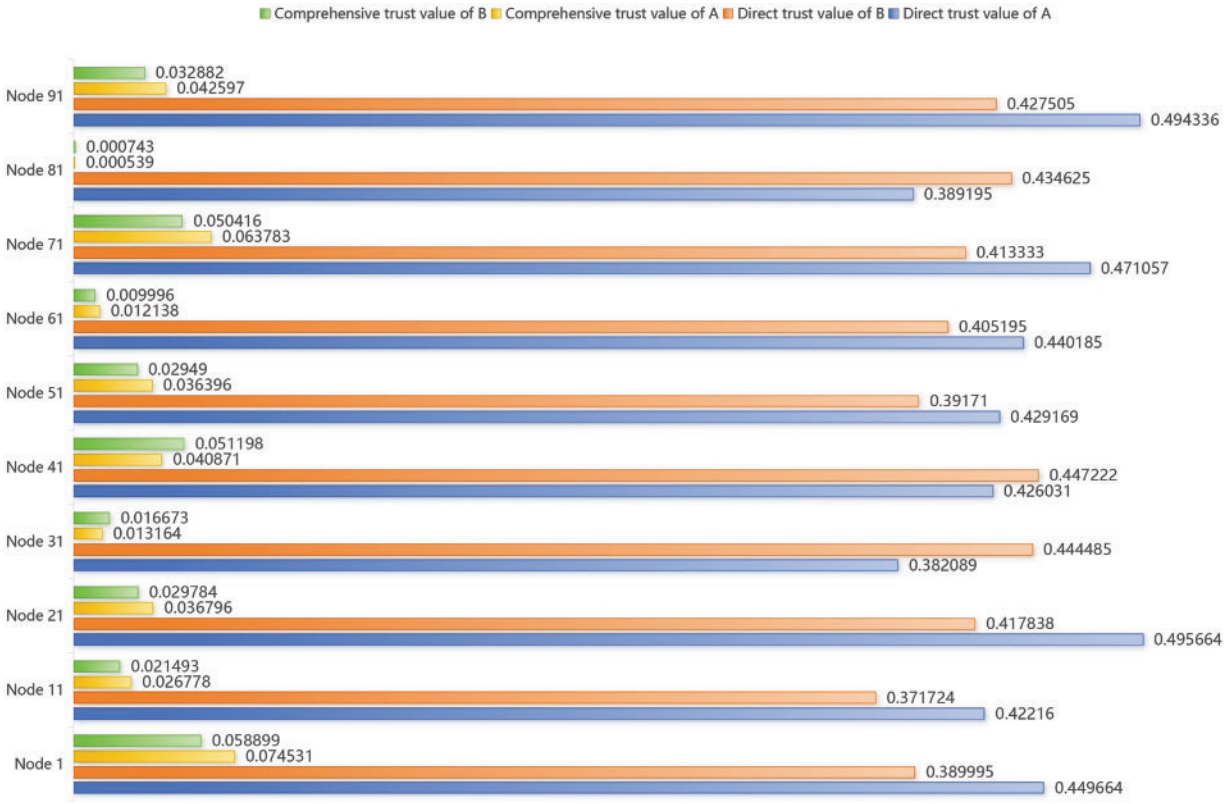

Figure 10: Comparison of comprehensive trust values for data sources

6 Conclusion

This paper proposes a data source reliability quantification mechanism for a big data environment, which includes a distributed data source trust assessment management framework, a data source reliability quantification model, and a dynamic adjustment mechanism for data source reliability. In the distributed data source trust assessment management framework, the nodes collectively form a trust network. These nodes gather trust data from data sources via a trust assessment agent, perform trust evaluations on the data sources, and store the assessment results in a trust database to respond to requests for data source reliability. By leveraging the collaborative efforts of multiple nodes in the distributed network, a comprehensive, dynamic, and objective trust assessment of the data sources is achieved. The data source reliability quantification model uses the AHP hierarchical analysis method to assign weights to the attributes of the data source. After applying these attribute weights to traditional D-S evidence theory, it generates the modified D-S evidence theory, calculates the direct trust value of the data source, determines the indirect trust value between the nodes in the trust network using the PageRank algorithm, and ultimately computes the comprehensive trust value of the data source based on both the direct and indirect trust values. The dynamic adjustment mechanism recalculates and adjusts the trustworthiness of the data source by checking if new historical evaluations have been generated compared to the previous request. This ensures that the trust assessment process is continuous and dynamic, maintaining real-time accuracy in the results. Experimental results show that, compared to traditional D-S evidence theory, our proposed data source reliability quantification mechanism not only reduces subjective interference, but also more objectively and accurately calculates the direct trust value of the data source node, as well as the indirect and comprehensive trust values.

Additionally, comparative tests demonstrate that the mechanism presented in this paper is significantly better at distinguishing whether a data source is trustworthy, with stronger robustness.

In practical applications of data asset trading platforms, the data source credibility quantification model may face various security threats and privacy concerns. First, malicious nodes may manipulate trust evaluations by falsifying data, boosting their own trust values, or diminishing the credibility of others through collusion attacks, which could undermine the fairness of data transactions. To address this, this study introduces a multi-condition D-S combination rule based on D-S evidence theory to enhance tolerance for conflicting data, and employs the PageRank algorithm to minimize the influence of individual nodes on the overall trust evaluation. Second, the data involved in the transaction process may contain sensitive information, which presents risks of leakage or unauthorized access. To mitigate this, differential privacy technology or secure multi-party computation methods can be applied to protect data source privacy when calculating trust values. Moreover, given the dynamic nature of the data asset trading environment, a time decay mechanism can be implemented to dynamically adjust trust values and prevent outdated data sources from affecting the accuracy of system decisions. These measures not only improve the accuracy of data source credibility evaluations but also strengthen the security and privacy protection features of the model.

In future work, we will apply the algorithm and model proposed in this paper to large-scale environments, such as data asset trading platforms, and explore their scalability and applicability. Specifically, we will investigate whether there are limits to the number of nodes and critical thresholds in real-world big data networks with vast numbers of nodes, and continue to explore other types of data or trust environments, such as those in IoT systems.

Acknowledgement: The authors would like to acknowledge all other participants.

Funding Statement: This work was funded by Haikou Science and Technology Plan Project (2022-007), in part by key Laboratory of PK System Technologies Research of Hainan, China.

Author Contributions: The authors acknowledge the following contributions to this paper: study conception and design: Gaoshang Lu; analysis and comparison of experimental results: Gaoshang Lu; experimental guidance and interpretability: Fa Fu; manuscript writing: Gaoshang Lu; manuscript checking: Fa Fu, Zixiang Tang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Sanger J, Richthammer C, Hassan S, Pernul G. Trust and big data: a roadmap for research. In: 2014 25th International Workshop on Database and Expert Systems Applications; 2014 Sep 1–5; Munich, Germany: IEEE. p. 278–82.
2. Xue Y, Lai Y. Convergence of big energy thinking and big data thinking (I) big data and power big data. Power System Automation. 2016;40(1):1–8.
3. Poonam KG, Misra M. Opinion based trust evaluation model in MANETs. In: Contemporary Computing: 4th International Conference, IC3 2011; 2011 Aug 8–10; Noida, India: Springer. p. 301–12.
4. Hosmani S, Mathapati B. R^{2}SCDT: robust and reliable secure clustering and data transmission in vehicular ad hoc network using weight evaluation. J Ambient Intell Humaniz Comput. 2023;14(3):2029–46. doi:10.1007/s12652-021-03414-3.

5. Xu Z. Research on software trustworthiness measurement evaluation model based on data driven. In: MATEC Web of Conferences; 2021 Feb 15; EDP Sciences; 2021. [cited 2025 Jan 30]. Available from: https://www.matec-conferences.org/articles/matecconf/abs/2021/05/matecconf_cscns20_08014/matecconf_cscns20_08014.html.
6. Xu L, Zheng X, Rong C. Trust evaluation based content filtering in social interactive data. In: 2013 International Conference on Cloud Computing and Big Data; 2013 Dec 16–19; Fuzhou, China: IEEE. p. 538–42.
7. Ansheng Y, Haiping H. Trusted network evaluation model based on comprehensive trust. Chin J Electron. 2021;30(6):1178–88. doi:10.1049/cje.2021.07.028.
8. Mythili V, Suresh A, Devasagayam MM, Dhanasekaran R. SEAT-DSR: spatial and energy aware trusted dynamic distance source routing algorithm for secure data communications in wireless sensor networks. Cogn Syst Res. 2019;58:143–55. doi:10.1016/j.cogsys.2019.02.005.
9. Rajasekaran M, Ayyasamy A, Jebakumar R. Performance and evaluation of location energy aware trusted distance source routing protocol for secure routing in WSNs. Indian J Sci Technol. 2020;13(39):4092–108. doi:10.17485/IJST/v13i39.1522.
10. Lv D, Zhu S. Achieving secure big data collection based on trust evaluation and true data discovery. Comput Secur. 2020;96(2):101937. doi:10.1016/j.cose.2020.101937.
11. Lopez J, Maag S, Morales G. Behavior evaluation for trust management based on formal distributed network monitoring. World Wide Web. 2016;19(1):21–39. doi:10.1007/s11280-015-0324-6.
12. Jayasinghe U, Otebolaku A, Um TW, Lee GM. Data centric trust evaluation and prediction framework for IoT. In: 2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K); 2017 Nov 27–29; Nanjing, China: IEEE. p. 1–7.
13. Li T, Huang G, Zhang S, Zeng Z. NTSC: a novel trust-based service computing scheme in social internet of things. Peer Peer Netw Appl. 2021;14(6):3431–51. doi:10.1007/s12083-021-01200-8.
14. Lin H, Garg S, Hu J, Wang X, Piran MJ, Hossain MS. Data fusion and transfer learning empowered granular trust evaluation for internet of things. Inf Fusion. 2022;78(9):149–57. doi:10.1016/j.inffus.2021.09.001.
15. Aldini A. A formal framework for modeling trust and reputation in collective adaptive systems. arXiv:1607.02232. 2016.
16. Feng R, Xu X, Zhou X, Wan J. A trust evaluation algorithm for wireless sensor networks based on node behaviors and ds evidence theory. Sensors. 2011;11(2):1345–60. doi:10.3390/s110201345.
17. Belov N, Schlachter J, Buntain C, Golbeck J. Computational trust assessment of open media data. In: 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW); 2013 Jul 15–19; San Jose, CA, USA: IEEE. p. 1–6.
18. Gao Y, Li X, Li J, Gao Y, Philip SY. Info-trust: a multicriteria and adaptive trustworthiness calculation mechanism for information sources. IEEE Access. 2019;7:13999–4012. doi:10.1109/ACCESS.2019.2893657.