Discover Computing

RESEARCH Open Access

Semantic data sharing and pricing in web 3.0 using blockchain

V. Sitharamulu[1], G. Sucharitha[2], Srihari Babu Gole[3], Hanumantha Rao Battu[4], Onur Osman[5*] and Jawad Rasheed[6,7,8,9]

*Correspondence:
Onur Osman
onurosman@topkapi.edu.tr

Full list of author information is available at the end of the article

Abstract
Web 3.0 represents an advanced internet framework that combines technologies such as blockchain and semantic communication to enable decentralized, transparent data exchange. While semantic communication reduces redundancy by transmitting contextual meaning rather than raw data and existing systems lack a cohesive integration strategy, further, the centralized models struggle with efficient semantic sharing and valuation. The work proposed integration framework of semantic web communication with blockchain to enhance reliable web 3.0 data sharing. A novel "Proof of Semantic" mechanism tackles data integrity issues, ensuring only high-quality semantic inputs. The approach also elucidates the implementation of the bottleneck method and state channels. The computational load is optimized using a semantic shared protocol across the web. Further, a hierarchical Stackelberg game model optimizes semantic pricing, balancing stakeholder incentives. Evaluations confirm the framework's efficacy, achieving reduction in network load without compromising accuracy compared to conventional approaches.

Keywords Web 3.0, Blockchain integration, Semantic communication, Decentralized data sharing, Incentive mechanisms

1 Introduction

Web 3.0 has evolved from its early focus on machine-readable semantics to a decentralized paradigm underpinned by blockchain, enabling user-controlled data management. Unlike Web 2.0's reliance on centralized cloud infrastructure, Web 3.0 leverages distributed ledgers and semantic technologies to foster trustless interactions. The work also deals with web3.0, working with decentralized blockchain systems to enhance reliable data sharing over complementary layers. Semantic communication enhances efficiency by transmitting task-specific meaning, allowing edge devices to reconstruct data using shared knowledge. This reduces latency and bandwidth demands, critical for intelligent networks. However, challenges persist in verifying semantic accuracy, securing decentralized knowledge updates, and fairly valuing user-generated content.

Blockchain's immutability and transparency address trust gaps in Web 3.0, ensuring tamper-proof data exchange across sectors like healthcare and IoT. Prior studies [1, 2] highlight its role in securing data ownership, while semantic communication optimizes

resource usage [3, 4]. Yet, merging these technologies remains underexplored, the methodology addresses the following core challenges of semantic sharing.

1. *Pre-Sharing Verification* Preventing unreliable semantic data from entering the system (garbage-in, garbage-out).
2. *Efficient Sharing* Extracting contextually relevant semantics while minimizing redundancy.
3. *Post-Sharing Valuation* Establishing equitable pricing to sustain stakeholder participation.

The approach integrates frameworks, sharing mechanisms, and Stackelberg models. Further, the methodology is entailed clearly below.

- A *blockchain-semantic integration framework* – an important framework that define the "proof of semantic" that works on the validation of data, data integrity and quality.
- A *state channel-driven sharing mechanism* using information bottleneck theory to prioritize task-critical semantics, reducing on-chain operations.
- Nash Algorithm is aptly used for producer-consumer incentives during the mechanism. A Multi-tier Stackelberg pricing model is deployed along with Nash Algorithm for balancing producer-consumer alternatives.
- Empirical validation is procured with 32.8% lower overhead compared to traditional methods, with competitive accuracy on benchmarks like FMNIST.

The paper is modelled as various sub sections; the Sect. 2 describes the reviews and related work. The Sects. 3–5 on the other side describes the experiments conducted and derived, Further the Sect. 6 concludes the complete approach. Therefore, the complete approach handles various challenges and integration framework using state channel -driven sharing mechanism and multi-tier Stackelberg pricing model. At the end it is observed that the challenges are handled with lowering overhead with FMNIST.

2 Related work

It is observed that there is a limited research scope with the integration part of semantic web with blockchain technology. Hence, the presentation also highlights the structure of the paper in three broad categories semantic web pricing, the semantic sharing and the web 3.0.

2.1 Web 3.0 innovation via blockchain and intelligent communication

Modern Web 3.0 is an emerging technology, similar to Web 2.0, but with enhanced features inspired and influenced by social media and cloud platforms. Several scientists examined and studied the fundamental approaches of semantic and blockchain to showcase their flexibility in the current scope of technical assets [5]. Keeping in view of the circumstances, the communication overhead is a typical factor in IoT networks. This lead to data integrity issues over semantic web [6]. has introduced few schemes for block chain which can be combined with semantic coding to increase integrity of data. Similarly, several others [7] also focused on designing a semantic differential transaction model to reduce redundant data exchange procedures [8]. entailed the factors of the integration of blockchain with web 3.0 and their decentralization techniques. This

reinforced the emerge of the block chain with semantic web. The [9] introduced Hyper Service, a cross-chain interoperability platform enabling seamless connections between several blockchains for Web 3.0 applications [10]. developed a consistency protocol utilizing Trusted Execution Environments (TEEs) to extend trust from blockchain networks to off-chain data sources [11]. developed a dual-layer scaling model that incorporates sharding and payment channels to facilitate large-scale wireless network transactions by integrating on-chain and off-chain components. Despite these advancements, most studies do not adequately address data authenticity on the blockchain, which is a crucial factor in maintaining system trustworthiness.

2.2 Semantic sharing in blockchain-enabled networks

While semantic communication aims to enhance the accuracy of information shared in Web 3.0, it often results in unnecessary data transmission and inefficient task execution [12]. introduced a federated learning-based auto encoder that applies wave-to-vector technology to enhance audio transmission over wireless networks [13]. developed a data adaptation framework for image transmission, ensuring data is transformed into a more relevant format. The other researchers during their servery found and introduced [14] a task-oriented multi-user semantic communication system, while [15] implemented a deep learning-powered model for optimizing speech transmission. Additionally [16], generated task-based learning frameworks to increase and improve mutual information between block chain and semantic sharing for edge inference, later extending it to a cooperative multi-device edge inference system aimed at reducing redundant features [17]. Despite these efforts, current methodologies still struggle with effectively extracting relevant semantic information from communication processes, leading to excessive data processing and increased energy utilization.

2.3 Mechanisms for semantic-based pricing

Multiple studies validate semantic pricing as a basis for enabling semantic information flow. Notably [18], introduced an attention-proximal policy optimization framework to assess semantic data importance [19]. designed a two-stage stochastic approach to optimize resource allocation for semantic communications [20]. proposed a resource distribution framework that prioritizes semantic spectral efficiency through effective channel allocation and symbol management, later, it is further incorporated with power management and Quality-of-Experience (QoE) optimization [21]- [22], developed a two-tier Stackelberg game-based model to enhance resource-sharing among mobile IoT devices [23]. introduced a QoE-driven, attention-based resource management framework to facilitate semantic communication within the metaverse ecosystem [24]. However, existing pricing models do not fully account for blockchain transaction costs and the computational overhead of semantic extraction, making them less suitable for a blockchain-integrated Web 3.0 environment. Therefore, integration mechanism is necessary to overcome the challenges of transaction costs and computational overhead.

3 Detailed challenges and gaps in the integration paradigm

Since blockchain lacks the capability to independently verify external semantic information, addressing the "garbage-in, garbage-out" problem remains a significant challenge [25]. To overcome this issue, we have combined the block chain and semantic web

communication with a specific web framework. The Figure 1 demonstrates the concepts of several semantic mechanisms and their threshold signatures. Original images are transmitted over a resource constrained wireless networks using block chain semantic web 3.0. Further, the semantic web framework is integrated with four key components. The key components include the wireless channel, semantic decoder and encoder, and proof of semantic."

Semantic Encoder Semantic Encoder extracts meaningful information from raw pictures. Edge devices generate datasets of pictures, denoted as $r = [r_1, r_2, r_3 \ldots .r_s]$ where $\mathbf{r}_s$ represents the **S**-th sample in a total of **S** inputs. The encoder processes these images using a neural network to obtain the corresponding semantic representations $z = [z_1, z_2, z_3 \ldots .z_s]$. The extraction process is parameterized by $\boldsymbol{\theta}_e$, leading to an output semantic representation $x = [x_1, x_2, x_3 \ldots .x_N]$, with all N number of elements and are corelated to the input sampling images. This transformation is mathematically expressed as $x = T\theta_e(r)$, where $T\theta_e(\cdot)$ Specifies the semantic encoder's multifunctional design and processing methodology.

Wireless Channel The parsed **s**emantic data is navigated through a dynamic wireless channel that operates under resource limitations and is susceptible to noise interference. The received semantic data, **y**, is influenced by the characteristics of the wireless channel and can be modelled as : $y = c.x + \sigma$, $where\ y$ denotes the received semantic data, which is affected by noise, **c** is the channel coefficient, $\boldsymbol{\sigma} \sim \mathcal{N}\left(0, \sigma^2 \mathbf{I}\right)$ denotes independent and identically distributed (i.i.d.) zero-mean Gaussian noise, σ^2 is the noise variance, and I represents the identity matrix. To quantify the efficiency of semantic transmission, the compression ratio (CR) is introduced, which measures the relationship between image resolution and raw data. It is defined as: CR = log(y/r). All the legacy systems existing so far rely on centralized entities for data exchange, making information transmission centralized, less controllable, and vulnerable to security risks. To ensure decentralization, transparency, and security, the generated semantic information should

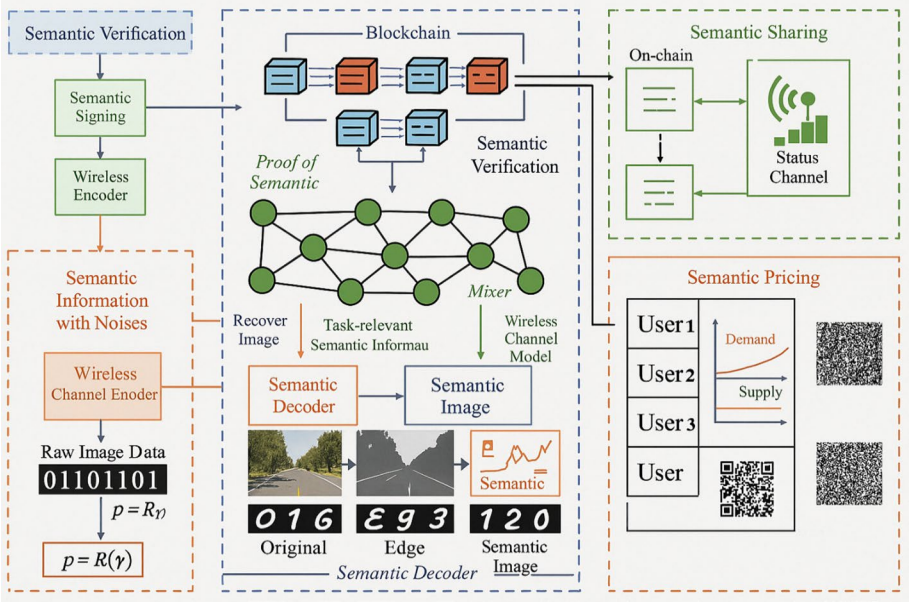

Fig. 1 Unified architecture combining blockchain and semantic communication

be recorded on the blockchain, enabling a more reliable and trustless communication framework.

3.1 Proof of semantic

Since blockchain cannot inherently validate external semantic data, an additional verification mechanism is necessary to ensure the integrity of information before it is recorded on blockchain. A typical mechanism termed as proof of semantic with threshold signature has been introduced to validate the verification mechanism, inspired by [26]. This approach enables secure verification of semantic data through decentralized validation and the protection of semantic attacks over the block chain. The proof of semantic maintains a key logging feature to list the semantic miners for its unauthorized usage.

All the existing threshold signature consensus mechanisms are less secured and unreliable over semantic web communication. The proposed PoS works balanced with semantic validation over the consensus mechanism featuring high reliability of data and less risk of unmeaningful information during enchain process. The method has following sequential steps.

Step 1: System Initialization A bilinear pairing function is a part of semantic mechanism used to establish secure mappings between different cyclic groups: e: G1 × G2 → GT where G_1 and G_2 are two multiplicative cyclic groups, and G_T represents a third group for the occurrence of mappings. Additionally, H: M → G_1 is deployed to transform semantic data M into unique representation within G_1, ensuring efficient verification.

Step 2: Key Generation A cryptographic key pair $\left(s_{i,}g^{si} \in G_2\right)$ is created by individual miners for the threshold signature process: **si** serves as the private key for signing semantic data, $\mathbf{g}^{si}$ acts as the corresponding public key for verification. Miners, including edge devices with sufficient computational capabilities validate the process of the semantic information and authenticate the complete process for merging with the blockchain. now semantic information is merged to blockchain.

Step 3: Semantic Signing Miners process y through a neural network-based model with parameters θ_b, generating an inference result z'. The result z' is obtained when the semantic data y is mapped to their neighbouring functions. To optimize data transmission and reduce overhead, miners compute a hashed identifier: h_i=H (y, z′) Each miner then signs the semantic information by computing: $\delta_i = (y, h_i)^{si}$ The signed message **δi** is then forwarded to a designated miner responsible for consensus.

Step 4: Semantic Verification To ensure the authenticity of the signatures, the selected miner verifies each **δi** using the bilinear pairing function: $e(\delta_i, g) = e((y, h_i), g^{si})$ which holds due to bilinear properties: $e((y, h_i)^{si}, g = e\left((y, h_i), g^{si}\right) = e((y, h_i), g)^{si}$ For the verification process to be considered valid, at least t-out-of-n miners must provide the valid signatures confirming the consistency of the semantic data y and its output label z. The consensus condition, adapted from [27], is expressed as: $\sum_{i=1}^{t} e((y, h_i), g^{si}) \geq h$ where: t represents the required minimum number of miner verifications, n is the total number of participating miners, and h denotes the available semantic information.

The threshold t, on the other hand can be finely tuned dynamically for the security requirements. The higher thresholds provide stronger security assurances but may

increase computational and communication costs. The selection of t can be guided by reputation-based models or other consensus mechanisms [27].

If the aggregated signature successfully passes verification, it confirms that a sufficient number of miners have validated the semantic data, ensuring its integrity, reliability, and resistance to tampering. This approach mitigates the garbage-in, garbage-out problem, improving trust in blockchain-based semantic communication.

In order to provide clarity and mathematical consistency throughout the proposed framework, Table 1 summarizes the general symbols and variables used within the formulation. This table provides a concise reference to notations being used in Sections, related to the Proof of Semantic, Information Bottleneck and Stackelberg game formulations.

3.2 Semantic decoder

The semantic decoder reconstructs images from received semantic data **y** using a neural network (θ_d). It generates a reconstructed image $\widehat{r}$ from the original **r**, expressed as $\widehat{r} = R_{\theta_d}(y)$ The goal is to minimize errors while maintaining decentralization. The loss function plays vital role in merging the MSE (Mean Square Error0 and the (CE) Cross Entropy: $L = \lambda L_{CE} + (1-\lambda)^2 L_{MSE}$ where λ = 1 - CR. Parameters $\theta = (\theta_e, \theta_d)$ are updated using Stochastic Gradient Descent (SGD)= $\theta^{(i+1)} = \theta^{(i)} - \eta \frac{1}{S} \sum (r_s - \hat{} r_s)^2$ Image recovery quality is matriculated by Peak Signal to Noise Ratio and is given by way of:

$$PSNR = 10\log_{10} \frac{\max(\widehat{r})^2}{MSE}$$

This proof of semantic approach ensures efficient and accurate image reconstruction and availsdata integrity through a secure mechanism.

4 State channel & task-based information bottleneck semantic sharing (SC & TB Ib)

All the existing errors in the Decentralized systems are handled using the proof of semantic mechanism. Despite, there are fewer limitations pertaining to cloud and storage. Miners may skip verification because they do not want to store shared knowledge. As a solution to this, (SC & TB Ib) Information Bottleneck methodology is elevated to

Table 1 Key mathematical notations

Notation	Definition	Notation	Definition
r	Input of the semantic encoder	z	Mapping pragmatic output of r
y	Noised semantic information	CR	Compression ratio
s_i, g^{si}	Key pairs for miner i	h_i	Hashed identities of semantic information
$\widehat{\mathbf{Z}}$	Mapping pragmatic output of $\widehat{r}$	$R_{\theta_d}(y)$	Semantic decoder function
$D_{KL}(.,.)$	Kullback–Leibler divergence	*N*	Producers set
p_n	Pricing strategy of producer n	e	Bilinear pairing function
N	Producers set	δ_i	Signed message
U_m	Utility of consumers	L (·)	Loss function
x	Semantic information	*M*	Consumers set
R_n	Utility of each producer	H	Hash function
H	Hash function	r̂	Distorted version of r
$I(.,.)$	Mutual information	x_m^n	Semantic demand of consumer m for producer n

optimize semantic sharing while reducing computational overhead. Some of the key Challenges include:

- How can miners verify semantic information efficiently while reducing on-chain overhead?
- What techniques enable semantic encoders to isolate task-relevant details while cutting down on redundancy?

A. Overhead Reduction semantic extraction

The Markov process includes the encoding and decoding of the semantic web dynamics. It also relates the relationship of encoding and decoding tasks with overhead ratings.

$\boldsymbol{Z} \rightarrow \boldsymbol{R} \rightarrow \boldsymbol{X} \rightarrow \boldsymbol{Y} \rightarrow (\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{Z}})$ are variables to represent different stages of information transformation. The probability function is denoted as: $p(\widehat{r} \mid r) = p_{\theta_d}(\widehat{r} \mid y) p_{\text{channel}}(y \mid x) p_{\theta_e}(x \mid r)$

To maximize task-relevant semantic information while minimizing encoding complexity, the Information Bottleneck (IB) principle is applied, balancing the trade-off between preserving meaningful information and reducing redundancy. The objective function is formulated as: $L_{IB} = I(Y,Z) - \beta\, I(Y,R)$, where I (Y, Z) represents the mutual information between the channel output Y and the extracted semantic features Z, while I(Y, R) quantifies the encoding complexity between Yand the original source data R The parameter β serves as a trade-off factor, ensuring that only the most relevant semantic information is retained while minimizing unnecessary overhead. During the implementation procedures, β is considered to be the empiral selection used for leveraging semantic compression and relevance. Further, strategic implications can also be used for future scope of work.

The lower bound of *l (Y, Z)* is derived using the Kullback-Leibler (KL) divergence and approximated with a variational distribution *q (X|Y)*. $I(Y,Z) \geq \int p(r)\, p(y \mid r) p(z \mid r) \log q(z \mid y) dr dy dz$. This ensures effective encoding of semantic data by preserving essential information. Similarly, the upper bound of *I (Y, R)* is approximated using another variational distribution function *r(y)*, $I(Y,R) \leq \int p(y,r) \log p(y \mid r) dy dr - \int p(y,r) \log r(y)\, dy dr$ representing the decoder. This approach is helpful to minimize unnecessary complexity in encoding and maintaining an optimal tradeoff balance between information retention and computational efficacy.

By combining these bounds, the final empirical estimation of L_{IB} can be computed as:

$$L_{IB} = \frac{1}{N} \sum_{n=1}^{N} \left[\frac{1}{K} \sum_{K=1}^{K} \log q\,(z_n \mid y_{n,k}) - \beta D_{KL}\,(p_{\text{channel}}\,(y|x)\; p_{\theta_e}\,(x|r_n) \;||\, r\,(y)) \right]$$

The Monte Carlo sampling is further used to approximate expectations. Reparameterization tricks are applied for efficient training. Channel noise is modelled as $y_{n,k} = c x_n + \sigma_{n,k}, \;\; \sigma \sim N(0, \sigma^2 I)$ and ccc is accounted as the channel coefficient. This approach improves the semantic communication efficiency by Extracting task-relevant information using the IB principle.

B. Off-chain semantic verification for redundancy.

The alternate mechanism to detect the semantic information on the blockchain is to make state channels move this process off-chain to reduce costs and data load. Semantic

information is shared among edge devices off-chain, and only after multiple updates, the accumulated data is submitted to the blockchain for consensus. This approach minimizes on-chain transactions and fees. A state channel is properly managed and organized by a smart contract, which requires edge devices to lock tokens before opening a channel. This prevents fraudulent submissions. Each channel facilitates semantic data sharing between devices and is represented as $C = \{d, s_o s_p s_c\}$, where: d = List of participating edge devices, s_o= Initial blockchain state, s_p = Off-chain state proofs from shared semantic data, s_c = On-chain state proved by miners.

Instead of sending every piece of semantic data to the blockchain immediately, devices initially share and validate it to off-chain multiple times. They then submit a compressed proof (sp) containing accumulated states, sequence numbers (to prevent replay attacks), Merkle proofs, and digital signatures of data producers and consumers. For example, a producer extracts semantic data from images and transmits it over a wireless network. The consumer verifies and signs the received data. This exchange continues until the party decides to update the blockchain, ensuring efficient and decentralized semantic verification. The smart contract then verifies the proof and updates the blockchain state accordingly.

5 Stackelberg game-based semantic pricing: a multi-leader multi-follower framework

The loss function on the other hand is used to calculate the already extracted semantic information from the web. It also measures the total amount of mutual information gathered. This allows semantic data to be securely and efficiently traded via blockchain. For example, in Web 3.0 urban planning, large-scale image data is needed to build digital world. Instead of transferring entire images, producers extract essential features (e.g., contours) and share them with consumers, reducing energy use and data redundancy. Optimizing revenue and efficiency is very essential and is successfully achieved by modelling the interaction between semantic consumers and producers based on a Stackelberg model with multiple leaders and followers. This game-theoretic approach helps establish fair pricing for semantic data while ensuring secure and efficient sharing.

A. Problem formulation: semantic pricing using stackelberg game.

In a blockchain-based semantic sharing network, all the clients associated to semantic encoders and decoders are considered as producers or consumers. The semantic network consists of N producers and M consumers, where each consumer *m* m requests semantic data from a producer *n* n at a price *pn*. The semantic demand represents the amount of transmitted data, and pricing strategies for producers and consumers are denoted as p = [p_1, …, p_n] → Prices set by producers, X = [x_1, …, x_m] → Amount of semantic data purchased by consumers. Each consumer has a budget *Bm*, and each producer has a data supply limit *Sn*. Since different producers contribute varying quality of semantic data, a consumer's utility depends on the quality of data received. If a producer provides higher-quality semantic data, consumers benefit from improved accuracy. The utility function follows a logarithmic model, which reflects diminishing returns as more data is acquired. The total utility for a consumer is represented as:

$$U_m = \mu_m \sum_{n=1}^{N} \log\left(1 + \log\left(1 + \eta_m x_m^n\right)\right) - \sum_{n=1}^{N} \left(p_n + k_m c_m f_m^2 + b_m\right) x_m^n$$

where: η_m → Producer's contribution to data quality, $k_m c_m f_m$→ Computational parameters affecting recovery costs, b_m → Fixed blockchain (on-chain) costs.

Similarly, the utility of a producer consists of revenue from selling data to the costs of semantic extraction and blockchain transactions:

$$R_n = \sum_{m=1}^{M} \left[p_n - k_n c_n f_n^2 - b_n\right] x_m^n$$

Where $k_n c_n f_n$→ Costs associated with extracting semantic data. Since both producers and consumers aim to maximize their respective utilities, we applied a multi-level Stackelberg game to describe the relationship between producers and consumers, featuring multiple leaders and followers. The Stackelberg game consists of two categories and are defined as under.

Category 1: Pricing Category (Leaders and Producers) Producers act as leaders and set prices p = [p_1, …, p_n] to maximize their utility R_n. The optimization problem for each producer n is: $\max_{p_n} R_n(p_n, p_{-n}, X)$ subject to: $p_n \geq k_n c_n f_n^2 + b_n, p_n \leq p_{max} = \varphi_n e^{-S_n/max(S_1,\ldots,S_N)}$, where p_{max} is constrained by available data supply S_n, thus, ensuring price fluctuations that reflect market conditions immediately. This pricing stage has a Nash equilibrium, that facilitates the best pricing of producer's decision compared over other prices.

Category 2: Semantic Demand Category (Followers and Consumers).

Here, Followers decide how much of semantic data need to be purchased: X = [x_1, …, x_m] based on the prices set by producers. The optimization problem for each consumer m is: $\max_{x_m^n} U_m(x_m^n, X_{-m}, p)$, subject to: $x_m^n \geq 0, \sum_{n=1}^{N} \left(p_n + k_m c_m f_m^2 + b_m\right) x_m^n \leq B_m, \sum_{n=1}^{N} x_m^n \leq S_n$ where X_{-m} represents the demands of all other consumers. Unique Nash equilibrium needs to be managed to definitely enhance stability in the semantic demand state. Rosen's concave game framework introduced in the year 1965 [28], is very unique for concavity and development of equilibrium. The Stackelberg game provides an optimal pricing and purchasing strategy for both semantic producers and consumers. Producers set competitive prices, and consumers allocate budgets efficiently. Since the utility functions are concave, the system naturally converges to an equilibrium where both parties maximize their benefits.

B. Problem Solution/ Follower Solution.

The existing problem as defined can be addressed using the Karush-Kuhn-Tucker (KKT) conditions. Let λ_m^1, λ_m^2, and λ_m^3, be the Lagrangian multipliers. The associated Lagrangian function becomes:

$$\begin{aligned} L_m(x_m^n) = {} & U_m(x_m^n) \\ & + \lambda_m^1 x_m^n + \lambda_m^2 \left[B_m - \sum_{n=1}^{N} \left(p_n + k_m c_m f_m^2 + b_m\right)\right] \\ & + \lambda_m^3 \left(S_n - \sum_{n=1}^{N} x_m^n\right). \end{aligned}$$

The primal feasibility conditions associated are:

$$x_m^n \geq 0,\ B_m - \sum_{n=1}^{N} \left(p_n + k_m c_m f_m^2 + b_m\right) x_m^n \geq 0,\ S_n - \sum_{m=1}^{M} x_m^n \geq 0.$$

The ideal and available dual feasibility conditions: $\lambda_m^1, \lambda_m^2, \lambda_m^3 \geq 0$

The prevailing laxity conditions:

$$\lambda_m^1 x_m^n = 0,\ \lambda_m^2 \left[B_m - \sum_{n=1}^{N} \left(p_n + k_m c_m f_m^2 + b_m\right) x_m^n\right] = 0,\ \lambda_m^3 \left(S_n - \sum_{m=1}^{M} x_m^n\right) = 0$$

x_m^n is limited within $[0, x_{max} = \min\{x_1, x_r\}]$. The optimum buying strategies can be categorized into the following cases:

Case 1 $x_m^n = 0$;
Conditions: $\lambda_m^2 = 0,\ \lambda_m^3 = 0,\ \lambda_m^1 = p_n + k_m c_m f_m^2 + b_m - \mu_m \eta_m$

To incentivize consumers: $\frac{\partial U_m}{\partial x_m^n}\Big|_{x_m^n=0} > 0 \Rightarrow p_n \leq \mu_m \eta_m - k_m c_m f_m^2 - b_m.$

Optimal price: $p_n = \mu_m \eta_m - k_m c_m f_m^2 - b_m$

Case 2 $\mathbf{x_m^n = x_1}$ Conditions: $\lambda_m^1 = 0, \lambda_m^3 = 0, \lambda_m^2 = -1 + \frac{\mu_m \eta_m}{(1+\eta_m x_m^n)[1+\log(1+\eta_m x_m^n)](p_n + k_m c_m f_m^2 + b_m)}$

Feasibility: $p_n \geq \frac{B_m - \sum_{n' \neq n}^{N} \left(p_{n'} + k_m c_m f_m^2 + b_m\right) x_m^{n'}}{S_n - \sum_{m' \neq m}^{M} x_{m'}^n} - k_m c_m f_m^2 - b_m,$

$p_n \leq \frac{\mu_m \eta_m}{(1+\eta_m x_m^n)[1+log(1+\eta_m x_m^n)]} - k_m c_m f_m^2 - b_m.$

Case 3 $\mathbf{x_m^n = x_r}$ Conditions:

$$\lambda_m^1 = 0, \lambda_m^2 = 0, \lambda_m^3 = \frac{\mu_m \eta_m}{(1+\eta_m x_m^n)\left[1+\log\left(1+\eta_m x_m^n\right)\right]} - \left(p_n + k_m c_m f_m^2 + b_m\right)$$

Feasibility:

$$p_n + k_m c_m f_m^2 + b_m \leq \min\left\{\frac{B_m - \sum_{n' \neq n}^{N} \left(p_{n'} + k_m c_m f_m^2 + b_m\right) x_m^{n'}}{S_n - \sum_{m' \neq m}^{M} x_{m'}^n}, \frac{\mu_m \eta_m}{(1+\eta_m x_m^n)\left[1+\log\left(1+\eta_m x_m^n\right)\right]}\right\}$$

Case 4 $\mathbf{0 < x_m^n = x_{max}}$ Conditions: $\lambda_m^1 = 0, \lambda_m^2 = 0, \lambda_m^3 = 0.$
Optimal price: $p_n + k_m c_m f_m^2 + b_m = \frac{\mu_m \eta_m}{(1+\eta_m x_m^n)[1+\log(1+\eta_m x_m^n)]}.$

Leader's Problem (Pricing Strategies).

The 1st and the 2nd order derivatives are very much essential to find out the optimal pricing strategy with the revenue function R_n. The derivatives are:

$$\frac{\partial^2 x_m^n}{\partial^2 p_n} = \frac{\mu_m}{D_1^4 (D_2+1)^2}\left[2D_1(D_2+1) + D_1^2 \frac{\eta_m \frac{\partial x_m^n}{\partial p_n}}{1+\eta_m x_m^n}\right] = \frac{\mu_m}{D_1^4 (D_2+1)^2}\left[2D_1(D_2+1) - \frac{\mu_m \eta_m}{e^{D_2-1}(D_2+1)}\right].$$

Where $D_1 = p_n + k_m c_m f_m^2 + b_m$ and $D_2 = 1 + \log(1 + \eta_m x_m^n)$.

For all cases, the second-order derivative of the revenue function satisfies: $\frac{\partial^2 R_n}{\partial^2 p_n} \leq 0$,ensuring that R_n is a concave function. The Rn function clarifies the uniqueness and existence of Nash equilibrium. The Nash equilibrium seems to be very unique around various pricing and semantic stages, and therefore the Mult fan Stackelberg game based semantic pricing also seems to be likely identical. Further, stability along with the simulation results are matriculated using Stackelberg method. It is observed that the convergence is gained with equilibrium outcome which are found to be very robust.

6 The performance analysis

To improvise performance, the proposed approach PoS (proof of semantic) is baselined and compared with several factors: PoTra (the traditional consensus without semantic validation), proof of work (PoW), Raft, Practical Byzantine Fault Tolerance (PBFT). The metrics were calculated for utility efficiency factors, consensus time, PSNR, the communication over head and accuracy levels.

A. Informal security analysis.

- *Blockchain-Related Threats* Replay attacks and privacy-preserving semantic pricing are the two critical security challenges for the integration part of block chain and semantic communication.A replay attack occurs when malicious actors repeatedly send semantic information to the receivers to gain unfair advantages. To counter this, the proposed framework introduces nonce values that represent the order of transactions. These nonce values are stored on the blockchain, ensuring each transaction is unique and thus, preventing the reusability of semantic data.

Privacy-Preserving Semantic Pricing: Storing semantic information on public blockchains makes it vulnerable to unauthorized copying, allowing consumers to access it without payment. Additionally, consumers cannot verify the authenticity of semantic data before making payments. To address these issues, the proposed solution incorporates Zero-Knowledge Proof (ZKP). ZKP allows semantic information to be encoded and decoded securely. The ZKP preserves the consumers authenticity by preserving privacy by validating the credentials.*b) Semantic-Related Threats* Insights were drawn from the proposed perspectives of PoS (proof of semantic mechanism) to measure the diversity of security [29]. This analysis identified potential vulnerabilities and attacks specific to semantic communication, ensuring the framework's resilience against such threats.

Targeted Semantic Attack It is one among the attacking techniques describing how attackers alter pixels to make manipulated semantic data appear akin to the original content. This prediction is very significant to safeguard the receiver's background knowledge.

Solution with the Proof of Semantic The idealized proposed methodology defends these attacks only after verifying semantic data before summing it to the blockchain. Let y = original semantic data, y' = tampered data. Let s_i = honest miner's signature, s'_i = attacker's signature. Miners check the validity of y using: $e\left((\mathbf{y}, h_i)^{s_i}, g\right) = e((\mathbf{y}, h_i), g)^{s_i} \neq e\left((\mathbf{y}', h_i)^{s_i'}\right), g)$.

Preventing Bypass Attacks Attackers may try to control miners to skip verification. PoS prevents this by requiring consensus among multiple miners. The calculations include Total miners (n), Honest miners: >2/3 n, Controlled miners (t'): ≤ 1/3 n, Consensus threshold (t). The verification ensures the following approach,

$$e\left(\prod_{i=0}^{t}(\mathbf{y},h_i)^{s_i},g\right)=e\left((\mathbf{y},h),g^{s_i}\right)$$
$$\neq e\left(\overbrace{(\mathbf{y},h_i)^{s_i}\cdots(\mathbf{y},h_i)^{s_i}}^{t-t'}\overbrace{(\mathbf{y},h_i)^{s_i}\cdots(\mathbf{y}',h_i)^{s_i'}}^{t'}\right).$$

The approach guarantees the non-manipulation of data after the miners are being authenticated thoroughly.

A. The Untargeted semantic attack.

In contrast to targeted attacks, the untargeted attacks deliberately maximize differences between authentic (y) and altered (y') semantic information. This degradation strategy disrupts proper functioning of semantic encoders and decoders.

6.1 Mitigation via Proof of Semantic

The proposed verification framework provides identical protection against untargeted attacks as it does for targeted ones, employing the same cryptographic validation and consensus protocol outlined previously.

B Experimental Setup and validation/ simulation results.

To validate the performance of our Proof of Semantic (PoS) mechanism and semantic communication system, we conducted extensive experiments using Go 1.18, PyTorch 1.11.0, and Torch vision 0.12.0. The consensus mechanism was implemented with a threshold of t = n/3 + 1, where n represents the total number of nodes. For evaluation purposes, we utilized two standard benchmark datasets - MNIST and Fashion-MNIST (FMNIST) - with each model undergoing 30 pretraining epochs followed by 30 training epochs. The MNIST and FMNIST are the commonly existing datasets relatively works with the wide variety of semantic approaches. Since, these are tiny datasets existing, they hardly compete with real world semantic web communication system. We considered these MNIST and FMNIST initially to validate the approach of proof of semantic and web framework integration activities [13, 14, 17]. The extension activity may be scoped involving higher valued datasets which may be for semantic activities. Some of the datasets used in the scope of semantic web are CIFAR – 10/100 imageNet. The neural architecture featured distinct configurations for different components. The PoS verification system employed a single fully connected layer in both encoder and decoder, and the semantic sharing mechanism utilized three fully connected layers for enhanced feature extraction.

Channel conditions were carefully controlled to simulate real-world scenarios, with Additive White Gaussian Noise (AWGN) channels configured at both low (3dB) and high (10dB) signal-to-noise ratios (SNR) to evaluate system performance across different quality regimes. The channel efficiency parameter was fixed at c = 1 throughout the experiments. The implementation procedure incorporated and extended previous work

from [13] and [30], particularly in developing the semantic communication framework for the optimal semantic pricing and sharing.

The Stackelberg game implementation employed the RMSProp optimization algorithm with parameters consistent with [27], featuring 10 leaders and 10 followers. The key game parameters included an initial price $p_n = 2$; uniform costs $c_m = c_n = 1 \times 10^4$; scaling factors $k_m = k_n = 1 \times 10^{-20}$; budget al.locations $f_m = f_n = 1 \times 10^8$; uniformly distributed parameters $\mu_m \sim U(10,20)$ and $\eta_m \sim U(0.7,0.9)$ and fixed values $\phi_n = 10$ and $B_m = 30$. System performance was evaluated across multiple dimensions, The Fig. 2 illustrates consensus efficiency (including both communication patterns and agreement speed) and Fig. 3 demonstrates the classification accuracy on both MNIST and FMNIST datasets under varying noise conditions.

Proof of Semantic Performance: The Fig. 2, constrasts the exhange overhead between the Proof of Semantic (PoSem) and traditional methods (PoTra) as implicated during the complete procedure. At a compression rate (CR) of 0.2, PoSem reduces communication overhead by up to 32.8% by eliminating unnecessary data during the procedure of consensus. Figure 2 shows PoSem's consensus time versus Raft, PBFT, and PoW. Semantic web may incur with several nodes, and these nodes may be prone to malicious acts. Using of Raft across these nodes may not be a viable step to handle these malicious acts. PoSem on the other hand, is very good at showcasing predictable performance with PBFT (practical Byzantine Fault Tolerance). It is clear that performance metrics is proportional to the total number of nodes in the semantic web. For scalability, the PoS (proof of semantic) is dependent on light weight semantic signature verification compared to $O(n^2)$ message complexity. The consensus attainment is considered once t – out- of -n signatures are aggregated. This mechanism is entailed in Fig. 2. It is also observed that the consensus time is lowered than PBFT which is better considered for large scale web 3.0 services for semantic web. Figures 3 and 4 depicts the comparisons of

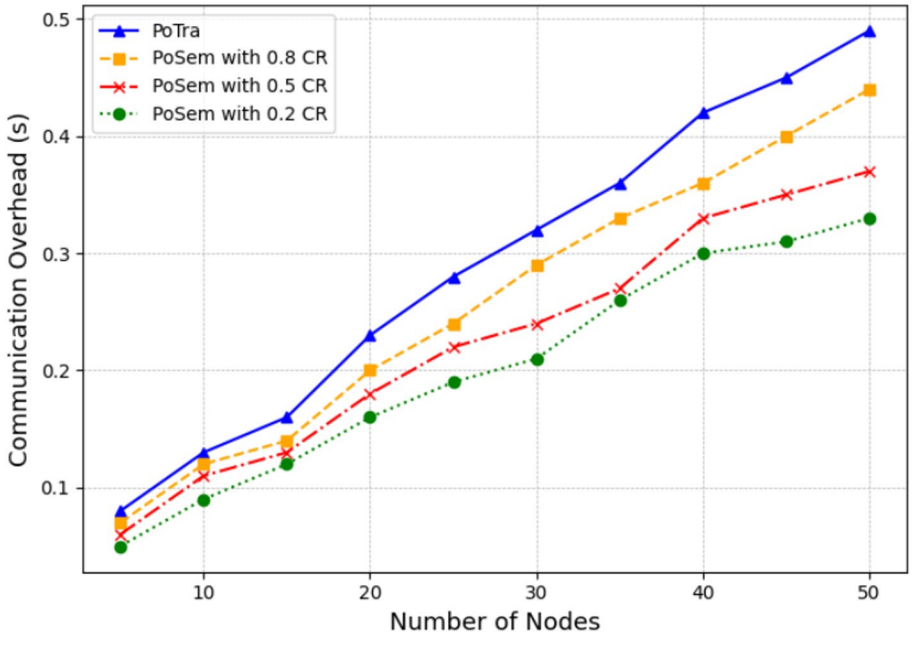

Fig. 2 Proof of semantic depicting the performance attributes (a) Communication

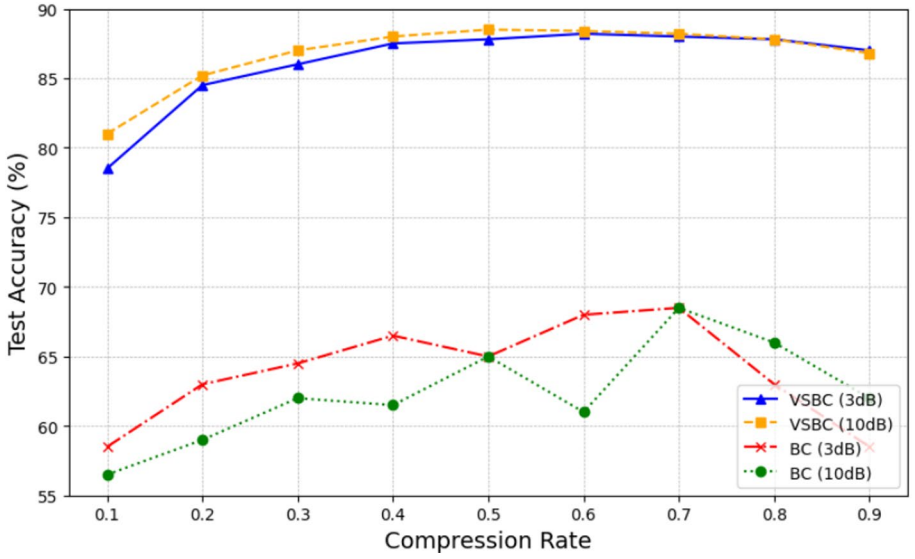

Fig. 3 Verifiable Semantic Block chain depicting Accuracy metrics (**a**) Accuracy with MNIST

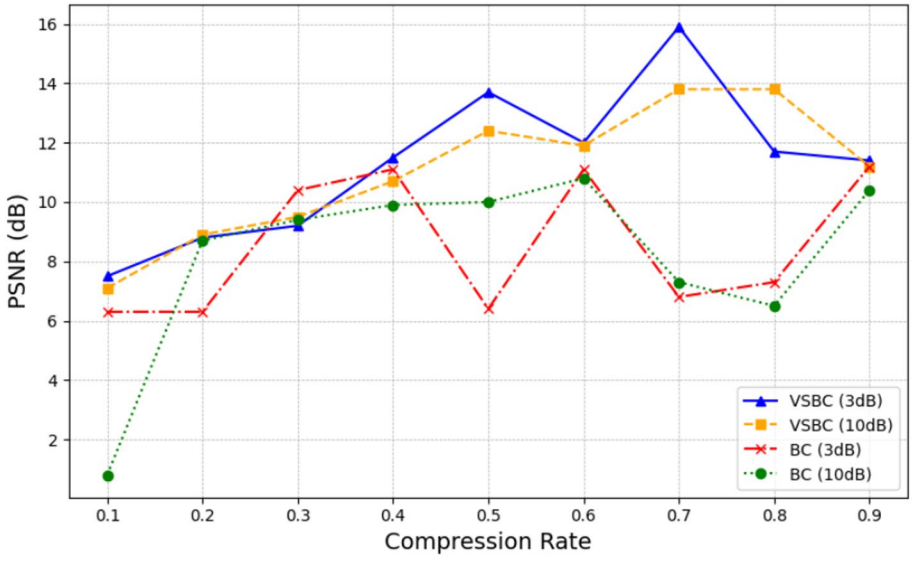

Fig. 4 Performance of MNIST with PSNR over verifiable semantic blockchain depicting reconstruction of semantic validations

semantic blockchain (SBC) over traditional practices of blockchain (BC) just by leveraging 30% of malicious data.

SBC maintains better accuracy with PSNR and MNIST/FMNIST datasets. It achieved up to 20% better accuracy and outperforms at both 3dB and 10dB SNR levels. Unlike standard semantic communication (which typically achieves 30-50dB PSNR without attacks), malicious data causes unpredictable encoder/decoder outputs, lowering PSNR levels. Verifiable Semantic Blockchain (VSBC) counters these unpredictable outputs by verifying all semantic data before blockchain inclusion, thus prevent corrupted inputs from affecting results. The complete semantic communication system is incomplete without the integration of Blockchain in terms of security. To increase the redundancy and security, integration framework would be the one among various approaches.

Semantic construction without the blockchain cannot enhance security measure and increased redundancy [13, 14, 17]. Figure 2, demonstrates the usage of Raft, giving more faster integration agreement but facilities less security. On the other hand, the proposed Posem generated similar efficacy with validations for security measures [18–20]. pricing and fixed strategies produced unstable outcomes and are very inconsistent. Figures 7, 8 and 9, the Stackelberg approach in our methodology and showcased the accurate costs over semantic sharing with blockchain.

Semantic Sharing Performance: The Fig. 5 depicts the comparison of the semantic sharing mechanism with VQA [31] following variable and dynamic perception. Both variable and dynamic perceptions perform equally well on MNIST. Semantic sharing mechanism (SemS) exhibits VQA on FMNIST by extracting more task-relevant semantic information. The Fig. 5 shows accuracy under fixed SNR conditions. The Fig. 6 demonstrates that PoSem and SNR are directly proportional to each other and increases their values proportionally. Together, these results show that our approach is magnificent in producing efficient semantic sharing with minimal accuracy loss in stable channels.

Semantic Pricing Performance: A detailed analysis is made on the game-based pricing mechanism and is clearly given in Figs. 7 and 8. In Fig. 7, the Prices and revenue initially rise before stabilizing at Nash equilibrium, Revenue drops in 7(b) due to competition from other leaders' pricing strategies, in 7(a), increased message availability (Sn) lowers prices but may boost revenue. The Fig. 8, shows the utility growth with parameters μm and ηm, confirming convergence. In Fig. 9, Our proposed method outperforms random/fixed strategies in utility efficiency. The results validate that our mechanism inherently adapts pricing based on message availability and competition.

From the above scenarios, all the comparisons clearly extract that the proposed method has better consensus for semantic web 3.0 with blockchain ensuring best reduced overhead, security, less redundancy and fair communication system.

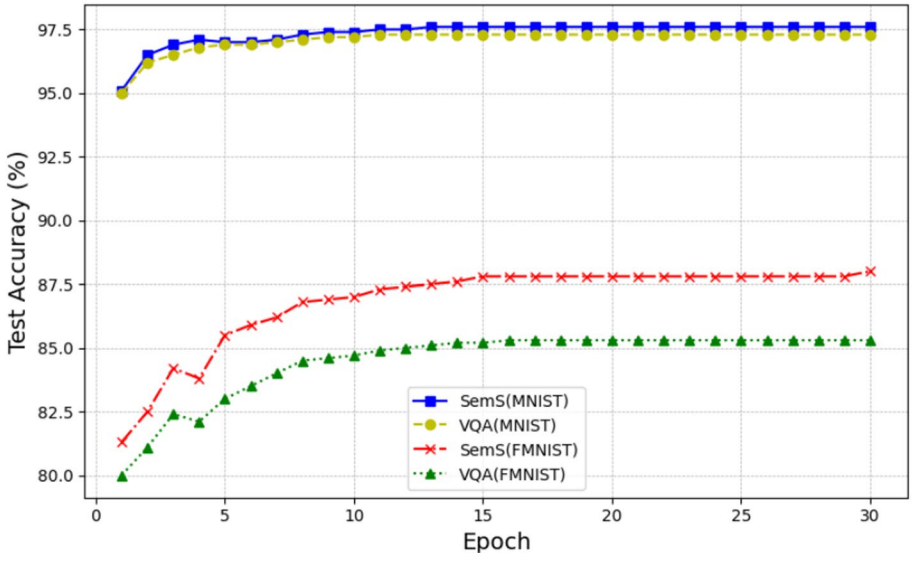

Fig. 5 Semantic Sharing depicting Accuracy metrics. **a** Accuracy with Dynamic Channel Conditions

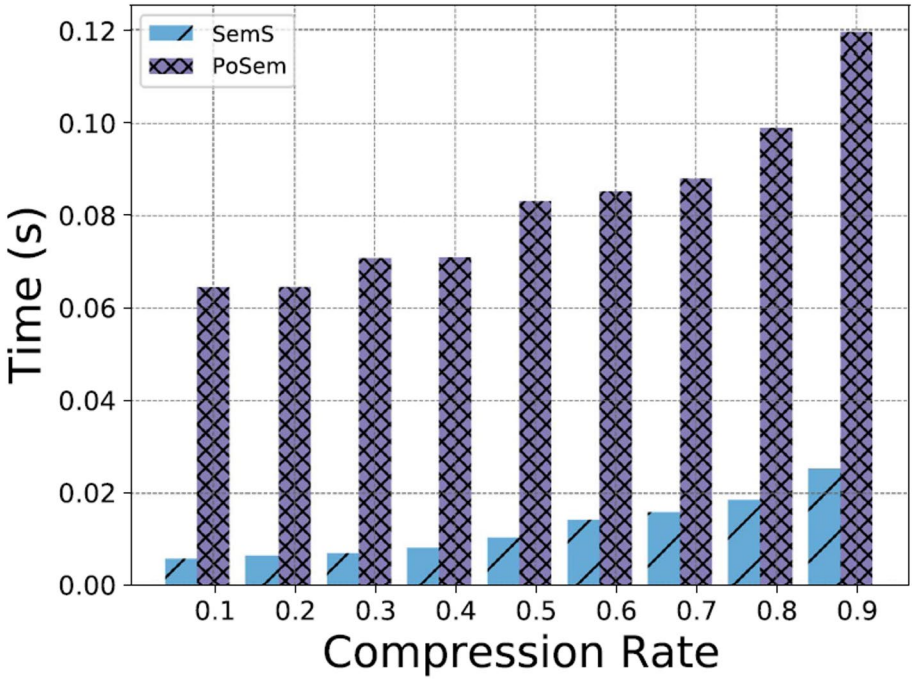

Fig. 6 (poSem) and Semantic sharing mechanism (SemS)depicting performance

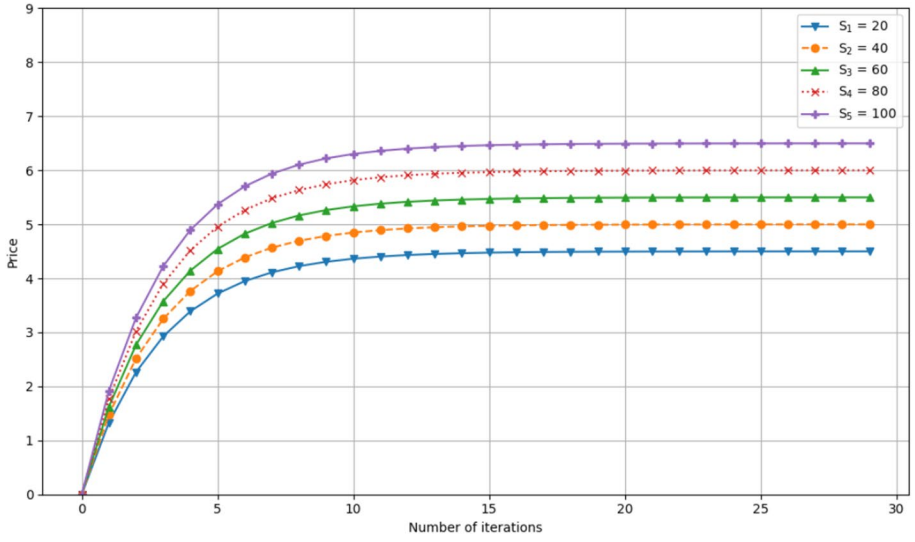

Fig. 7 Stackelberg game's convergence pricing strategies with number of iterations under equilibrium conditions

7 Limitations

Despite, the proposed method promotes reduced communication overhead and controlled simulations with accuracy, there are still few limitations which are on hold. Initially, the experiment methodology carried out is limited to datasets being used only on MNIST and FMNIST. These datasets may not fully represent the web 3.0 environment. Secondly, the dynamic network environment is not analysed for heterogeneity and scalability. Thirdly, the security patch levels are at the primary level as this is the preliminary approach for PoS. Finally, the usage and implementation of Stackelberg pricing model may be deeply investigated to the robustness of integration framework. The study did

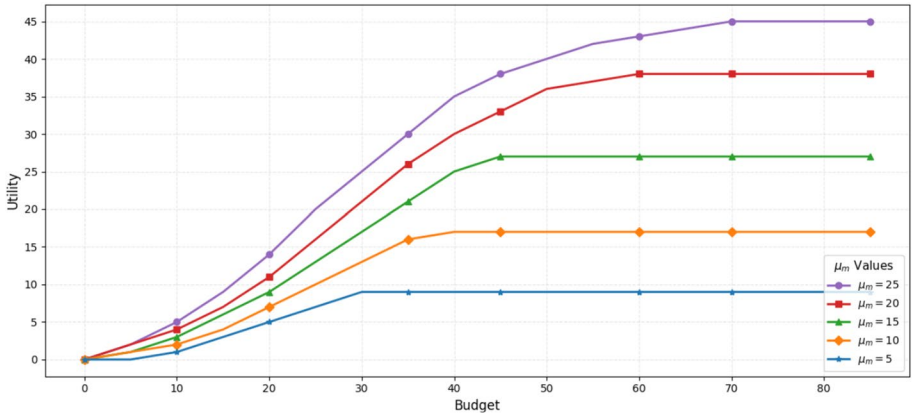

Fig. 8 Follower depicting Convergence metrics

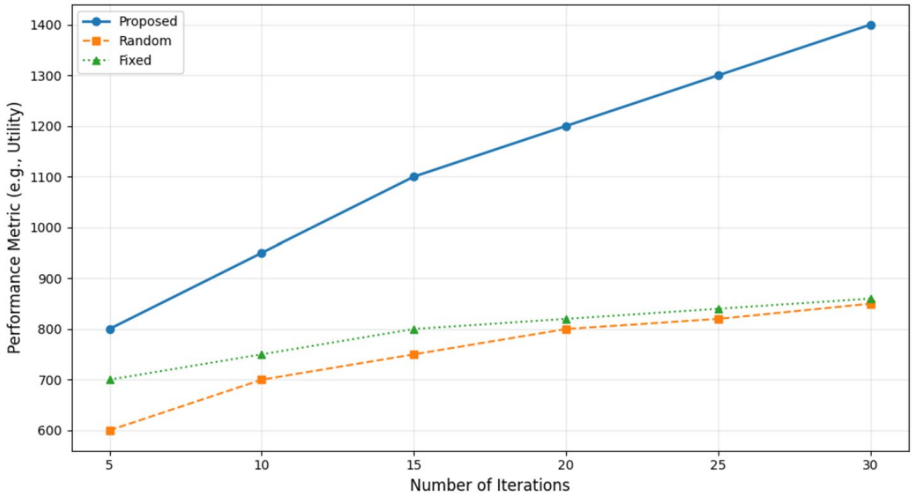

Fig. 9 Semantic Pricing depicting Total utilities

not lay emphasis on regulatory implications of proof of semantic approach and will be addressed in future work.

8 Conclusion

In conclusion, this paper introduced a novel framework for Web 3.0 that leverages blockchain and semantic communication to enable secure, decentralized, and transparent sharing of semantic information. At the core of this system, a Proof of Semantic based mechanism is deployed for threshold signatures, to ensure the reliability of shared data and addresses issues related to data availability. Additionally, we paired a semantic pricing strategy, modelled through a multi-leader, multi-follower Stackelberg game to optimize benefits for both information providers and consumers. The implementation results showcased that the proposed approach have evolved with reduced overhead with improvised efficacy compared to base line methods. Our methods also stated that the base lined datasets are utilized for semantic web sharing 3.0 with blockchain. Moving forward, our research will explore Zero-Knowledge Proof techniques to further guarantee the authenticity of semantic data in future generations.

Author contributions
VS - Framework Construction, model development, GS- Software development, Srihari BG- Literature Survey, HR Battu-Documentation, OOsman- Model deployment and error finding, JRasheed- Result Analysis and Comparisons.

Funding
The research has not received any funding.

Data availability
The datasets generated and/or analysed during the current study are available in the [Kaggle] repository, **[** MNIST: [https://www.kaggle.com/datasets/hojjatk/mnist-dataset](https:/www.kaggle.com/datasets/hojjatk/mnist-dataset)FMNIST: [https://www.kaggle.com/datasets/zalando-research/fashionmnist](https:/www.kaggle.com/datasets/zalando-research/fashionmnist) **]**.

Declarations

Ethical approval and consent to participate
Not Applicable.

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Computer Science and Engineering, GITAM School of Technology, GITAM (Deemed to be University), Hyderabad, India
[2]Computer Science and Engineering, School of Engineering, Anurag University, Hyderabad, India
[3]Computer Science and Engineering (Data Science), Geethanjali College of Engineering & Technology, Hyderabad, India
[4]Computer Science and Engineering, Koneru Lakshmaiah Education Foundation(KLEF), Guntur, India
[5]Department of Electrical and Electronics Engineering, Istanbul Topkapi University, Istanbul, Turkey
[6]Department of Computer Engineering, Istanbul Sabahattin Zaim University, Istanbul 34303, Turkey
[7]Department of Software Engineering, Istanbul Nisantasi University, Istanbul 34398, Turkey
[8]Research Institute, Istanbul Medipol University, Istanbul 34810, Turkey
[9]Applied Science Research Center, Applied Science Private University, Amman, Jordan

Received: 4 July 2025 / Accepted: 20 November 2025
Published online: 02 December 2025

References

1. Guo Y, Xie H, Miao Y, andX CW, Jia. FedCrowd: afederated and privacy-preserving crowdsourcing platform on blockchain. IEEE Trans Serv Comput. 2020;15(4):2060–73.
2. Wang M, Guo Y, Zhang C, Wang C, Huang H, Jia X. MedShare: A privacy-preserving medical data sharing system by using blockchain. IEEE Trans Serv Comput. 2021;16(1):438–51.
3. Wanting Yang H, Du ZQ, Liew WYB, Lim Z, Xiong D, Niyato X, Chi X, Shen, and Chunyan Miao. 2023. Semantic Communications for Future Internet: Fundamentals, Applications,Challenges. Commun. Surveys Tuts. 25, 1 (Firstquarter 2023), 213–250. https://doi.org/10.1109/COMST.2022.3223224
4. Aiting Yao S, Pal X, Li Z, Zhang C, Dong F, Jiang. Xiao Liu,A privacy-preserving location data collection framework for intelligent systems in edge computing, ad hoc Networks, 161, 2024,103532, ISSN 1570–8705, https://doi.org/10.1016/j.adhoc.2024.103532
5. Cano-Benito J, Cimmino A, García-Castro R. Towards blockchain and semantic web. In: Abramowicz W, Corchuelo R, editors. Business information systems Workshops. BIS 2019. Lecture Notes in Business Information Processing. Volume 373. Cham: Springer; 2019. https://doi.org/10.1007/978-3-030-36691-9_19.
6. Aiting Yao S, Pal X, Li Z, Zhang C, Dong F, Jiang X, Liu. A privacy-preserving location data collection framework for intelligent systems in edge computing. Ad Hoc Netw, 161,2024,103532, ISSN 1570 8705, https://doi.org/10.1016/j.adhoc.2024.103532
7. Chen C, et al. When digital economy Meets Web3.0: applications and challenges. IEEE Open J Comput Soc. 2022;3:233–45. https://doi.org/10.1109/OJCS.2022.3217565.
8. Xiangjuan X, Hong W, Guang L, Shouyi Z. Integration and innovation of blockchain in web 3.0. World Wide Web. 2024. https://doi.org/10.1007/s11280-024-01319-7.
9. Liu Z, et al. MakeWeb3. 0 connected. IEEE Trans Dependable Secure Comput. 2022;19(5):2965–81.
10. Liu C, et al. Extending On-chain trust to Off-chain–Trustworthy blockchain data collection usingtrusted executionenvironment (TEE). IEEE Trans Comput. 2022;71(12):3268–80.
11. Cai T, et al. Scalable on-chain and off-chain blockchain for sharing economy in large-scale wireless networks. IEEE Wirel Commun. 2022;29(3):32–8.
12. Tong H, Yang Z, Hu SWY, Saad W. and C.Yin, Federated learning-based audio semantic communication over wireless networks, in Proc.IEEE Glob. Commun. Conf., 2021, pp. 1–6.
13. Zhang H, Shao S, Tao M, Bi X, Letaief KB. Deep learningenabled semantic communication systems with task-unaware transmitter and dynamic data. IEEE J Sel Areas Commun. 2023;41(1):170–85.

14. Bourtsoulatze E, Kurka DB, Gunduz D. Deep joint source-channel coding for wireless image transmission. IEEE Trans Cognit Commun Netw. 2019;5(3):567–79.
15. Molchanov P, Tyree S, Karras T, Aila T, Kautz J. Pruning convolutional neural networks for resource efficient inference. ArXiv Preprint arXiv:1611.06440. 2016.
16. Samuel N, Diskin T, Wiesel A. Learning to detect. IEEE Trans Signal Process. 2019;67(10):2554–64.
17. Shao J, Mao Y, Zhang J. Task-oriented communication for multidevice cooperative edge inference. IEEE Trans Wirel Commun. 2023;22(1):73–87.
18. Jankowski M, Gündüz D, Mikolajczyk K. Deep joint source-channel coding for wireless image retrieval, in Proc. Int. Conf. Acoust., Speech Signal Process., Barcelona, Spain, 2020, pp. 5070–5074.
19. Cheng R, Wu N, Chen S, Han B. Will metaverse be nextg internet? Vision, hype, and reality. IEEE Netw. 2022;36(5):197–204.
20. Tong W, Li GY. Nine challenges in artificial intelligence and wireless communications for 6G, IEEE Wireless Commun., vol. 29, no. 4, pp. 140–145, Aug. 2022.
21. Ye H, Liang L, Li GY. Decentralized Learning With Unreliable Communications, IEEE J. Select. Topics Signal Proc., Apr. 2022, pp. 487–500.
22. Yang Z, Liu K, Chen Y, Chen W, Tang M. Two-level Stackelberg game for IoT computational resource tradingmechanism: Asmart contract approach. IEEE Trans Serv Comput. 2022;15(4):1883–95.
23. Du H, et al. Attention-aware resource allocation and QoE analysis for metaverse xURLLC services. IEEE J Sel Areas Commun. 2023;41(7):2158–75.
24. Du H, et al. Exploring attention-aware network resource allocation for customized metaverse services. IEEE Netw Early Acess. 2022. https://doi.org/10.1109/MNET.128.2200338.
25. Li X, Yin X, Ning J. Trustworthy announcement disseminationscheme with blockchain-assisted vehicular cloud. IEEE Trans Intell Transp Syst. 2023;24(2):1786–800.
26. Weiss MB, Werbach K, Sicker DC, Bastidas CEC. On the application of blockchains to spectrum management. IEEE Trans Cogn Commun Netw. 2019;5(2):193–205.
27. Lin Y, et al. A novel architecture combining oracle with decentralized learning for IIoT. IEEE Internet Things J. 2023;10(5):3774–85.
28. Rosen JB. Existence and Uniqueness of Equilibrium Points for Concave N-Person Games. Econometrica, vol. 33, no. 3, 1965, pp. 520–34. JSTOR. https://doi.org/10.2307/1911749. Accessed 1 Oct. 2025.
29. Du H, et al. Rethinking wireless communication security in semantic internet of things. IEEE Wirel Commun. 2023;30(3):36–43.
30. Li Z, Wang W. Multi-operator dynamic spectrum sharing for wireless communications: A consortium blockchain enabled framework. IEEE Trans Cogn Commun Netw. 2023;9(1):3–15.
31. Xie H, Qin Z, Tao X, Letaief KB. Task-oriented multi-user semantic communications. IEEE J Sel Areas Commun. 2022;40(9):2584–97.

Publisher's note