SCIENCE CHINA
Information Sciences

• LETTER •

November 2024, Vol. 67, Iss. 11, 219303:1–219303:2
https://doi.org/10.1007/s11432-024-4159-2

Trustworthy DNN partition for blockchain-enabled digital twin in wireless IIoT networks

Xiumei DENG[1], Jun LI[6*], Long SHI[2*], Kang WEI[3], Ming DING[4], Yumeng SHAO[2], Wen CHEN[5] & Shi JIN[6]

[1]*Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore* 487372*, Singapore;*
[2]*School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing* 210094*, China;*
[3]*Department of Computing, Hong Kong Polytechnic University, Hong Kong* 999077*, China;*
[4]*Data61, CSIRO, Sydney* NSW 2015*, Australia;*
[5]*Department of Electronics Engineering, Shanghai Jiao Tong University, Shanghai* 200240*, China;*
[6]*National Mobile Communications Research Laboratory, Southeast University, Nanjing* 210096*, China*

Received 5 May 2024/Revised 17 August 2024/Accepted 18 September 2024/Published online 9 October 2024

The integration of artificial intelligence (AI) and digital twin (DT) technology has revolutionized the industrial Internet of Things (IIoT), enabling advanced automation and intelligent manufacturing [1]. Through sophisticated interactions between physical entities and their virtual counterparts, AI-driven DTs facilitate performance monitoring, analysis, simulation, and optimization of physical assets, enabling predictive maintenance and informed decision-making [2]. However, state-of-the-art deep neural network (DNN) architectures that demand substantial computational resources result in prolonged execution times, posing challenges for IIoT networks that necessitate timely completion of each processing step during manufacturing [3, 4]. Furthermore, the extensive data requirements of AI-driven DTs render IIoT networks vulnerable to device malfunctions and cyber-attacks. Integrating blockchain technology in DT-assisted IIoT networks mitigates these risks by ensuring transparent and secure data tracking, promoting trustworthy interactions among DTs. Nevertheless, blockchain-enabled DTs for IIoT networks face two primary challenges: 1) lightweight IIoT devices lack sufficient computing power for resource-intensive DNN inference tasks, and 2) conventional blockchain consensus algorithms, such as proof of work (PoW), are computationally prohibitive for these resource-constrained devices.

To address these challenges, we propose a novel three-tier blockchain-enabled DT (B-DT) framework for wireless IIoT networks. Our main contributions are as follows: 1) a DNN partitioning scheme where gateways execute bottom-layer inference tasks and offload top-layer tasks to access points (APs); 2) a reputation-based consensus mechanism that evaluates the off-chain reputation of each AP based on its computational contributions to DNN tasks and utilizes the off-chain reputation as a stake to adjust the block generation difficulty of the on-chain PoW; 3) a Dynamic DNN Partitioning and Resource Allocation (DPRA) algorithm to jointly optimize communication (i.e., partition point) and computation resource allocation (i.e., computation frequency of APs for top-layer DNN inference and block generation) for a communication and computation efficient wireless B-DT system; and 4) an analysis of the DPRA algorithm demonstrating its asymptotic optimality and characterizing an $[\mathcal{O}(1/V), \mathcal{O}(V)]$ trade-off between system scalability and trustworthiness.

System model. We consider a blockchain-enabled DT (B-DT) system in wireless IIoT networks. The B-DT system consists of three tiers: the physical twin tier with multiple IIoT devices, the digital twin tier with multiple gateways, and the edge server tier with multiple APs. Each device holds a local dataset that is continuously collected from its equipped sensors and running applications, and keeps synchronizing its collected data with the corresponding DT maintained on its associated gateway. Considering that the system operates in slotted time, we denote the set of time indices by $\mathcal{T} = \{1, 2, ..., t, ..., T\}$, where t represents an individual time slot index, and T is the total number of time slots. At the beginning of each time slot, each gateway runs its locally maintained DTs to perform the bottom-layer DNN inference, and transmits the forward output of the bottom-layer DNN inference to its associated AP. Upon receiving the forward output of the bottom-layer DNN inference from the gateways, each AP performs the top layers of the DNN inference, and finally outputs its inference results. Then, each AP encrypts its DNN inference results by a unique digital signature, and exchanges the DNN inference results with the other APs over the peer-to-peer network. After that, the APs add the verified inference results to their respective candidate block, and compete to generate a new block with the proposed consensus mechanism.

Let $\mathcal{N} = \{1, 2, ..., N\}$, $\mathcal{M} = \{1, 2, ..., M\}$, and $\mathcal{J} = \{1, 2, ..., J\}$ denote the index sets of the devices, gateways, and APs, respectively. Define an $N \times M$ connection matrix $\boldsymbol{a}$ with entries $a_{n,m} \in \{0, 1\}$, $\forall n \in \mathcal{N}$ and $m \in \mathcal{M}$. If $a_{n,m} = 1$, the n-th device is deployed with the m-th gate-

[*] Corresponding author (email: jleesr80@gmail.com, slong1007@gmail.com)

way on the same factory floor and communicates with the m-th gateway to maintain its corresponding DT, referred to as the n-th DT. Note that each device's corresponding DT is maintained by a single gateway, i.e., $\sum_{m=1}^{M} a_{n,m} = 1$, $\forall n \in \mathcal{N}$.

Reputation based consensus mechanism. We develop a reputation-based hybrid consensus mechanism by integrating PoS and PoW. Calculating the top-layer DNN reference tasks offloaded to each AP, we evaluate the off-chain reputation of the j-th AP in the t-th time slot as

$$U_j(t) = g\left(O_j(t)\right), \tag{1}$$

where $g(\cdot)$ is a generic off-chain reputation evaluation function, which can be further customized according to specific reputation evaluation rules in diverse networks,

$$O_j(t) = \sum\nolimits_{m=1}^{M} \sum\nolimits_{n=1}^{N} b_{m,j} a_{n,m} D_n(t) \sum\nolimits_{l=l_n(t)+1}^{L_n} \chi_n^l, \tag{2}$$

is the top-layer DNN reference tasks offloaded from the associated gateways to the j-th AP, and χ_n^l denotes the FLOPs required by the n-th DT to perform the l-th layer of the DNN inference for each data point. From (1), the AP with more computational contributions to the DNN inference tasks can achieve higher off-chain reputation values. The core of the proposed reputation-based consensus mechanism is that the off-chain reputation can be used as the stake to adjust the block generation difficulty of on-chain PoW. In this way, the degradation of PoW consensus security is compensated by off-chain reputation stake, which improves the consensus efficiency while ensuring the trustworthiness on the chain. The block generation difficulty is inversely proportional to the off-chain reputation of the APs. From [5], the relationship between the block generation difficulty and off-chain reputation is represented as

$$\gamma_j(t) = e^{-\alpha U_j(t) - \beta}, \tag{3}$$

where α and β control the influence of the off-chain reputation values on block generation difficulty and the final convergence of block generation difficulty, respectively.

Problem formulation. To obtain a communication and computation efficient wireless B-DT system, we jointly optimize the communication resource (i.e., DNN partition point) and the computation resource (i.e., computation frequency for DNN inference and block generation) allocation. Let $\boldsymbol{X}(t) = [\boldsymbol{l}(t), \boldsymbol{f}^{\mathrm{A}}(t), \boldsymbol{f}^{\mathrm{bloc}}(t)]$, where $\boldsymbol{l}(t) = [l_1(t), l_2(t), ..., l_N(t)]$ denotes the number of bottom layers of the DNN inference tasks performed at the gateways at time slot t, $\boldsymbol{f}^{\mathrm{A}}(t) = \left[f_1^{\mathrm{A}}(t), f_2^{\mathrm{A}}(t), ..., f_M^{\mathrm{A}}(t)\right]$ denotes the computation frequency of the APs for DNN inference tasks offloaded from each gateway at time slot t, and $\boldsymbol{f}^{\mathrm{bloc}}(t) = \left[f_1^{\mathrm{bloc}}(t), f_2^{\mathrm{bloc}}(t), \ldots, f_M^{\mathrm{bloc}}(t)\right]$ denotes the computation frequency of APs for block mining at time slot t, respectively. The stochastic optimization problem is formulated as

$$\mathbf{P0}: \min_{\boldsymbol{X}(t)} \overline{\tau} = \frac{1}{T}\sum\nolimits_{t=1}^{T} \tau(t) \tag{4}$$

$$\begin{aligned}
\text{s.t.}\quad & \mathbf{C1}: 1 \leqslant l_n(t) \leqslant L_n, \forall n \in \mathcal{N}, t \in \mathcal{T},\\
& \mathbf{C2}: 0 \leqslant \sum\nolimits_{m=1}^{M} b_{m,j} f_m^{\mathrm{A}}(t) \leqslant f_j^{\max}, \forall j \in \mathcal{J}, t \in \mathcal{T},\\
& \mathbf{C3}: 0 \leqslant f_j^{\mathrm{bloc}}(t) \leqslant f_j^{\max}, \forall j \in \mathcal{J}, t \in \mathcal{T},\\
& \mathbf{C4}: 0 \leqslant e_m^{\mathrm{G}}(t) \leqslant E_m^{\mathrm{G}}(t), \forall m \in \mathcal{M}, t \in \mathcal{T},\\
& \mathbf{C5}: 0 \leqslant e_j^{\mathrm{A}}(t) \leqslant E_j^{\mathrm{A}}(t), \forall j \in \mathcal{J}, t \in \mathcal{T},\\
& \mathbf{C6}: U^{\min} \leqslant \frac{1}{T}\sum\nolimits_{t=1}^{T} U_j(t) \leqslant U^{\max}, \forall j \in \mathcal{J},
\end{aligned}$$

where the constraints **C1**–**C3** bound the ranges of the variables $\boldsymbol{l}(t)$, $\boldsymbol{f}^{\mathrm{A}}(t)$, and $\boldsymbol{f}^{\mathrm{bloc}}(t)$, respectively. **C4** and **C5** are the energy consumption constraints for devices and gateways in each time slot, respectively. In order to guarantee both the scalability and trustworthiness of the B-DT system, the long-term constraint **C6** is adopted to bound the average off-chain reputation of each AP.

Problem solution. We propose a dynamic DNN partitioning and resource allocation (DPRA) algorithm to solve the long-term stochastic optimization problem formulated in **P0**. By leveraging the Lyapunov optimization method, DPRA first transforms the long-term stochastic optimization problem into a sequence of one-shot static optimization problems. Subsequently, the algorithm solves the transformed deterministic problem in each time slot.

Experimental results. For comparison purpose, we simulate two baselines as follows: (a) computation resource allocation policy without DNN partitioning point optimization (WDPO), and (b) DNN partitioning point optimization and computation resource allocation policy without reputation based consensus mechanism (WTCM). DPRA shows a lower system latency than baseline schemes. Compared with WTCM, DPRA adjusts the block generation difficulty according to the off-chain reputation, which reduces the system latency while guaranteeing the trustworthiness of the B-DT system.

Conclusion. This paper proposes a communication and computation efficient B-DT framework for wireless IIoT networks, incorporating a DNN partitioning method. Experimental results demonstrate DPRA's superiority in reducing latency and ensuring B-DT system trustworthiness compared to baseline approaches. Future work includes conducting real-world experiments to measure DNN inference task latency and energy consumption, and developing lightweight consensus mechanisms for resource-constrained IoT devices to enhance blockchain efficiency and security.

Acknowledgements This work was supported in part by Key Technologies R&D Program of Jiangsu (Prospective and Key Technologies for Industry)(Grant Nos. BE2023022, BE2023022-2), in part by National Natural Science Foundation of China(Grant No. 62371239), and in part by Jiangsu Specially-Appointed Professor Program 2021.

Supporting information Appendixes A–H. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Wu Y, Zhang K, Zhang Y. Digital twin networks: a survey. IEEE Internet Things J, 2021, 8: 13789–13804

2 Guo D, Zhong R Y, Rong Y, et al. Synchronization of shop-floor logistics and manufacturing under IIoT and digital twin-enabled graduation intelligent manufacturing system. IEEE Trans Cybern, 2023, 53: 2005–2016

3 Lu Y, Huang X, Zhang K, et al. Communication-efficient federated learning for digital twin edge networks in industrial IoT. IEEE Trans Ind Informatics, 2021, 17: 5709–5718

4 Guo Q, Tang F, Kato N. Federated reinforcement learning-based resource allocation for D2D-aided digital twin edge networks in 6G industrial IoT. IEEE Trans Ind Informatics, 2023, 19: 7228–7236

5 Yang Z, Yang K, Lei L, et al. Blockchain-based decentralized trust management in vehicular networks. IEEE Internet Things J, 2019, 6: 1495–1505