Received: 11 November 2023 | Revised: 2 February 2024 | Accepted: 5 February 2024

DOI: 10.1002/itl2.510

LETTER

WILEY

PRICELESS: Privacy enhanced AI-driven scalable framework for IoT applications in serverless edge computing environments

Muhammed Golec[1,2] | **Mustafa Golec**[3] | **Minxian Xu**[4] | **Huaming Wu**[5] | **Sukhpal Singh Gill**[1] | **Steve Uhlig**[1]

[1]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK

[2]Abdullah Gul University, Kayseri, Turkey

[3]Faculty of Engineering, Computer Engineering, Dumlupınar University, Kütahya, Turkey

[4]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

[5]Center for Applied Mathematics, Tianjin University, Tianjin, China

Correspondence
Muhammed Golec, School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road, Bethnal Green, London E14NS, UK.
Email: m.golec@qmul.ac.uk

Funding information
Chinese Academy of Sciences President's International Fellowship Initiative, Grant/Award Number: 2023VTC0006; National Natural Science Foundation of China, Grant/Award Number: 62071327

Abstract

Serverless edge computing has emerged as a new paradigm that integrates the serverless and edge computing. By bringing processing power closer to the edge of the network, it provides advantages such as low latency by quickly processing data for time-sensitive Internet of Things (IoT) applications. Additionally, serverless edge computing also brings inherent problems of edge and serverless computing such as cold start, security and privacy that are still waiting to be solved. In this paper, we propose a new Blockchain-based AI-driven scalable framework called PRICELESS, to offer security and privacy in serverless edge computing environments while performing cold start prediction. In PRICELESS framework, we used deep reinforcement learning for the cold start latency prediction. For experiments, a cold start dataset is created using a heart disease risk-based IoT application and deployed using Google Cloud Functions. Experimental results show the additional delay that the blockchain module brings to cold start latency and its impact on cold start prediction performance. Additionally, the performance of PRICELESS is compared with the current state-of-the-art method based on energy cost, computation time and cold start prediction. Specifically, it has been observed that PRICELESS causes 19 ms of external latency, 358.2 watts for training, and 3.6 watts for prediction operations, resulting in additional energy consumption at the expense of security and privacy.

KEYWORDS

artificial intelligence, cold start, IoT, privacy, security, serverless edge computing

1 | INTRODUCTION

Edge computing is a paradigm that provides benefits such as low latency and bandwidth savings by moving processing power and storage closer to where data is produced.[1] Edge nodes may consist of various hardware and sensors, including but not limited to Internet of Things (IoT) devices, routers, and switches.[2] In addition to the advantages it offers, edge computing typically comprises devices characterized by restricted processing capabilities and storage capacities when compared to traditional cloud systems.[3] For this reason, it may be insufficient to respond to operations that require high

processing power. To solve these concerns of edge computing and also to benefit from the advantages of both edge and cloud computing, the concept of serverless edge computing has emerged, combining these two computing paradigms.[4] Serverless computing is one of the latest service models of cloud computing, which offers advantages such as dynamic scalability, economical pricing model, and abstraction of infrastructure from the customer by integrating Function as a Service (FaaS) and Backend as a Service (BaaS) service models.[1] In serverless edge computing, serverless platforms are deployed on edge nodes, providing great advantages at the edge of the network. Some of these advantages are[3]: (i) dynamic scalability, where resources can be scaled when necessary, (ii) an event-driven architecture that provides advantages for real-time applications, and (iii) an economic model where resources are charged only for time they are used. While serverless edge computing includes the benefits inherent in both paradigms, it also inherits challenges from each, for example, cold start latency, security, and privacy.[5] These challenges are as follows:

1. Cold Start Latency: In serverless computing, the execution of each function involves assigning it to a container, and once the execution process is completed, these containers are released to prevent unnecessary resource consumption (scaling to zero).[4] However, to restart these released containers when required, certain operations, including preparing the runtime environment and establishing function libraries, must be performed. The delays incurred by these processes are termed cold start latency. In contexts such as virtual reality and autonomous vehicle scenarios, where swift response times are critical, these delays can negatively affect Quality of Service (QoS) parameters, particularly user experience. Consequently, ongoing research endeavors are focused on resolving the cold start latency challenge in serverless computing.
2. Security and Privacy: Serverless edge computing brings security and privacy problems, given its reliance on resource-constrained devices and the dispersed nature arising from its heterogeneous composition.[5] Since edge devices are used in scenarios such as healthcare applications, they process sensitive data (biometric data) and disclosure or capture of this data may cause many privacy problems.[6] Therefore, strong security mechanisms such as authentication and secure communication protocols must be implemented.

1.1 | Motivation and contributions

Serverless edge computing is a new paradigm that combines serverless and edge computing paradigms, bringing functions closer to the edge of the network where data is processed. It has advantages such as faster processing of data with lower latency rate (edge computing) and economic model-dynamic scalability (serverless computing). In addition to these advantages, it also brings with it problems such as cold start latency, security, and privacy inherited from edge and serverless computing. In our previous work, we developed ATOM framework that predicts cold start latency and request number using a Deep Deterministic Policy Gradient (DDPG) in serverless edge computing.[7] These two prediction results are very important as they provide a basis for future cold start prevention studies. In another study, we developed a framework BlockFaaS that provides security and privacy for serverless-based IoT healthcare applications.[8] In BlockFaaS framework, Blockchain module provides data immutability to offer security and privacy. The security of all communication channels is protected by Transport Layer Security (TLS) (version 1.3), thus preventing privacy issues. To provide security and privacy in serverless edge computing environments while focusing on cold start latency of IoT healthcare application, we developed PRICELESS framework to integrate BlockFaaS with ATOM framework. The main contributions of this paper are: (1) Propose a PRICELESS to ensure security and privacy for serverless edge computing environments and predicts cold start latency for IoT healthcare application, (2) Investigate the impact of blockchain on cold start latency estimation, and (3) Compare the performance of PRICELESS framework is compared with the current state-of-the-art method based on energy cost, computation time and cold start prediction.

2 | RELATED WORK

Literature reports that research on serverless computing is evolving. Golec et al.[9] presented iFaaSBus, a framework that provides security and privacy for serverless computing environments. iFaaSBus framework uses the TLS protocol for the security of communication channels; they use the OAuth-2.0 Authorization Protocol and JSON Web Token Structure to ensure privacy. Vahidinia et al.[10] proposed a two-layer approach to reduce cold start frequency in serverless computing. In the first layer, it uses a Reinforcement Learning (RL) based model to learn request patterns. The second layer decides the number of containers to be kept warm with the Long Short-Term Memory (LSTM) model. It should not be forgotten that

TABLE 1 Comparison of PRICELESS with existing works.

Work	Privacy	Environment	Model	Scalability	IoT	Energy	Generating dataset
Golec et al.[9]	✓	Serverless computing	ML	✓	✓		
Vahidinia et al.[10]		Serverless computing	RL & DL	✓			
Aslanpour et al.[2]		Serverless edge computing	WS & SO	✓		✓	
Agarwal et al.[11]		Serverless computing	RL	✓		✓	✓
Jegannathan et al.[12]		Serverless computing	ML	✓		✓	
PRICELESS	✓	Serverless edge computing	DRL	✓	✓	✓	✓

keeping containers warm for the next requests will reduce cold start. Aslanpour et al.[2] proposed an energy-aware scheduling mechanism for serverless edge computing. Their innovation involved crafting region-specific and priority-driven algorithms designed to manage nodes operating under resource constraints and renewable energy sources. To achieve this, they deployed strategies centered on Warm Scheduling (WS) and Sticky Offloading (SO) and tested its performance by implementing on the Raspberry Pi devices. In Agarwal et al.[11] conducted a study aiming to reduce the frequency of cold starts in serverless computing with a RL-based model (i.e., Q-Learning). The state and reward in the RL algorithm are designed by monitoring the CPU usage. To reduce the frequency of cold starts, function instances are prepared in advance depending on the result of the RL model. It is a known fact that another factor affecting cold start latency is container preparation time. Jegannathan et al.[12] utilized a time-series-based (SARIMA) model to reduce the container preparation time, that is, the cold start frequency, by estimating the time of requests. Table 1 shows a comprehensive analysis of the reviewed literature studies and comparison with PRICELESS. To the best of our knowledge, there is no study in the literature that considers cold start latency, security, and privacy simultaneously in serverless edge computing.

3 | METHODOLOGY

In this paper, a cold start dataset[i] is created using a heart disease risk-based IoT application[8] and deployed using Google Cloud Functions (GCF). Further, we have developed a new framework called PRICELESS by integrating a Blockchain module of BlockFaaS[8] with the ATOM framework[7] to provide security and privacy in serverless edge computing environments while focusing on the cold start latency of IoT healthcare applications. It has been identified that this module causes additional delays in the cold start dataset. While these delays are measured in milliseconds, their significance becomes more pronounced in time-sensitive scenarios. Consequently, to quantify and comprehend this additional latency, identical experiments were conducted to generate the cold start dataset using GCF. The experimental results revealed that the blockchain module induced an extra delay of 19 ms in the cold start dataset. The comparison between the cold start datasets from ATOM and PRICELESS post-deployment of the blockchain module is illustrated in Figure 1.

3.1 | PRICELESS: Proposed model

PRICELESS framework provides security and privacy for serverless edge-based applications while predicting the cold start latency. In scenarios where sensitive data, such as biometric data, is applied, it is important to ensure the security of communication channels and biometric data.[8] PRICELESS uses the TLS 1.3 protocol to secure communication channels. TLS encrypts data to ensure confidentiality in communication channels between the edge and serverless layer and ensures confidentiality by preventing this data from being viewed by unauthorized persons. In addition, incoming and outgoing data via API Gateway, which is used in communication between the serverless layer and the outside world, is encrypted by TLS. It should be noted that TLS is used in the HTTP protocol in the PRICELESS. The PRICELESS uses Blockchain to ensure data immutability (security). To prevent data manipulation, Blockchain obtains the hash values of the data transmitted through the hash algorithm and connects these values together in the form of blocks. This can be easily noticed as the smallest change in any data will affect the hash value. Blockchain inherently works in perfect harmony with decentralized architectures such as serverless computing. In time-sensitive scenarios such as Virtual Reality and autonomous vehicles in serverless edge computing, latency is desired to be as low as possible. However, in the serverless paradigm, cold start latency due to the scale-to-zero feature may pose a problem for these scenarios. PRICELESS can monitor the serverless edge environment, detect cold starts in advance, and warn the user.

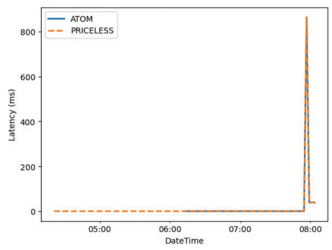

FIGURE 1 Impact of the Blockchain module on Cold Start Latency.

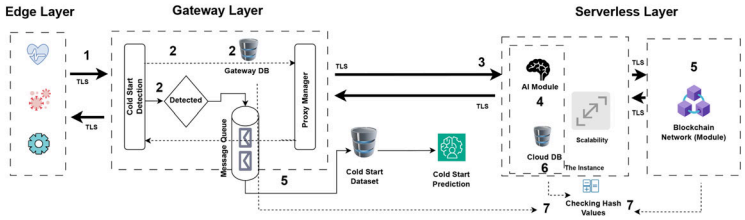

FIGURE 2 Architecture of PRICELESS.

Figure 2 shows the architecture of the PRICELESS framework, which is positioned between the edge layer and serverless layer and creates a dataset with transaction information by constantly monitoring the communication channel. This dataset is used to train the DDPG model, which will be used to predict cold start latency. The edge layer consists of various numbers of edge nodes and can host various applications, such as health applications and predictive maintenance. And at the serverless layer are the AI models that will make sense of the data coming from these applications. It was mentioned that within the framework of PRICELESS, TLS protocol is used to ensure confidentiality in all communication channels, and blockchain is used to ensure security. Now let's explain the working steps of PRICELESS:

1. Blockchain Module for Security: (i) Data sent from the Edge Layer is first sent to the Gateway node and a copy of this data is stored in the Gateway database (dB). (steps 1–2) (ii) The data collected in this layer is sent to the AI module in the Serverless Layer (step 3). The AI module contains an AI model that is responsible for interpreting the data sent and returning a result. (In this scenario, the health data sent from the Gateway node is processed in the Light Gradient Boosting Machine model in the AI module to determine the risk of heart disease.) (iii) The data coming to the AI module and the result are sent to the Blockchain network (module) to obtain a hash value (step 5). (In this scenario, the blockchain network connects the health data and the heart disease risk detection result and obtains a hash value. This process is repeated for all users and the resulting hash values are connected to the next block). (iv) A copy of this hash value is stored in the cloud database (step 6). The result obtained from the AI model with the data previously stored in Gateway dB is sent to the Blockchain module and a second hash value is obtained. (v) In the last step, the hash value obtained in the first process (Cloud DB) is compared with the hash value obtained in the second process (Gateway DB) (step 7). Matching the two hashes guarantees data immutability between the Gateway Layer and the Serverless Layer. Additionally, all obtained hash values can be stored in all edge nodes and used as a distributed ledger.
2. Cold Start Prediction: Predicting the cold starts that occur is important in terms of laying the foundation for future resource-sensitive cold start prevention studies. That's why PRICELESS uses the DDPG model, which has proven to be successful in cold start prediction.[7] To train the DDPG model, a dataset created with transaction information logged by the PRICELESS framework (steps 2–5) is used.

3.2 | Evaluation metrics

The formulas used to calculate the energy consumption and latency for the ATOM and PRICELESS frameworks are as follows:

$$E_{\text{cost}} = \rho \times t, \tag{1}$$

$$C_{\text{Latency}} = \iota_{\text{first}} - \iota_{\text{second}}. \tag{2}$$

where energy consumption E_{cost} is obtained by multiplying the thermal design power of the processor CPU_{TDP} and the processing time t. The latency $C_{Latency}$ that occurs on a serverless platform is calculated by subtracting the first request from the second one.

4 | EXPERIMENTS

In this section, we observed the amount of additional latency that PRICELESS brings to the cold start dataset and its impact on cold start prediction performance in the ATOM framework. Additionally, the amount of energy consumed by the Blockchain module is calculated by comparing the energy costs of PRICELESS and ATOM framework. The system configurations on which all experiments were performed are as follows: CPU: Intel Core i7-10 750, Frequency Speed: 2.6–5.00 GHz, RAM: 16 GB, TDP (Watt): 45. Additionally, the model hyperparameter values for the DDPG model are as follows: "Nf = 2, LRa = 0.0001, LRc = 0.01, Nah = 30, Nch = 30, M AXep = one hundred". Serverless environment parameters values are as follows: "Platform: – GCP, Region: Europe-west2b, Runtime Python 3.11, Function trigger type: HTTP, Memory 256 MB".

4.1 | Impact of blockchain on cold start latency

Table 2 shows the observed delay for proposed PRICELESS and existing ATOM framework. The results show an additional latency of 19 ms in the PRICELESS framework, which is due to the time it takes in the blockchain to hash each block and add it to the chain.

4.2 | Impact of blockchain on cold start prediction

As shown in Table 3, there was a decrease in prediction performance because of Blockchain's impact on the correlation between time series.

4.3 | Impact of blockchain on QoS

These results are obtained using Equation 1 are shown in Figure 3. The times spent for training and prediction in the ATOM Framework are 118.76 and 0.12 s, respectively. This corresponds to ~5344.2 Watts and 5.4 Watts of energy consumption. Likewise, the times spent on training and prediction for the PRICELESS Framework are 126.92 and 0.20 s, respectively. This corresponds to ~5702.4 Watts and 9 Watts of energy consumption, respectively. In other words, the additional load that the Blockchain module brings to the ATOM framework at the expense of security and privacy is 358.2 Watts for training and 3.6 Watts for prediction operations. As a result, the PRICELESS framework provides higher security and privacy than the ATOM framework, thanks to blockchain and TLS protocol. However, it causes external latencies caused

TABLE 2 The impact of Blockchain on latency.

Work	Normal latency (ms)	Cold start latency (ms)
ATOM[7]	18–32	800–2000
PRICELESS	37–51	800–2000

TABLE 3 The impact of Blockchain on cold start prediction.

Work	MAE	RMSE	R2 Score
ATOM[7]	51.57	148.76	0.071
PRICELESS	52.30	142.52	0.186

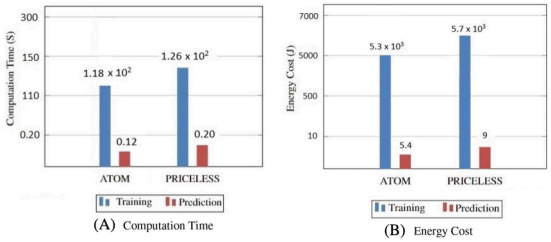

FIGURE 3 Comparison of PRICELESS and ATOM frameworks in terms of computation time and energy consumption.

by the blockchain module, and these latencies affect the dataset used in the training of the DDPG model and reduce the performance of cold start prediction. Additionally, it may require additional energy consumption for blockchain-based operations at the edge layer, which includes resource-constrained devices such as IoT.

5 | CONCLUSIONS

This paper proposes a new privacy-enhanced, AI-driven scalable framework called PRICELESS, which offers privacy and security in serverless edge computing and enables cold start prediction by integrating the BlockFaaS framework into the ATOM framework. We measured the impact on cold start prediction performance and the additional time introduced by the Blockchain module. The performance of the proposed framework (PRICELESS) is compared with the existing framework (ATOM) in terms of performance parameters such as computation time, energy consumption, and cold start latency. These results show that PRICELESS ensures privacy and security by increasing energy usage during training and prediction. In the future, PRICELESS can be extended to improve cold start prediction performance and reduce energy consumption, which can be utilized in energy-sensitive cold start prevention studies.

ACKNOWLEDGMENTS

Muhammed Golec would express his thanks to the Ministry of Education of the Turkish Republic, for funding. This work is partially supported by Chinese Academy of Sciences President's International Fellowship Initiative (No. 2023VTC0006) and National Natural Science Foundation of China (No. 62071327).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available at https://github.com/MuhammedGolec/ColdStart-Dataset.

ENDNOTE

[i] https://github.com/MuhammedGolec/ColdStart-Dataset

ORCID

Muhammed Golec https://orcid.org/0000-0003-0146-9735
Sukhpal Singh Gill https://orcid.org/0000-0002-3913-0369

REFERENCES

1. Cao K, Liu Y, Meng G, Sun Q. An overview on edge computing research. *IEEE Access*. 2020;8:85714-85728.
2. Aslanpour MS, Toosi AN, Cheema MA, Gaire R. *Energy-Aware Resource Scheduling for Serverless Edge Computing*. IEEE; 2022:190-199.
3. McGrath G, Brenner PR. *Serverless Computing: Design, Implementation, and Performance*. IEEE; 2017:405-410.
4. Baresi L, Mendonça DF. *Towards a Serverless Platform for Edge Computing*. IEEE; 2019:1-10.
5. Xie R, Tang Q, Qiao S, Zhu H, Yu FR, Huang T. When serverless computing meets edge computing: architecture, challenges, and open issues. *IEEE Wirel Commun*. 2021;28(5):126-133.
6. Khan AA, Wagan AA, Laghari AA, Gilal AR, Aziz IA, Talpur BA. BIoMT: a state-of-the-art consortium serverless network architecture for healthcare system using blockchain smart contracts. *IEEE Access*. 2022;10:78887-78898.
7. Golec M, Gill SS, Cuadrado F, et al. ATOM: AI-powered sustainable resource management for serverless edge computing environments. *IEEE Trans Sustain Comput*. 2023;1-13. doi:10.1109/TSUSC.2023.3348157
8. Golec M, Gill SS, Golec M, et al. BlockFaaS: blockchain-enabled serverless computing framework for AI-driven IoT healthcare applications. *J Grid Comput*. 2023;21(4):63.
9. Golec M, Ozturac R, Pooranian Z, Gill SS, Buyya R. IFaaSBus: a security-and privacy-based lightweight framework for serverless computing using IoT and machine learning. *IEEE Trans Industr Inform*. 2021;18(5):3522-3529.
10. Vahidinia P, Farahani B, Aliee FS. Mitigating cold start problem in serverless computing: a reinforcement learning approach. *IEEE Internet Things J*. 2022;10(5):3917-3927.
11. Agarwal S, Rodriguez MA, Buyya R. *A Reinforcement Learning Approach to Reduce Serverless Function Cold Start Frequency*. IEEE; 2021:797-803.
12. Jegannathan AP, Saha R, Addya SK. *A Time Series Forecasting Approach to Minimize Cold Start Time in Cloud-Serverless Platform*. IEEE; 2022:325-330.

How to cite this article: Golec M, Golec M, Xu M, Wu H, Gill SS, Uhlig S. PRICELESS: Privacy enhanced AI-driven scalable framework for IoT applications in serverless edge computing environments. *Internet Technology Letters*. 2024;e510. doi: 10.1002/itl2.510