A Study of Universal Zero-Knowledge Proof
Circuit-based Virtual Machines that Validate General Operations & Reduce Transaction Validation

Soonhyeong Jeong¹ and Byeongtae Ahn²,*

¹ Onther Inc., 527, Gangnam-daero, Seocho-gu, Seoul, Republic of Korea
kevin.j@onther.io
² Faculty of Liberal & Arts College, Anyang University, 22, 37-Beongil, Samdeok-Ro,
Manan-Gu, Anyang 430-714, Republic of Korea
ahnbt@anyang.ac.kr

Abstract. Recently, blockchain technology accumulates and stores all transactions. Therefore, in order to verify the contents of all transactions, the data itself is compressed, but the scalability is limited. In addition, since a separate verification algorithm is used for each type of transaction, the verification burden increases as the size of the transaction increases. Existing blockchain cannot participate in the network because it does not become a block sink by using a server with a low specification. Due to this problem, as the time passes, the data size of the blockchain network becomes larger and it becomes impossible to participate in the network except for users with abundant resources. Therefore, in this paper, we studied the zero knowledge proof algorithm for general operation verification. In this system, the design of zero-knowledge circuit generator capable of general operation verification and optimization of verifier and prover were also conducted. Also, we developed an algorithm for optimizing key generation. Based on all of these, the zero-knowledge proof algorithm was applied to and tested on the virtual machine so that it can be used universally on all blockchains.

Keywords: Zero-Knowledge, validation, transaction, BlockChain, Ethereum

1. Introduction

The blockchain-based distributed application market is expected to grow from about $3.2 billion in 2019 to more than $60 billion in 2024. Among them, the market with 'transaction processing' as a profit model is expected to reach 55% of the total. This means that blockchain-based distributed applications are generally provided on the basis of open source, so transaction fees rather than content usage fees are inevitably accepted by users. Therefore, the economic value of the technology to efficiently process transactions is very positive. In the past decade, numerous blockchain implementations have emerged as platforms, but there has been no significant innovation in terms of accumulating and storing transactions, but rather the burden of verifying the chain data

*Corresponding author
has increased as it supports complex operations. Therefore, it is necessary to lead the structural innovation of the blockchain by creating a verification module that can be commonly used in various blockchain platforms that will appear in the future.

Currently, the domestic blockchain technology is mainly biased toward mainnet-based technologies such as distributed ledgers and consensus algorithms, but the area where the domestic technology ecosystem is likely to lead in the global market is the distributed application area rather than the mainnet. And there is currently no virtual machine based on zero-knowledge proof that can efficiently verify complex operations required for distributed applications, not just bookkeeping. Blockchain is a decentralized digital ledger that secures the integrity of transaction details and allows participants to share details without the involvement of a trusted third party in a peer-to-peer (P2P) network. A typical example of applying blockchain is cryptocurrency such as Bitcoin and Ethereum. Ethereum introduced the Ethereum virtual machine (hereafter EVM). With EVM, users can program their own way, rather than performing a predefined set of tasks. However, EVM is very inefficient compared to existing virtual machines such as Java Virtual Machine (JVM). And it is difficult to support a complex application environment. Therefore, it is necessary to lead the structural innovation of the blockchain by creating a verification module that can be commonly used in various blockchain platforms.

Blockchain is classified as a simple type of blockchain made of UTXO (Unspent Transaction Output) represented by Bitcoin, and a complex type of blockchain that deals with a state tree such as Ethereum. Currently, zero-knowledge proof is used only when processing some transactions in a simple form of blockchain. Zero-knowledge proof is a system that proves to the verifier that the proofer knows that knowledge without revealing the knowledge and information he knows [1]. The proofer is the subject that proves that he / she knows the knowledge, and the verifier is the subject that verifies that the proofer knows the knowledge. When zero-knowledge proof technology is applied to storage of transaction data, data storage space can be saved by compressing data in a way that pruning actual data and leaving only proof of data. As time goes by, the data of the blockchain will gradually accumulate, and accordingly, the computing resources required to operate the full node are gradually increasing.

In the case of Ethereum, it is already difficult for an individual to operate a full node, and in the future, only a large company or large hands that can have sufficient computing resources can operate the full node. These factors will lead to the centralization of the blockchain, and this problem can be solved by reducing the resources required for data storage and verification through a virtual machine with zero knowledge proof technology. Therefore, in this paper, we developed a zero-knowledge proof algorithm capable of general operation verification and designed a zero-knowledge circuit generator capable of general operation verification. Also, by applying and testing the zero-knowledge proof algorithm to the virtual machine, the performance of the transaction can be improved. Section 2 of this paper introduces related research, and Section 3 introduces domestic and foreign cases. In section 4, an algorithm capable of transaction verification is proposed, and in section 5, a zero-knowledge circuit to be applied to a virtual machine is designed. Finally, Section 6 presents conclusions and future tasks.
2. Related Studies

Blockchain technology can be divided into a simple type of blockchain made of UTXO (Unspent Transaction Output) and a complex type of blockchain that deals with the State Tree [2]. Currently, in the simple form of blockchain, zero-knowledge proof is used at the protocol level only in some transaction processing. However, although some complex forms of blockchain use smart contracts using smart contracts, there are limitations in terms of performance and utilization because they are implemented in the upper layer. Proof size of a single operation created through the proposed SNARKs algorithm is about 1,500 bytes (1.5 kbytes) [3].

- **Bullet Proof algorithm.**
  - Transaction size of UTXO-based blockchain is measured in \((\text{in} \times 254 \times 146 + \text{out} \times 254 \times 33 + 10)\) bytes, and increases arithmetically according to the number of \(\text{in}, \text{out}\) used.
  - It occupies about 45,000 bytes (45kbytes) based on 1 \(\text{in}\) and 1 \(\text{out}\).

- Regardless of the type of transaction, the transaction size can be fixed to 1.5 kbytes, and even the simplest transaction standard is more than 70% economical.

- The blockchain-based distributed application market is expected to grow from about $3.2 billion in 2019 to more than $60 billion in 2024 (Blockchain Market Shares, Market Strategies, and Market Forecasts, 2018 to 2024, IBM, 2018). Among them, the market with 'transaction processing' as a profit model is expected to reach 55% of the total.

Since such a blockchain-based distributed application is generally provided on the basis of open source, transaction fees rather than content usage fees are inevitably accepted by users. Therefore, the economic value of the technology to efficiently process transactions is very positive. Even if all verification nodes do not participate in block verification, the general operation is verified with the same security strength as all nodes participated and verified using zero-knowledge proof technology, thereby providing the same effect as saving the entire transaction without saving all transaction data [4].

Currently, as the value of using personal information increases, discussions on how to provide personal information have been actively conducted. Currently, one of the most common methods of providing personal information is a group that uses personal information to obtain personal consent and use personal information. However, the above method has two problems. First, information that is more than the information required by the institution for personal information is being exposed. Second, whenever a company requests personal information, there is a problem that a trusted party must provide authentication information for the information to the company. In order to solve the above problems, this paper proposes a privacy-protected personal information management method using zk-SNARK (zero-knowledge Succinct Non-interactive ARgument of Knowledge) technique and blockchain[5]. zk-SNARK is a modification of the existing ZKP to be more succinct and applicable in a non-interactive environment. This logic was first proposed in 2012, and due to its characteristics, ZKP can be implemented in a blockchain environment. In the case of a blockchain transaction using zk-SNARK, the validity of the transaction can be communicated to nodes other than the sending and receiving node without exposing information such as a receiver, a sender, and a transfer amount. ZCash is the first application of zk-SNARK, and related contents
Soon yeong Jeong and Byeongtae Ahn

were applied to Ethereum's Byzantium hard fork [6]. The zk-SNARK is largely divided into two parts, one is the process of converting the problem to be proved to a specific form, and the other is the process of actual proofing using the converted problem. The privacy-protected personal information management technique can guarantee the privacy and reliability of information when providing personal information through zk-SNARK. In addition, it is possible to manage personal information data while ensuring the integrity of the data through the blockchain, and sharing personal information can be performed more easily than the existing authentication method. The privacy-protected personal information management technique can guarantee the privacy while ensuring privacy when providing personal information through zk-SNARK. In addition, it is possible to manage personal information data while ensuring the integrity of the data through the blockchain, and sharing personal information can be performed more easily than the existing authentication method [7].

3. Domestic & International cases

There are several companies with blockchain virtual machine technology. The most representative virtual machine is EVM, Ethereum's virtual machine. EVM is the first blockchain virtual machine and based on EVM, Ethereum has grown into a basic platform for smart contracts, tokens, and decentralized applications (Dapps). And many blockchain projects are using Ethereum's EVM when creating the mainnet. Currently, Ethereum is planning to upgrade to Ethereum 2.0, and when Ethereum 2.0 is introduced, the current virtual machine EVM will be converted to eWASM [7].

EOS-VM is a virtual machine created by EOS.IO and is not limited to the blockchain industry, and is expected to be used in traditional software development fields such as game engines, databases, and web frameworks. EOS-VM is a virtual machine dedicated to the blockchain system, and it can be expected to save development resources (CPU), improve blockchain scalability, and improve development efficiency compared to the first blockchain virtual machine, EVM [8].

Tron's virtual machine TVM is developed based on Ethereum's EVM and is characterized by being compatible with Ethereum. By designing a unique virtual memory mechanism, the amount of memory actually used can be greatly reduced, and the operation cost of a decentralized application can be greatly reduced by providing developers with almost unlimited memory capacity. And you can save resources by optimizing the compiler.

Table 1 shows domestic and International cases as a table.
Currently, the domestic blockchain technology is mainly biased to the underlying technologies related to the main net, such as distributed ledgers and consensus algorithms. Due to the nature of the domestic technology ecosystem, the area that can lead in the global market is the distributed application area rather than the mainnet area. And there is currently no zero-knowledge proof-based virtual machine that can efficiently verify complex operations required for distributed applications, not just bookkeeping.

Therefore, by designing a system for improving the amount of code verification based on zero-knowledge proof applicable to various distributed applications and smart contract execution environments, it will become a distributed application-based technology with great growth potential in the future [9].

4. Algorithm for Transaction Validation

Proof of zero knowledge must satisfy the following three conditions.

* completeness: If a condition is true, a trusted verifier must be able to understand this by a trusted prover.
* soundness: When a condition is false, a dishonest verifier can never convince the verifier that the condition is true by lying.
* zero-knowledge: When a condition is true, the verifier knows nothing other than the fact that this condition is true.

The study intends to utilize zero-knowledge proofs for various types of transactions that the user wants, as well as predefined types of transactions. Circuits that can produce evidence of current zero-knowledge proofs can only perform operations in a predefined form. In order to be able to utilize this in various types of transactions desired by users, a circuit capable of verifying general operations is required. General operation means universal and various operations, not specific predefined operations. Therefore, the research team researched a circuit capable of verifying general operations and designed the method to apply it to the virtual machine. In addition, by using the zero-knowledge proof technology, even if a blockchain participant does not know the contents of the block, it can quickly verify that the contents of the block are not forged or tampered with by the node performing the proof and reporting role among all nodes. Also, by rapidly increasing the block sync speed for participants, new participants can quickly join the network [10].
Fig. 1. is about the existing transaction verification and data storage method. The verification amount increases by re-executing the block for all nodes except the minor. And as the chain data connection increases, the data size also increases linearly.

Fig. 1. Existing Transaction Validation & Data Storage Method

Fig. 2. improves the existing transaction verification and data storage method. This problem was solved by reducing the resources required for data storage and verification through a virtual machine with zero knowledge proof technology.

Fig. 2. Advanced Transaction Validation & Data Storage Method
The requirements of the zero-knowledge proof-based algorithm capable of general operation verification are as follows.

The circuit to be created in this study should be arithmetic. In the finite field F, the F-arithmetic-circuit takes the element in F as the input value, and the output value is also the element in F. The circuit consists of a gate and a wire, and the gate takes two numbers as inputs, adds or multiplies numbers, and outputs the result through the output wire. The wire is responsible for passing the value into and out of the gate.

Circuits must: Local Consistency Check:
Verify that all gate equations are met.
Global Consistency Check: Verify that the wires correctly connect the gates together to form the circuit.

The zk-SNARK structure considered in this study is based on cryptographic pairing. zk-SNARK for F-arithmetic-circuit receives key generator G, attester P, and verifier V as input values [11].

The key generator G samples the proof key pk and verification key vk using the security parameters λ and F-arithmetic-circuit C: Fn * Fh → Fl.

The above values are public parameters of the verification system that needs to be generated only once per circuit. Once set, anyone can generate non-interactive evidence using pk, and anyone can verify this evidence using vk. This is a study on a universal zero-knowledge proof algorithm that can perform general operation verification [12].

A new circuit creation method and a new zk-SNARK for the circuit are needed.

The existing zk-SNARK has the following problems.
* A new setup is required for each new program and a new key needs to be generated accordingly.
* Memory access or the number of repetitions of a loop cannot be changed depending on the input value of the program.
* Even if you allow data dependence on memory access, you have to go through heavy tasks such as verifying the Merkle Pass.
* Circuit size increases inefficiently in proportion to the program size even if an arbitrary program is supported.

The universal circuit should run on any program with less than L instructions, less than T machine steps, and less size. And it should be available in all cases with one key generation. Accordingly, the circuit that satisfies this must satisfy C_L,n,T is O(L+n+T) log(L+n+T) gates [13].

Previously studied universal zero-knowledge proof algorithms, the size of data increases to the size of L * T according to the program size. In this case, as the program size increased, the storage cost increased significantly.

In the case of the newly created zero-knowledge proof algorithm, the size of the data increases in the form of L + T.

The circuit generator and zero-knowledge algorithm are independent of each other. If the circuit generator and the zero-knowledge proof algorithm to be applied to the circuit are independent, a more flexible system can be built [14].

Fig. 3. shows a combination of two elements, a circuit generator and a key generator, for general operation verification. The output C of the circuit is universal because it does not depend on the program or main input values, but only on the L, n, and T values. In this case, as the program size increased, the storage cost increased significantly. In the case of the newly created zero-knowledge proof algorithm, the size of the data increases.
The circuit generator and zero-knowledge algorithm are independent of each other. If the circuit generator and the zero-knowledge proof algorithm to be applied to the circuit are independent, a more flexible system can be built. Fig. 3 shows a combination of two elements, a circuit generator and a key generator, for general operation verification. The circuit's output C is universal because it depends only on the values, not on the program or the main input values. When combined with a circuit verification system such as zk-SNARK, the parameters of the verification system are also universal. In this case, all programs can be verified with a single key generation, and after that, a key suitable for a given calculation range can be selected. Therefore, the cost of generating keys for each program can be reduced [15].

![Fig. 3. Key Generator for General Operation Validation](image)

In Fig. 4, permission for block generation is granted through a validator and a verifier for general operation verification.
The verifier $V$ takes the verification key $v_k$, $\hat{x} \in \mathcal{I}$, and the evidence $\pi$ as input values to verify that the evidence $\pi$ is valid. The operation on $V$ consists of two parts:

* Calculate $v_k x := v_{k1} c_1 + \sum_{i=1}^{n} x_i v_k$ by entering part of $v_k$ and $\hat{x}$.
* Enter $v_k$, $\hat{x}$, and $\pi$ to be able to calculate 12 pairings and perform necessary checks. Regarding the first part of $V$, the variable-based multi-scalar multiplication technique can be used to reduce the amount of computation required for $v_k$ calculation. With respect to the second part of $V$, even if the pairing evaluation takes a certain amount of time regardless of the input size $n$, these evaluations are very expensive and dominate for the small $n$. Our goal is to minimize the cost of these pairing assessments.

The research for the optimizer optimization is as follows.

The proofer $P$ takes the proof key $PK$ (including circuit $C$), $\hat{x} \in \mathcal{I}$, and the witness $\hat{z} \in \mathcal{Z}$ as input values. Proofer $P$ creates evidence $\pi$ and testifies that it is $\hat{x} \in \mathcal{I}$.

The operation on $P$ consists of two parts:

* Calculate the coefficient $H(z) := \frac{a(z)(z)}{z(z)}$ of the polynomial , where $A, B, C \in \mathbb{F}_p$ is derived from the QAP instances $(\hat{A}, \hat{B}, \hat{C}, \hat{z}) := QAP_{\text{inst}}(C, \hat{x})$ and QAP evidence $\hat{z} := QAP_{\text{wiz}}(C, \hat{x})$.
* Calculate $\pi$ by using coefficient $T$, evidence of QAP, and public key $pk$. In particular, in relation to the first part of $P$, the coefficient $T$ is efficiently calculated through the FFT technique of [BCGTV13a].

---

**Fig. 1.** Prover and verifier for general operation validation
5. Design of Zero-knowledge circuit

Due to the nature of the blockchain that stores all transaction data, the data on the blockchain continues to increase over time. When zero-knowledge proof technology is applied to storage of transaction data, data storage space can be saved by compressing the data by pruning actual data and leaving only proof of data. As time goes by, the data of the blockchain will gradually accumulate, and accordingly, the computing resources required to operate the full node will gradually increase [19].

In the case of Ethereum, it is already difficult for an individual to operate a full node, and in the future, it is expected that only large companies or large hands that can have sufficient computing resources can operate the full node. This will cause the centralization of the blockchain, and this problem can be solved by reducing the resources required for data storage and verification through a virtual machine with zero knowledge authentication technology [20].

Fig. 5, is designed to apply the zero-knowledge proof algorithm to the virtual machine. It shows the flow of the operation method of the Ethereum virtual machine for applying zero knowledge proof technology. Since Solidity, the smart contract language of Ethereum, is a language created for human understanding, it needs to be changed to a machine language understandable by a virtual machine in order to operate in a virtual machine. Code written in Solidity is converted to Ethereum bytecode by the compiler. This bytecode is executed by EVM, Ethereum's virtual machine. When a specific bytecode is executed, all nodes in the Ethereum network execute the same bytecode respectively to verify the transaction. At this time, if zero-knowledge proof technology that can perform general operation verification is applied to the virtual machine, even if the virtual machine does not execute the transaction, it is possible to know whether the corresponding transaction is the correct transaction by performing verification on the zero-knowledge evidence [21].
In order to modify the virtual machine, it is necessary to understand the structure. Therefore, Figure 6 shows the architecture and execution flow diagram of the Ethereum virtual machine. Once you understand how the virtual machine is running, you need to figure out what parts of the virtual machine need to be modified to apply zero-knowledge techniques [22].
Fig. 6. Architecture & Execution Flow Chart of Ethereum Virtual Machine
In order to execute the transaction, it needs to be changed to Ethereum bytecode as mentioned. These bytecodes are decomposed into what are called opcodes, stacked on the stack and executed one by one. You must subtract the gas cost for running the virtual machine before the opcode runs. The opcodes are now executed if the gas cost is not insufficient. Fig. 7 shows the parts that need to be changed in the virtual machine when the opcodes are executed [23].

![Diagram showing opcodes and gas cost](image)

**Fig. 7. Change Part in Virtual Machine**

In the figure above, the part marked with a red box is the part to which zero-knowledge proof technology should be applied, and the part to create a universal circuit that can execute the opcode. Fig. 8 shows the change in data stored after the zero-knowledge proof technology is applied. After the zero-knowledge proof technology is applied to the part that performs the opcode, TX data among the data stored in the existing storage is replaced with the proof of the zero-knowledge proof. And we will create and test a virtual machine with zero knowledge proof [24].
Fig. 8. Data Changes after zero knowledge proof technology

6. Conclusion and future works

In this paper, a circuit with zero-knowledge proof algorithm for general operation verification was developed. The core of this paper solved the problem of increasing chain data size and block verification amount in the existing system. In addition, a zero-knowledge proof algorithm was designed for general operation verification [25].
Finally, a study was conducted to optimize the validator and the validator, and a study was conducted to optimize the key generation. This study is a practical example of applying the zero-knowledge proof algorithm capable of semi-operational verification, and can develop two different blockchains in the future. In addition, Crypto Currency implementation with zero-knowledge proof algorithm for general operation verification can be developed. It can also create new blockchain business opportunities, such as platform services, where DApps linked to public blockchains can be integrated with each other. However, anyone can participate in the blockchain network if a blockchain with a zero-knowledge proof-based virtual machine capable of verifying general operations is born [26].

When developing the world's first virtual machine technology based on zero-knowledge proof that can be applied to various distributed applications and smart contract execution environments, Korean companies will secure the foundation technology that can lead the global market in the distributed application market, which has great growth potential in the future. In addition, new blockchain business opportunities are provided, such as platform services where DApps linked to the public blockchain can be integrated with each other. In addition, exports of related products and services will be expanded by revitalizing the blockchain industry and strengthening cooperation with global companies based on core technologies. The increase in expertise of domestic blockchain R & D personnel and the internalization of technology development will also create jobs for R & D personnel. Opportunities for technological innovation on the blockchain and high potential for use in other fields are also provided.

As a future task, based on this design, we will develop a zero-knowledge proof system capable of general operation verification. And based on this research, we will develop our own platform.

Acknowledgement. This research was supported by the Institute for Information &Communication Technology Planning & an evaluation grant funded by the Korea Ministry of Science and ICT (No. 2020-0-00105).

References


Soonhyeong Jeong is a co-organizer of <Seoul Ethereum Meetup>, and this meetup started from November 15, 2014 (even before the launching of ethereummainnet). Kevin is a CEO at the Ethereum blockchain R&D startup OntherInc in South Korea. He received a bachelor’s degree in economics and a master’s degree in Software from Graduate School of Software in SoongSil University in Seoul. Nowadays, he’s interested in Plasma, EVM and Gas System and privacy in public blockchain.

Byeongtae Ahn received his B. Eng. Degree in Digital Media Technology in 2017 from North China University of Technology, Beijing, The People’s Republic of China. Since September 2017, he is in the M. Eng. course at the Department of Multimedia Engineering, Dongguk University, Seoul, Republic of Korea. He has done some works mainly associated with NUI (Natural User Interface) applications, VR (Virtual Reality) applications and artificial intelligence with deep learning.

Received: March 220, 2020; Accepted: February 01, 2021.