A Study on Characteristics and Identification of Smart Ponzi Schemes

LINGLING SONG1 AND XIAOHAN KONG2
1Department of Information Engineering, Binzhou Polytechnic College, Binzhou 256603, China
2Haide College, Ocean University of China, Qingdao 266100, China
Corresponding author: Lingling Song (songlingling@bzpt.edu.cn)

ABSTRACT A smart Ponzi scheme (SPS) is a financial Ponzi scheme that is implemented and deployed in blockchain through smart contract technology. It is built on treachery and lies, by which, the organizers and speculators jointly deceive innocent investors by fostering a belief in obtaining the expected benefits. The occurrence of SPSs is originated from the vulnerability of the supervision mechanism on the blockchain. Although there are many excellent studies, these contributions overemphasized the methods themselves, did not describe the characteristics of the SPS well, and had certain limitations in practice. We made a thorough study on the characteristics of an SPS and brought out the vital features to identify an SPS for an investor. Based on the analysis of the contributions of predecessors, we propose an approach to test whether a contract is an SPS. This approach could deal with two situations, the contracts to be deployed and the long-run contracts respectively. Of the approach, the priori method can be exploited to distinguish whether the contract to be deployed is an SPS for the runner of a blockchain; the posterior method could protect an investor from being trapped in a fraud. At the same time, the posterior method can also be extended to monitor some contracts dynamically to alert users with the probability to be fallen into SPSs.

INDEX TERMS Blockchain, digital currency, smart contract, smart Ponzi scheme.

I. INTRODUCTION

Ponzi scheme is an ancient form of fraud that many people hate and seek after. The scam is built on treachery and lies, by which, the organizers and speculators jointly deceive innocent investors by fostering beliefs in obtaining the expected benefits. Ponzi schemes have entered the digital finance from the traditional financial along with times, and are called smart Ponzi scheme (SPS) [1] in the digital/virtual currency world.

SPSs are rooted in the blockchain ecosystems, especially in those implementing smart contracts. Blockchain technology enlarges the hiddenness of targets invested to, at the same time, smart contracts provide the flexibility to hidden fraud by information technology; the investors couldn’t master or are difficult to know the true information of the initiators in the digital token world. Nevertheless, as a decentralized ledger, blockchain has stored not only full information of all the transactions of the attenders but also the opcode and Application Binary Interface (ABI) of deployed smart contracts. Therefore, many researchers exploit this information to build various methods to protect investors from SPSs.

From the view of researchers, the detection or identification of an SPS is a class-imbalanced problem [18]. Many of them put their focus on how to build an efficient and accurate classification method, such as Machine learning, ANN methods, based on static data. In contrast, some persons try to build a simulation system to run the contracts for distinguishing whether they are SPSs.

In the view of an investor, it is a decision problem. The investors need all the accurate information to determine the probability that the new fund is an SPS, and to trade off the possible loss when investing or not investing. However, most of the researchers could not conduct an in-depth analysis and summary of the characteristics of SPSs although they have built a lot of efficient methods. Therefore, it is very necessary to study the characteristics and natures of SPSs in detail while bringing out a new method.

This paper aims to provide a full of characteristics of an SPS based on bibliometric analysis and to establish a reasonable approach for investors to distinguish whether a fund is an SPS within a blockchain. We make a two steps route, that is from the analysis of the
achievements of predecessors to establishing an approach ourselves.

Therefore, we search the important references, articles, and datasets. Firstly, by inputting “Ponzi scheme smart contract blockchain” as keywords for “subject/title/abstract” into the search engine of EI Village, we got 19 results and obtained 15 articles by excluding the non-relative and the repeated eventually. Then by entering “TS = (Ponzi scheme) and (smart contract) and (blockchain)” as expressions into the search engine of Web of Science, we got 13 results, and we obtained 4 papers by removing the repetitions to EI results and the non-relative. At last, we got 19 papers that talk about the smart Ponzi scheme and have important effects.

The paper is structured as follows: Section II provides the analysis of the full characteristics of an SPS. Section III introduces the current methods to detect SPSs. Section IV talks about the proposed approach to identify an SPS. Section V gives a test to verify the proposed approach. In the last section, VI makes a conclusion.

II. CHARACTERISTICS OF AN SPS

A. DEFINITION OF AN SPS

In 2017, Bartoletti and his colleagues [1] made a concept description to “smart Ponzi scheme” as Ponzi schemes implemented as smart contracts firstly. Then, in 2018, Chen et al. referred explicitly to those Ponzi schemes that disguised themselves within smart contracts as smart Ponzi schemes [2]. Therefore, a smart Ponzi scheme is a financial Ponzi scheme that is implemented and deployed in blockchain through smart contract technology; Smart contracts are merely media and tools to realize Ponzi schemes. As for, an SPS has not only newly added features but also all the characteristics of a traditional Ponzi scheme.

B. ORIGINAL FEATHERS OF A PONZI SCHEME

According to the definition in the U.S. Securities and Exchange Commission and the item of Ponzi scheme in Wikipedia, a Ponzi scheme has features as follows.

- It is investment fraud.
- Its returns to existing investors come from the new investors.
- Its organizer has little or no legitimate earnings.
- Its organizer promises high rewards to new investors.
- It can maintain a sustainable business until it cannot pay off withdrawals demanded by investors because there are not enough new patricians.
- A normal fund would become a Ponzi scheme while its legal income sources are permanently cut off.

C. FEATHERS OF AN SPS

1) ANONYMITY

The creator of each smart Ponzi scheme is anonymous, could be anywhere, and so do the other participants underlying the blockchain [3]–[5]. The anonymity protects against incrimination of the initiators while increasing the difficulty to identify an SPS for an investor.

2) IMmutability

While a smart contract is deployed, it is impossible to modify and is potentially unstoppable, so it is impossible to break the scheme. Therefore, the participants have a high degree of confidence that the return will continue [4]. Once an investor takes part in there is no easy way out [3]–[5].

3) METASTASIS

To cut down their losses, the participants have to continually invite new investors to the scheme while they are aware of being trapped in an SPS. This further hides the creator’s malice and makes it more difficult for investors to distinguish [3].

4) TRUSTWORTHINESS

The appearance of the availability of the source code of a smart contract and the automation of the contract make users feel more reliable and credible. The investors could not believe that it is easy to deploy a scam in this transparent field, while they can see their invested money clearly shown in there [3]–[5].

5) INDISTINGUISHABILITY

It is difficult to extract the features of the Ponzi scheme from the smart contracts though with source code [6]. Like traditional Ponzi schemes, indistinguishability is a vital characteristic of an SPS. Therefore, many people manage to develop various methods or tools to help distinguish whether a scheme is a Ponzi scheme. However, these efforts did not lead to their targets.

6) OPENNESS AND TRANSPARENCY

All deployed contracts, related transactions, and accounts are accessible to anyone linked in the blockchain. All these records the whole information of the contracts and their historical activities, these are the basis to build identification methods.

D. OPERATION FEATURES OF AN SPS

The whole operation of an SPS can be divided into three explicit phrases: bootstrap, hyper-operation, and collapse, which picture the process of a scheme from starting to falling away [7].

- Bootstrap: It is the initial phrase that the fund is known by a very small amount of investors who are waiting.
- Hyper-operation: During this stage, more and more investors take part in the scheme, they all believe that they could get high returns, and invite new investors sincerely.
- Collapse: This phrase starts at the investors can’t withdraw their investment and lose their confidence in the scheme, for, at that time, not enough entered new investments to support the rewards to early investors. In general, before this stage, the cheater will run away, which would accelerate this phrase to come.
E. KINDS OF AN SPS

As we all know, the owner’s income in an SPS comes from investors, so how does he accumulate himself amount, allocate investments and return to investors? The internal structure of the contract answers these questions while determining their kinds. Based on the logical relations of accounts inside the contract, we group all the SPSs into two categories, tree-shaped, and array-shaped. Of each category, there are four important issues, return route, return ratio, reward time, and owner’s income which expose the characteristics of an SPS. With this information, an investor could manually analyze the source code of the smart contract to determine whether it is an SPS.

1) TREE-SHAPED SCHEMES
Within this type of scheme, all the investors are arranged into a tree data structure, the initiator acts as the root; every new investor has a parent node that stores the inviter. This is to say that a new investor must have an inviter. And his investment will be divided into different parts to all his predecessors according to the pre-set ratios.
- Return route: along the branch of the tree to the initiator.
- Return ratio: vary according to his descendants; different branch has a different ratio.
- Reward time: when each descendant registers.
- Owner’s income: a fixed or float ratio of newly entered investment.
Etheramid and DynamicPyramid are good examples of this type [1].

2) ARRAY-SHAPED SCHEMES
All the investors are linearly arranged into an array or list data structure. Generally, the creator takes the place of the head. With this type of scheme, the new investor may have or not have an inviter. Although this is a simple structure, the assignment of returns can change variously depending on the specific distribution algorithm.
- Return route: along the line to the head or from the head.
- Return ratio: fixed or vary according to his descendants.
- Reward time: when each descendant registers or at a certain point matches the pre-set conditions
- Owner’s income: a fixed ratio of newly entered investment.
Bartoletti and his colleagues [1] had found three common subtypes according to their investigation of SPSs.
- Chain-shaped: of which, each of the early investors would get 100% reward on his investment if there enough new investors were going to take part in. Doubler, DianaEtherum, and ZeroPonzi are examples.
- Waterfall schemes: the distribution of return always began from the head of the list with a fixed ratio to initiator balance, as a result, earlier investors got more rewards, and the later entered maybe get no return. TeasureChest and PiggyBack are such types of schemes.
- Handover schemes: the investment will be increased when a new investor comes, and most of the investment will return to the previous. KingOfTheEtherThrone is the representative example.

III. DETECTION METHODS OF AN SPS

Methods to detect an SPS can be grouped into two categories, static process methods, and dynamic process methods. Most of the current research belongs to static process methods, only a few consider the dynamic process methods.

A. STATIC PROCESS METHODS

All the data used by a method is static and not changed during the whole operation process. Source code, bytecode, transactions, and so on are fixed while the method begins. The focus of these methods is how to extract or build features of the dataset and adopt which algorithms to classify the data, and so are the research hotspots. This type can be divided into several subtypes based on the data that it depends mainly.

1) SOURCE CODE BASED

The source code of the contracts contains the whole logic to realize the organizer’s intent. By analyzing the source code, it is easy to distinguish the rewards distribution map. Using multi-channel textCNN and transformer [8] to improve Ponzi scheme contract detection is a good example to establish a method based on the source codes of the contracts. This method can result in a more accurate Ponzi scheme prediction. Both pros and cons of this method are too dependent on source code.

However, the methods that relied on source codes couldn’t process the new deployed contacts yet since source codes don’t be opened to extern nodes.

2) BYTECODE BASED

Almost all the static methods adopt the features of the bytecode of the contracts because the bytecode is the entity that represents the contracts completely, and contains the whole information of the contracts. For examples, Shen, X. et al. provided a method with OC-SVM and IForest algorithms to detect the Ponzi features on the bytecode [4]. This method innovatively transforms the classification of Ponzi scheme contracts into anomaly detection problems. However, the extracted features are difficult to reflect the real meaning of the bytecode. Lou, Y. et al. brought out an improved convolutional neural network with a high F-score of 0.959 [6]. But, the authors did not show a discussion of the features that they adopted in the method. Fan, S. et al. proposed an anti-leakage detection method by segmenting the opcode sequence through n-gram with 0.96 F-score [9]. This method shows that the best gram length to extract the opcode features is 2. Fan, S. et a. brought out a detection method by transforming the opcode of every contract into eigenvectors with 0.98 F-score [10]. Chen, W. et al. proposed a novel semantic-aware approach: SADPonzi [11]. This method builds a static analysis tree and identifies smart contracts according to the presetting scam model. Although this method executes the bytecode statically, it is limited to
the calling relationship of the related “functions” and does not reproduce the execution purpose of the bytecode. Peng, J. and Xiao, G. gave the detection method using opcode [15]. Within the method, the extremely randomized trees algorithm is used as the classifier and 669 opcode sequences are used as the features to detect frauds.

3) TRANSACTIONS BASED
Transactions are the operation results of a smart contract, to a certain extent, they are the agents of the contracts, so they imply the information of an SPS. For an instance, Yu, S. et al. provided a Ponzi scheme detection method based on the GCN model using two types of information: transaction amount and transaction time [12]. The number of the features, which come from transactions, adopted by the method is about 14 and all the features are real and observable. However, this method does not involve other related information such as the relationships among the transactions.

4) BYTECODE AND TRANSACTIONS BASED
Many researchers think that the information from bytecode is incomplete, it needs more information to build a good method in fact, and this information comes from the transactions of the contracts. Therefore, transactions are added to the dependent dataset. For examples, Chen, W. et al. used a boosting-based algorithm to detect smart Ponzi schemes on Ethereum [13]. This method builds account features by extracting information from transaction history and achieves high accuracy for practical use. Fan, S. et al. proposed a smart Ponzi scheme detection method using Federated Learning with 0.9655 F-score [14]. The method does not require the client to share the original training data so that it can protect the private data of all parties from leaking.

5) ACCOUNT INFORMATION ADDED
A part of researchers points out that only the contacts and their transactions could not reflect all the information to determine an SPS and account information could reveal the real intent. For instance, Wang, L. et al. proposed a method that exploits opcode and account features via oversampling-based Long Short-Term Memory (LSTM) to detect smart contracts [16]. Within this method, the oversampling technique is utilized to sample feature data and the LSTM model is trained for future Ponzi scheme detection. Liu L. et al. construct a heterogeneous graph transformer network to discover Blockchain-enabled fraud with account feature and code feature [19]. This method extracts features and constructs a heterogeneous information network, and uses the obtained node embedding to perform smart contract anomaly detection effectively and stably.

6) OTHER COMBINATIONS BASED
Each data related to smart contracts is a part of the whole scheme, and it carries more or less information. Therefore, the combination of the data varies according to certain research. For examples, the image-based scam detection method proposed by Bian, L. et al. relies on the combination of bytecode and ABI [17]. This method uses the bytecode and ABI of the contract for detection and analysis to improve upon the limitation resulting from only using the source code of the contract. However, the shortcoming of the method is that the training time required for the model is relatively long. Zhang, Y. et al. brought out an improved LightGBM algorithm to detect Ethereum Ponzi schemes exploiting the features extracted from opcodes, bytecodes, and user information [18]. This method extracts the bytecode similarity feature innovatively and combines it with the existing user transaction record and opcode frequency as a combined feature to test the Ponzi scheme.

Each combination of information data has its advantage in revealing the information of an SPS, it could not say which one is better than the others. According to the investigation of the methods on different data combinations by Ibba, G., Pierro, G. A., and Francesco, M. D., the F1-Score does not have extreme differences [5].

B. DYNAMIC PROCESS METHODS
A Ponzi scheme itself has dynamics, for some schemes are initialized as normal funds and are obliged into Ponzi schemes, the operation processes of each method of the contract realize the purpose of the creators at last. Sun and his colleagues provide a dynamic method to early detect SPSs based on behavior forest similarity [3]. By this method, an investor could inspect the contract without relying on anything else except for the smart contract itself.

IV. NEW APPROACH
There are two situations when investors want to judge whether a fund is a scam. One is that the scheme has run for a long time, at the hyper-operation or collapse stage; the other is that the scheme just begins or is at the bootstrap phase. In the first situation, it can detect the features of an SPS from the existing transactions; in contrast, In the second situation, the investors could not decide on seldom data. Therefore, how to judge whether the fund is a scam requires a comprehensive approach. When investors face the second situation they should adopt methods that mainly rely on bytecode information. And when with the first situation, they should exploit the methods that depend on transaction information merely.

A. POSTERIOR METHOD
Throughout all the attributes of an SPS, we can see that the source of the organizer’s income and the source of the investor’s rewards are the two most important characteristics, which determine whether the scheme is a Ponzi scheme, to a certain extent. If both are from investors (or new investors), it is firmly a scam. Therefore, we propose an approach that exploits this feature to detect an SPS within a scenario. This method firstly makes a basis selection whether there is a possibility to be an SPS based on checking the operation...
profile of the scheme. If so, then, build the ecosystem of the SPS. At last, determine whether it is an SPS by examining the two income sources.

1) OPERATION PROFILE
We can see that there is an exponential increase [7] in the super-operation phase. This profile is explicitly different from the normal funds for high rewards. This can be taken as the first metric to judge an SPS. So, we divide these features into three indexes to show the detailed information of this metric.

a: DAILY TRANSACTION VOLUME (DTV)
DTV is a function of time to represent the amount of all the transactions in a day. The DTV of an SPS appears as a special curve and can be divided into three parts, the first part is a linear curve segment, the second part is an exponent curve segment, and the third part is a linear curve segment.

\[
T(t) = \begin{cases} 
  at + b & t < t_0 \\
  ke^{mt} & t_0 \leq t < t_1 \\
  c - dt & t \geq t_1 
\end{cases} \quad (1)
\]

where, \(t_0\) refers to the time that the amount of transactions boosts, \(t_1\) refers to the time that the scheme begins to collapse, \(m\) refers to increase speed, and \(k\) refers to the initial constant.

We can exploit DTV to measure whether the transaction of a fund is normal and to determine the probability that the fund is an SPS.

b: DAILY INVESTORS NUMBER (DIN)
DIN is a function of time to represent the number of new investors entering the scheme, \(I(t)\). We suppose that the change of DIN depends on three causes, one is the information transmission rate, the second is the attending rate of known information, and the last one is the attractiveness of returns.

Suppose that, the information transmission rate can be calculated according to the following methods. \(A(n)\) represents the total number of those who are aware of the fund at \(n\) times transmission. And about \(x\) percent of \(A(n)\) will send this information to his friends, while, about \(y\) percent of \(\Delta A(n)\) will invest the fund. \(A_0\) is the number of persons that the fund got by advertisement.

\[
\begin{align*}
A(0) &= A_0 \\
I(0) &= y \Delta A(0) = yA_0 \\
A(1) &= A(0) + A(0)x = A_0(1 + x) \\
I(1) &= I(0) + y \Delta A(1) = yA_0(1 + x) \\
A(2) &= A(1) + A(1)x = A_0(1 + x)^2 \\
I(2) &= I(1) + y \Delta A(2) = yA_0(1 + x)^2 \\
&\vdots \\
A(n) &= A_0(1 + x)^n \\
I(n) &= yA_0(1 + x)^n
\end{align*}
\]

The transmission times, \(n\), will decrease half of before in a day. Let initial value \(n\) be \(N_0\), then \(n\) is,

\[
n(t) = \sum_{i=0}^{t} \frac{N_0 2^{-i}}{2N_0 \left[1 - 2^{-0.5}ight]} 
\]

where \(t\) is the number of days from starting. At last, we can obtain the total quantity of investors, \(I(t)\),

\[
I(t) = yA_0(1 + x)^2N_0[1 - 2^{-0.5}]
\]

DIN shows the change in the number of new investors and reflects the possibility that the fund is a scam.

c: DAILY VALUE FLOW (DVF)
DVF is a function of time to picture the amount of transaction income value in a day. We don’t provide the formal equation because DVF is so complex that it is difficult to represent simply. But, it is important to outline the scam graphically and its curve is similar to the curve of a DTV.

When an investor finds that the transaction trend of the fund appears the features of DTV, DIN, and DVF, he or she should be vigilant and careful to avoid falling into a scam while making an investment decision.

2) THE ECOSYSTEM OF A SMART PONZI SCHEME
To realize its purpose, an SPS has to establish its ecosystem, as shown in Fig. 1. From the figure, we can see three sections that marked a circled number. No.1 indicates the entities controlled by the organizer of the scheme. Within it, the master account runs for creating the contract interfaced with investing accounts, exchanges Tokens with exchangers, and transacts with partners. The contract is the address deployed as shown in Fig. 1. From the figure, we can see three sections that marked a circled number. No.1 indicates the entities controlled by the organizer of the scheme. Within it, the master account runs for creating the contract interfaced with investing accounts, exchanges Tokens with exchangers, and transacts with partners. The contract is the address deployed by the ‘fund’ to investors to perform transactions, deposits, or withdrawals. The most important are partner accounts, which are used to cover up the secret of the frauds and are the companies to implement the Ponzi scheme. No.2 shows the entities of victims, which are directive accounts and indirective investment accounts. No.3 is the Exchangers who are the places victims buy Token and the fund gets real currency.

In the figure, all the arrows represent transactions among all the entities in the ecosystem. The single-line dark red arrows represent transactions between the accounts and exchangers. The double-line green arrows give the transactions between the contract and investing accounts, and between the contract and partner accounts. The single-dot-line black arrow represents the transactions between the master account and other related accounts. The single-dot-line orange arrow gives the transactions between the contract and exchangers. At last, the three-line blue arrow shows the transactions between the master account and the contract.

We can build the ecosystem based on the transactions of the given contact according to their route paths. The first work is to list all possible exchanger accounts by searching
related websites. And then pick out different accounts from
the transactions.

a: FIND THE MASTER ACCOUNT
As long as you find the creator of the contract, the address of
the creator is the master account. It is easy to do so for there
is only one “create transaction” related to the contract.

b: FIND INVESTING ACCOUNTS AND PARTNER ACCOUNTS
Firstly, build two sets, one for the accounts that as senders
to the contract, and the other one for the accounts as a
receiver from the contract. Secondly, exclude the accounts
that could be found in the exchanger list from the two sets.
Thirdly, make a new set that contains the shared accounts
by the two sets. Fourthly, exclude the accounts that could
be found in the sender set from the receiver set, and get the
pure partner accounts. Fifthly, check the shared set to find
out all the accounts that of each one its amount of received
values is bigger than its quantity of sending values, which
can be reasonably taken as a partner account, and move these
accounts into the partner set. At last, move the rest of the
shared set into the sender set, and then get investing accounts.

c: FIND OTHER ACCOUNTS
Check the receive transactions of each one of investing
accounts, and get the sender accounts. Excluding exchanger
accounts would get the indirective victim accounts.

3) EXAMINE INCOME SOURCES
To return to the investors, the scheme must have continuous
incomes. If it can prove that these incomes come from the
investors, in other words, the scheme income and the rewards
to the investors are the same from the investors themselves,
then it concludes that the scheme must be a finance fraud
whether it is an SPS. The analysis process is shown as follows.

1. Collect all the transactions that the partner accounts
send value to the contract. And calculate the total
amount of the values sent, referred to as \( S_p \).
2. Collect all the transactions that the exchangers send
value to the contract. And calculate the total amount
of the values sent, referred to as \( S_e \).
3. Collect all the transactions that the contract sends
the value to the investors. And calculate the total amount
of the values sent, referred to as \( S_{ir} \).
4. Collect all the transactions that the investors send value
to the contract. And calculate the total amount of the
values sent, referred to as \( S_{is} \).
5. Collect all the transactions that the contract sends the
value to the externals. And calculate the total amount
of the values sent, referred to as \( S_{cs} \).
6. if \( S_p + S_e \ll S_{ir} \) and \( S_{cs} + \text{balance} < S_{is} \) then it is a scam.

B. PRIORI METHOD
Faced a newly or to be deployed contract, there are seldom
transactions, we must adopt a priori way to test the possibility
of being a fraud for the contract. Because all the income of
an SPS comes from the investors, we could establish a virtual
scenario to figure out this feature and determine whether it is
an SPS.

Give a contract, the balance, \( M \), of its owner can be
represented as,

\[
M = \sum_{i=1}^{n} (kI_i - F_i) - \sum_{j=1}^{l} (R_j + F_j)
\]  

(4)

where \( k \) is the nominal ratio for the owner income; \( I \) is the
investment of an investor; \( F \) is the transaction fee; \( R \) is the
return to an investor; \( n \) is the amount of the transactions; \( l \) is
the number of investors that got returns.

If we make each investment much larger than every
possible transaction fee and be identical, we can get,

\[
M \approx nkI - \sum_{j=1}^{l} R_j
\]

(5)

We can use formula (5) to test whether the contract is an
SPS. If the input data and the balance of the owner fit into
this formula and there are not any other functions or methods
that can send value to the contract, we firmly think that it is a
fraud.

The following process is summarized to realize the priori
method which exploits the dynamical behaviors of the
contract.

1. Build a private operation network.
2. Translate the ABI into a public function list, find all the
send methods and get the value functions.
3. Deploy the contract in a private network, and give a list
of addresses with a fixed balance.
4. Run one of the send methods for each address with a
fixed value, and record the value and the balance.
TABLE 1. Transaction values.

| Date | $S_p$ | $S_t$ | $S_r$ | $S_{cs}$ | balance | $S_p+S_{se}$ | $S_{cs}+balance$ |
|------|-------|-------|-------|----------|---------|-------------|-----------------|
| 1    | 0     | 0     | 0.048 | 0        | 0.192   | 0           | 0.192           |
| 2    | 0     | 0     | 1.51706 | 7.5678 | 0       | 6.05074    | 6.05074         |
| 3    | 0     | 0     | 2.773011 | 13.84654 | 0 | 11.07353 | 11.07353        |
| 4    | 0     | 0     | 4.113017 | 20.09157 | 0       | 15.97855  | 15.97855       |
| 5    | 0     | 0     | 6.911091 | 31.1556 | 0       | 24.24451  | 24.24451        |
| 6    | 0     | 0     | 10.09329 | 44.30439 | 0 | 34.2111  | 34.2111         |
| 7    | 0     | 0     | 12.76333 | 55.66839 | 0       | 42.90506  | 42.90506        |
| 8    | 0     | 0     | 23.26018 | 100.3764 | 0       | 77.1121   | 77.1121         |
| 9    | 0     | 0     | 32.29946 | 132.7892 | 0       | 100.4997  | 100.4997        |
| 10   | 0     | 0     | 50.99363 | 210.9226 | 0       | 159.939  | 159.939         |
| 11   | 0     | 0     | 61.78189 | 240.6687 | 0       | 178.8968  | 178.8968        |
| 12   | 0     | 0     | 74.73663 | 277.2757 | 0       | 202.5491  | 202.5491        |
| 13   | 0     | 0     | 97.15867 | 354.9578 | 0       | 257.8091  | 257.8091        |
| 14   | 0     | 0     | 114.6476 | 393.5422 | 0       | 278.9045  | 278.9045        |
| 15   | 0     | 0     | 134.1946 | 442.3184 | 0       | 308.1338  | 308.1338        |
| 16   | 0     | 0     | 154.7431 | 492.6322 | 0       | 337.8991  | 337.8991        |
| 17   | 0     | 0     | 183.183  | 545.812  | 0       | 362.639   | 362.639         |
| 18   | 0     | 0     | 198.416  | 557.4897 | 0       | 359.0837  | 359.0837        |
| 19   | 0     | 0     | 221.3292 | 592.2552 | 0       | 370.936   | 370.936         |
| 20   | 0     | 0     | 221.3292 | 592.9306 | 371.6102  | 0.000713  | 371.6109        |
| 21   | 0     | 0     | 221.3292 | 593.3102 | 371.6102  | 0.000713  | 371.6109        |

5. Analyze the result according to formula (5), if the result does not fit the formula, go to step 7, it says that the contract has a little possibility to be an SPS.
6. Repeat step 4 until all the send methods are performed. Here you could think that the contract is much more possible a scam.
7. End

V. EXPERIMENTS
The purpose of these experiments is to verify the feasibility of the proposed method, provide users with demonstrative examples, and compare it with some existing methods. We selected the current popular dataset as test objects of the new approach, which comes from cn.therscan.com. The dataset contains 50 contracts that were labeled as Ponzi schemes.

A. DEMONSTRATIVE EXPERIMENTS
Firstly, we chose InfinitiMoney (address @ 0 × 53.6a6ba0d 913d5d6a4ce2c6eb7ed0de3c0f0b89e) as the analysis object to verify the posterior method. We got a total of 3231 transaction original records from cn.therscan.com and obtained 2586 successful transactions after excluding the errors. The first transaction is the create of the contract, so we get the master account @0xb648cc3367571c9166e966d1-eb97de3f9ebf8fd. The rest of all the transactions are sent to the contract. We pictured the number of transactions and their values in Fig. 2. The horizontal axis corresponds to the days of transactions and the left vertical axis refers to the number of transactions, and the right vertical axis represents the values of transactions in Fig. 2. We exploit the $DTV$ index to explain the curve of the quantity of the transactions. It explicitly has the same tendency as the $DTV$ index. In addition, the transaction values, as the $DVF$ index, has complex changes than the $DTV$ index, but it clearly shows that it has three phrases. So that we can think that the scheme is a fraud.

We don’t show the process that how to build the ecosystem of InfinitiMoney since it is relatively easy and verbose.

We adopted the analysis process mentioned in part 3) of subsection A in section IV to collect, calculate and arrange the data into Table 1. We can find that $S_p + S_{se} \ll S_{ir}$ and $S_{cs} + \text{balance} < S_{ir}$, so we regard reasonably InfinitiMoney as a scam.
Secondly, we select Wallie.me (address @0x0b52b760-55c392d679322ae7737cdb6d42133) as the object to test the priori method. We got the bytecode and ABI from cn.therscan.com and deployed them into our private aleth-PM network. We took 2999 addresses as the investment accounts each of which was allocated 0×8000000000000000 as its balance. Each account invested one time with value, 0×2000000000, and made a one-time withdrawal. We got raw data of all the accounts as shown in Table 2. In Table 2, OB refers to the balance of the contract and IB represents the balance of an investment account. From Table 2, we can see that the OB increases with IB according to a certain rule. To show this rule, we rearrange the data into Table 3.

### Table 2. Account balance.

| No | Address | Before investment | After investment | After withdraw |
|----|---------|------------------|-----------------|---------------|
|    |         | OB               | IB              | IB            |
| 1  | 0x03065c5ba4e24325580.2505ba937123e98009b | 0 | 92237320384577004 | 922373189445158 |
| 2  | 0x009b258392c0.917e0955f6762e6b707c3 | 133205242711 | 92237320384577004 | 922373189445158 |
| 3  | 0x0476457b64e5cha5b5e4057e81c519613b4 | 92237320384577004 | 276644377190 | 922373189445158 |
| 4  | 0x0606dcad9c1e23a518c0e12582a8af4cc | 399232222961 | 92237320384577004 | 922373189445158 |
| 5  | 0x009b258392c0.917e0955f6762e6b707c3 | 133205242711 | 92237320384577004 | 922373189445158 |
| 6  | 0x0476457b64e5cha5b5e4057e81c519613b4 | 92237320384577004 | 276644377190 | 922373189445158 |
| 7  | 0x0606dcad9c1e23a518c0e12582a8af4cc | 399232222961 | 92237320384577004 | 922373189445158 |
| 8  | 0x009b258392c0.917e0955f6762e6b707c3 | 133205242711 | 92237320384577004 | 922373189445158 |

### Table 3. Analysis data.

| No | Address | OB | AR | TI | OB+AR | TI-(OB+AR) |
|----|---------|----|----|----|-------|------------|
| 1  | 0x03065c5ba4e24325580.2505ba937123e98009b | 0 | 423350368 | 173438953472 | 137438954086 |
| 2  | 0x009b258392c0.917e0955f6762e6b707c3 | 133205242711 | 8657250304 | 274877906944 | 274877906875 |
| 3  | 0x0476457b64e5cha5b5e4057e81c519613b4 | 92237320384577004 | 13084329984 | 412316860416 | 41231682945 |
| 4  | 0x0606dcad9c1e23a518c0e12582a8af4cc | 399232222961 | 1570609504 | 549755813888 | 549755801085 |
| 5  | 0x009b258392c0.917e0955f6762e6b707c3 | 133205242711 | 532249091581 | 16130472730 | 159951351298 |
| 6  | 0x0476457b64e5cha5b5e4057e81c519613b4 | 92237320384577004 | 22314975070 | 687149767360 | 687149740830 |
| 7  | 0x0606dcad9c1e23a518c0e12582a8af4cc | 399232222961 | 26622329280 | 824633720832 | 824633679774 |
| 8  | 0x009b258392c0.917e0955f6762e6b707c3 | 133205242711 | 798410385171 | 30516659089 | 962072654240 | 962072652246 |
| 9  | 0x03065c5ba4e24325580.2505ba937123e98009b | 0 | 423350368 | 173438953472 | 137438954086 |
| 10 | 0x009b258392c0.917e0955f6762e6b707c3 | 133205242711 | 8657250304 | 274877906944 | 274877906875 |
| 11 | 0x0476457b64e5cha5b5e4057e81c519613b4 | 92237320384577004 | 13084329984 | 412316860416 | 41231682945 |
| 12 | 0x0606dcad9c1e23a518c0e12582a8af4cc | 399232222961 | 1570609504 | 549755813888 | 549755801085 |
| 13 | 0x009b258392c0.917e0955f6762e6b707c3 | 133205242711 | 532249091581 | 16130472730 | 159951351298 |
| 14 | 0x0476457b64e5cha5b5e4057e81c519613b4 | 92237320384577004 | 22314975070 | 687149767360 | 687149740830 |
| 15 | 0x0606dcad9c1e23a518c0e12582a8af4cc | 399232222961 | 26622329280 | 824633720832 | 824633679774 |

VOLUME 10, 2022
see that all the differences are small and accumulated slowly. In fact, these differences are the transaction fees. It also shows that the contract income is all from the investors for the balance of the contract is nearly equal to the investments plus all the returns. So we can determine that Wallie.me is a scam.

By the above two examples, we explained and verified the test process of our proposed approach. Then, we tested all other contracts with a large number of transactions in the dataset, which shows that our approach is feasible and practical.

### B. COMPARATIVE EXPERIMENTS

To further demonstrate the effectiveness of the proposed method, we selected four methods for comparison. The first is textCNN [8] which uses the feature information contained in the source code of smart contracts to classify the contracts with a neural network algorithm and is a static method. The second is SADPonzi [11] which exploits bytecode features, this method uses the inherent relationship of the bytecode to build a static analysis tree and identifies smart contracts according to the presetting scam model. Although this method executes the bytecode statically, it is limited to the calling relationship of the related “functions” and does not reproduce the execution purpose of the bytecode. It is still a static method, but it is a step towards a dynamic method. The third method is the improved LightGBM [18] with Smote + Tomek, which further expands the source of features. In addition to using the features of source code and bytecode, 7 user features are extracted and added, which greatly extend the range of features and belong to static methods. Among the added user features, Paid_rate and Kr could be used as training features of the detection model. The fourth method is S_HGTNs [19] which further expands the range of features, including account, transaction, and other information, but does not contain bytecode features. The authors exploit these features to construct a heterogeneous information network so that they can obtain some meta paths as the input of the convolution network to perform the classification process. And again, it is a static method.

#### 1) EXPERIMENTAL DATA AND RESULTS

The comparative dataset consists of 50 contracts labeled SPS and 150 normal contracts. Among them, smart contracts without source code will be converted into equivalent source code by the decompile tools. 25 SPS contracts and 50 normal contracts were selected as the training set. The classification test data is the entire comparison dataset. The classification results, calculated by F1-score, are shown in Table 4, where the proposed method has not a training phase.

#### 2) ANALYSIS OF COMPARATIVE RESULTS

It can be seen from Table 4 that the experimental results of method 1 are the worst, with relatively low precision and recall rate, which shows that it is impossible to accurately determine whether a contract is a scam only against the source code features. Since the source code is the text that represents the contract functionality, the text itself is difficult to fully represent the essential characteristics of the contract’s functionality, so the three indicators are all low. Because method 2 adopts the calling structure features of the bytecode, which reflects the functionality of the smart contract, the experimental results are improved compared with those methods depending on the source code, especially the precision is the highest among all methods. However, some smart contracts with the same structural features are not all SPSs. Method 3 further improves the recognition results due to the adoption of more feature combinations, with the highest F1-score. Method 4 extracts a wider range of features, however, due to ignoring the features of the bytecode, the experimental results are not as ideal as method 3, the recall rate is low, and the results indicate that the features of the bytecode play an important role in judging an SPS. The experimental results of the proposed method are roughly equivalent to the results of method 4. The precision ranks second among the five methods, the recall rate ranks third, and F1-score ranks second. From the result, the proposed method has the same effectiveness as method 3 and method 4 in calculating the test set.

Although the four methods are different in the specific process and method of feature selection and extraction, they are all classification methods based on machine learning algorithms, and there is a learning phase that requires certain samples for training.

The proposed method reflects the essential characteristics and network characteristics of SPSs, and it is convenient for investors to operate, and has the following advantages.

1. It does not require a learning and training procedure before real use, can directly detect smart contracts, and is convenient for users to master.
2. It is highly adaptable and can deal with variants of the scam without other processing. Methods based on the training set need to be retrained in the face of new changes.
3. The curves provided by the three indicators, DTV, DIN, and DVF, can intuitively allow investors to screen whether the smart contract is a scam, which is simple and easy to use.

### VI. CONCLUSION

The occurrence of SPSs is originated from the vulnerability of the supervision mechanism on a blockchain. In other words, there is not a safety system to ensure vicious addresses or contracts are not deployed. We propose an approach to test...
whether a contract is a fraud. This approach could deal with two situations, the contracts to be deployed and the long-run contracts respectively. Of the new approach, the priori method can be exploited to test the risk of being an SPS of a contract for the runner of a blockchain, and the posterior method can protect an investor from being trapped in a scam. At the same time, the posterior method can also be extended to monitor some contracts dynamically so that to alert users with the probability to be fallen into smart Ponzi schemes.

Although the three indicators, $DTV$, $DIN$, and $DVF$ are convenient for investors to observe the operation status of the fund, the information of their curves cannot be fully automated, and there is still a certain gap in deployment as the fund, the information of their curves cannot be fully convenient for investors to observe the operation status of the fund, and at the same time develop the method into an automated application service plug-in that is convenient for research, and at the same time develop the method into an automated daemon service program. In addition, we have not given a formal model of $DVF$. To this end, we will further improve various indicator models and related algorithms in future research, and at the same time develop the method into an automated application service plug-in that is convenient for investors to use.

REFERENCES

[1] M. Bartoletti, S. Carta, T. Cimoli, and R. Saia, “Dissecting Ponzi schemes on Ethereum: Identification, analysis, and impact,” Future Gener. Comput. Syst., vol. 102, pp. 249–277, Jan. 2020.

[2] W. Chen, Z. Zheng, J. Cui, E. Ngai, P. Zheng, and Y. Zhou, “Detecting Ponzi schemes on Ethereum: Towards healthier blockchain technology,” in Proc. World Wide Web Conf. (WWW), Lyon, France, 2018, pp. 1409–1418.

[3] W. Sun, G. Xu, Z. Yang, and Z. Chen, “Early detection of smart Ponzi scheme contracts based on behavior forest similarity,” in Proc. IEEE 20th Int. Conf. Softw. Quality, Rel. Secur. (QRS), Macau, China, Dec. 2020, pp. 297–309.

[4] X. Shen, S. Jiang, and L. Zhan, “Mining bytecode features of smart contracts to detect Ponzi scheme on blockchain,” Comput. Model. Eng. Sci., vol. 127, no. 3, pp. 1069–1085, 2021.

[5] G. Ibba, G. A. Pierro, and M. Di Francesco, “Evaluating machine-learning techniques for detecting smart Ponzi schemes,” in Proc. IEEE/ACM 4th Int. Workshop Emerg. Trends Softw. Eng. Blockchain (WETSEB), Madrid, Spain, May 2021, pp. 34–40.

[6] Y. Lou, Y. Zhang, and S. Chen, “Ponzi contracts detection based on improved convolutional neural network,” in Proc. IEEE Int. Conf. Services Comput. (SCC), Beijing, China, Nov. 2020, pp. 353–360.

[7] Y. Boshmaf, C. Elvitigala, H. A. Jawaheri, P. Wijesekera, and M. A. Sabah, “Investigating MMM Ponzi scheme on bitcoin,” in Proc. 15th ACM Asia Conf. Comput. Commun. Secur., New York, NY, USA, Oct. 2020, pp. 519–530.

[8] Y. Chen, H. Dai, X. Yu, W. Hu, Z. Xie, and C. Tan, “Improving Ponzi scheme contract detection using multi-channel TextCNN and transformer,” Sensors, vol. 21, no. 19, pp. 1–20, Sep. 2021, doi: 10.3390/s21196417.

[9] S. Fan, S. Fu, H. Xu, and X. Cheng, “AI-SPSD: Anti-leakage smart Ponzi schemes detection in blockchain,” Inf. Process. Manage., vol. 58, no. 4, pp. 1–13, Jul. 2021, doi: 10.1016/j.ipm.2021.102587.

[10] S. Fan, S. Fu, H. Xu, and C. Zhu, “Expose your mask: Smart Ponzi schemes detection on blockchain,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2020, pp. 1–7, doi: 10.1109/IJCNN48605.2020.9207143.

[11] W. Chen, X. Li, Y. Sui, N. He, H. Wang, L. Wu, and X. Luo, “SADPonzi: Detecting and characterizing Ponzi schemes in Ethereum smart contracts,” in Proc. ACM Meas. Anal. Comput. Syst., May 2021, pp. 1–30, doi: 10.1145/3410220.3460105.

[12] S. Yu, J. Jin, Y. Xie, J. Shen, and Q. Xuan, “Ponzi scheme detection in Ethereum transaction network,” 2021, arXiv:2104.08456.

[13] W. Chen, Z. Zheng, E. C.-H. Ngai, P. Zheng, and Y. Zhou, “Exploiting blockchain data to detect smart Ponzi schemes on Ethereum,” IEEE Access, vol. 7, pp. 37575–37586, 2019.

[14] S. Fan, H. Xu, S. Fu, and M. Xu, “Smart Ponzi scheme detection using federated learning,” in Proc. IEEE 22nd Int. Conf. High Perform. Comput. Commun., IEEE 18th Int. Conf. Smart City, IEEE 6th Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS), Yauca Island, Fiji, Dec. 2020, pp. 881–888.

[15] J. Peng and G. Xiao, “Detection of smart Ponzi schemes using opcode,” Commun. Comput. Inf. Sci., vol. 1267, pp. 192–204, Nov. 2020.

[16] L. Wang, H. Cheng, Z. Zheng, A. Yang, and X. Zhu, “Ponzi scheme detection via oversampling-based long short-term memory for smart contracts,” Knowl.-Based Syst., vol. 228, no. 9, pp. 1–12, Jul. 2021.

[17] L. Bian, L. Zhang, K. Zhao, H. Wang, and S. Gong, “Image-based scam detection method using an attention capsule network,” IEEE Access, vol. 9, pp. 33654–33665, 2021.

[18] Y. Zhang, W. Yu, Z. Li, S. Raza, and H. Cao, “Detecting Ethereum Ponzi schemes based on improved LightGBM algorithm,” IEEE Trans. Computat. Social Syst., vol. 9, no. 2, pp. 624–637, Apr. 2022, doi: 10.1109/tcss.2021.3088145.

[19] L. Liu, W.-T. Tsai, M. Z. A. Bhuiyan, H. Peng, and M. Liu, “Blockchain-enabled fraud discovery through abnormal smart contract detection on Ethereum,” Future Gener. Comput. Syst., vol. 128, pp. 158–166, Mar. 2022, doi: 10.1016/j.future.2021.08.023.

LINGLING SONG was born in Binzhou, Shandong, China, in 1974. She received the master’s degree from the China University of Petroleum, China. She is currently an Associate Professor with the Department of Information Engineering, Binzhou Polytechnic, China. Her research interests include data mining, big data analysis, computational intelligence, network technology, and information security.

XIAOHAN KONG was born in Binzhou, Shandong, China, in 2001. He is currently pursuing the bachelor’s degree with the Haide College, Ocean University of China. His research interests include economics and informatics.