Received May 4, 2019, accepted May 20, 2019, date of publication June 7, 2019, date of current version June 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2921555

Securing Data With Blockchain and AI

KAI WANG[1,2], (Member, IEEE), JIAQING DONG[1], YING WANG[3], AND HAO YIN[1], (Member, IEEE)

[1]Research Institute of Information Technology, Tsinghua University, Beijing 100084, China
[2]School of Computer and Control Engineering, Yantai University, Shandong 264005, China
[3]School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China

Corresponding author: Hao Yin (h-yin@mail.tsinghua.edu.cn)

This work was supported in part by the China Postdoctoral Science Foundation, under Grant 2017M620786, Shandong Provincial Natural Science Foundation, China, under Grant ZR2017BF018, and National Natural Science Foundation of China (NSFC), under Grant 61702439.

ABSTRACT Data is the input for various artificial intelligence (AI) algorithms to mine valuable features, yet data in Internet is scattered everywhere and controlled by different stakeholders who cannot believe in each other, and usage of the data in complex cyberspace is difficult to authorize or to validate. As a result, it is very difficult to enable data sharing in cyberspace for the real big data, as well as a real powerful AI. In this paper, we propose the ***SecNet***, an architecture that can enable secure data storing, computing, and sharing in the large-scale Internet environment, aiming at a more secure cyberspace with real big data and thus enhanced AI with plenty of data source, by integrating three key components: 1) blockchain-based data sharing with ownership guarantee, which enables trusted data sharing in the large-scale environment to form real big data; 2) AI-based secure computing platform to produce more intelligent security rules, which helps to construct a more trusted cyberspace; 3) trusted value-exchange mechanism for purchasing security service, providing a way for participants to gain economic rewards when giving out their data or service, which promotes the data sharing and thus achieves better performance of AI. Moreover, we discuss the typical use scenario of SecNet as well as its potentially alternative way to deploy, as well as analyze its effectiveness from the aspect of network security and economic revenue.

INDEX TERMS Data security, data systems, artificial intelligence, cyberspace.

I. INTRODUCTION

With the development of information technologies, the trend of integrating cyber, physical and social (CPS) systems to a highly unified information society, rather than just a digital Internet, is becoming increasing obvious [1]. In such an information society, data is the asset of its owner, and its usage should be under the full control of its owner, although this is not the common case [2], [3].

Given data is undoubtedly the oil of the information society, almost every big company want to collect data as much as possible, for their future competitiveness [4], [5]. An increasing amount of personal data, including location information, web-searching behavior, user calls, user preference, is being silently collected by the built-in sensors inside the products from those big companies, which brings in huge risk on privacy leakage of data owners [6], [7]. Moreover, the usage of those data is out of control of their owners, since currently there is not a reliable way to record how the data is used and by who, and thus has little methods to trace or punish the violators who abuse those data [8]. That is, lack of ability to effectively manage data makes it very difficult for an individual to control the potential risks associated with the collected data [9]. For example, once the data has been collected by a third party (e.g., a big company), the lack of access to this data hinders an individual to understand or manage the risks related to the collected data from him. Meanwhile, the lack of immutable recording for the usage of data increases the risks to abuse them [10].

If there is an efficient and trusted way to collect and merge the data scattered across the whole CPS to form real big data, the performance of artificial intelligence (AI) will be significantly improved since AI can handle massive amount of data including huge information at the same time, which would bring in great benefits (e.g., achieving enhanced security for data) and even makes AI gaining the ability to exceed human capabilities in more areas [11]. According to the research in [12], if given large amount of data in

The associate editor coordinating the review of this manuscript and approving it for publication was Chi-Yuan Chen.

an orders of magnitude more scale, even the simplest AI algorithm currently (e.g., perceptrons from the 1950s) can achieve fanciest performance to beat many state-of-the-art technologies today. The key lies in how to make data sharing trusted and secured [13]. Fortunately, the blockchain technologies may be the promising way to achieve this goal, via consensus mechanisms throughout the network to guarantee data sharing in a tamper-proof way embedded with economic incentives [14], [15]. Thus, AI can be further empowered by blockchain-protected data sharing [16]–[18]. As a result, enhanced AI can provide better performance and security for data.

In this paper, we aim at securing data by combining blockchain and AI together, and design a **Sec**ure **Net**working architecture (termed as ***SecNet***) to significantly improve the security of data sharing, and then the security of the whole network, even the whole CPS.

In SecNet, to protect data, one of the biggest challenges is where and how to store data, because users have to give their data to service providers if they want to use certain services or applications [1], [3]. This is caused by the inherent coupling of user data and application in current service mechanisms, which significantly hinders the development of data protection and application innovation. Inspired by the concept of Personal Data Store (PDS) from openPDS [5] and the Private Data Center (PDC) from HyperNet [1], *SecNet* finally inherits and adopts PDC instead of PDS, as PDC is more suitable to deploy and to deal with this problem, since it provides more secure and intelligent data storage system via physical entities instead of software-based algorithms as in openPDS. Each PDC actually serves as a secured as well as centralized physical space for each SecNet user where his/her data lives in. Embedding PDC into SecNet would allow users to monitor and reason about what and why their data is used as well as by who, meaning the users can truly control every operation on their own data and achieve fine-grained management on access behaviors for data. Actually, besides PDC, other choices can also be applied for the data storing in SecNet according to certain requirements (see Section V).

The trust-less relationship between different data stakeholders significantly thwarts the data sharing in the whole Internet, thus the data used for AI training or analyzing is limited in amount as well as partial in variety. Fortunately, the rise of Blockchain technologies bring in a hopeful, efficient and effective way to enable trust data sharing in trust-less environment, which can help AI make more accurate decisions due to the real big data collected from more places in the Internet. SecNet leverages the emerging blockchain technologies to prevent the abuse of data, and to enable trusted data sharing in trust-less or even untrusted environment. For instance, it can enable cooperations between different edge computing paradigms to work together to improve the whole system performance of edge networks [19]. The reason why blockchain can enable trusted mechanisms is that it can provide a transparent, tamper-proof metadata infrastructure to seriously recode all the usage of data [17]. Thus, SecNet introduces blockchain-based data sharing mechanisms with ownership guarantee, where any data ready for sharing should be registered into a blockchain, named Data Recording Blockchain (DRB), to announce its availability for sharing. Each access behavior on data by other parties (not the data owner) should also be validated and recorded in this chain. In addition, the authenticity and integrity of data can only be validated by DRB as well. Besides, SecNet enables economic incentive between different entities if they share data or exchange security service, by embedding smart contract on data to trigger automatic and tamper-proof value exchange. In this way, SecNet guarantees the data security and encourages data sharing throughout the CPS.

Furthermore, data is the fuel of AI [11], and it can greatly help to improve the performance of AI algorithms if data can be efficiently networked and properly fused. Enabling data sharing across multiple service providers can be a way to maximize the utilization of scattered data in separate entities with potential conflicts of interest, which can enables a more powerful AI. Given enough data and blockchain-based smart contract [20] on secure data sharing, it is not surprised that AI can become one of the most powerful technologies and tools to improve cybersecurity, since it can check huge amount of data more quickly to save time, and identify and mitigate threats more rapidly, and meanwhile give more accurate prediction and decision support on security rules that a PDC should deploy. Besides, embedded with Machine Learning [21] inside, AI can constantly learn patterns by applying existing data or artificial data generated by GAN [22] to improve its strategies over time, to strengthen its ability on identifying any deviation on data or behaviors on a 24/7/365 basis. SecNet can apply these advanced AI technologies into its Operation Support System (OSS) to adaptively identify more suspicious data-related behaviors, even they are never seen before. In addition, swarm intelligence can be used in SecNet to further improve the data security, by collecting different security knowledge from huge amount of intelligent agents scattered everywhere in the CPS, with the help of trusted exchange mechanisms for incentive tokens [23].

The rest of this paper is organized as follows. Section II overviews related works. Section III presents the SecNet architecture. Section IV gives a typical use scenario of SecNet on medical care area. Section V discusses an alternative way to deploy a different data storage model in SecNet. Section VI provides the analysis on both security improvement of the network system and the incentive for users to share learned security rules. Finally, section VII concludes this paper and gives some future directions.

II. RELATED WORK

Data security is among key concerns of any network architectures, and is the base for AI algorithms to improve due to its requirement for huge amount of data from as much as possible places in Internet. Meanwhile, with a more powerful AI, data security can be further protected at a higher level as an enhanced AI can figure out advanced and complicated threats more easily than normal AI.

To enhance *the security of data* in CPS, numbers of efforts are conducted. The work in [3] presents an architecture named Amber to enable decoupling data from the web applications, which gives control ability to web users over their personal data, as well as provides a powerful web-wide query function to search personal data. To extend the decoupling mechanism of data and applications from only web services to all kinds of applications, the research group from the Media Lab in Massachusetts Institute of Technology designs the openPDS [5], acting as a secured virtual space for users to collect, store and manage their data, separating all kinds of applications from operating on data directly. In addition, openPDS introduces a new service paradigm named SafeAnswer, to dynamically protect data privacy by reducing the dimensions of personal data.

Besides, the emerging blockchain technology provides an efficient and effect way to guarantee *the security of data* in CPS, by providing tamper-proof and traceable recording features as well as incentive mechanisms. The authors in [8] develop the OriginChain system to realize the transparency and tamper-proof features of the metadata when the supply chain traces products. OriginChain enables all related parties to obtain the same trusted data and adapt to dynamic environment and regulations. The authors in [10] propose a blockchain-based MeDShare system to effectively manage and protect medical records, as well as share medical data among cloud repositories, with guarantees on data provenance, auditing and controlling. The work in [17] overviews the background of blockchain and Intrusion Detection System (IDS) in details, and discusses how to apply blockchain technologies to IDS, as well as gives reasonable guesses about possible hidden dangers in this direction. Besides, the work in [15] designs a blockchain-based incentive mechanism for crowdsensing applications, with privacy preserving and data security guaranteeing.

Furthermore, AI is also a promising way to enhance *data security* in CPS, since it can deeply analyze huge amount of data, learn hidden patterns and then make accurate predictions, with the help of availability of enormous data and increased computational power. The work in [11] has made a detailed overview about the use of AI for big data as well as the use of big data for AI, and also put forward some development directions including how to improve the data security by AI. The work in [16] highlights AI can gain better performance if provided huge amount of data to achieve a better base model, and appeals to develop more efforts for building larger valuable datasets, to empower the AI for better security of data. Furthermore, the work in [21] overviews and presents a comprehensive survey on AI methods for cyber security. In addition, the work in [20] aims at creating a market where participants can exchange machine learning modes for rewards, making AI more practical and accessible to everyone, and thus providing more AI solutions for better security of data.

All these ideas and solutions above propose to protect data security, by designing a new service paradigm supporting the decoupling of data and application, or by designing a specific blockchain to meet demands of certain applications, or by integrating AI algorithms as a functional component to analyze data security. However, none of them treats the problem of data security from the view of architecture. To fill this gap, SecNet tries to construct a common and general networking architecture by combining the power of AI and blockchain together at a large scale, which can support dynamic update of all these functional component separately at any time as needed, to efficiently and effectively improve the data security for all applications.

It is worth noting that SecNet is different from HyperNet [1]. For instance, firstly, AI in HyperNet mainly acts as the virtual personal assistant to protect privacy of a single PDC user while AI in SecNet is also in charge of generating artificial data for training more robust security rules, which can be used to enhance AI again. Secondly, how to securely sharing security rules with the help of a detailed on-chain smart contract is given in SecNet, yet HyperNet lacks. In addition, SecNet aims at achieving a more secure cyberspace by sharing not only user data but also security rules produced by AI, while HyperNet only aims at securely sharing user data. Last but not least, PDC is only one of the data storing solutions for SecNet (see Section V), yet is the only solution for HyperNet.

III. THE SECNET ARCHITECTURE

SecNet is build as an architecture for a more secure cyberspace, by integrating three key components: 1) blockchain-based data sharing with ownership guarantee; 2) AI-based secure computing platform based on big data to produce intelligent and dynamic security rules; 3) trust value-exchange mechanism for purchasing security services.

Figure 1 illustrates the overall architecture of SecNet. Nodes in SecNet are connected with Blockchain-based Networking. In the network, nodes communicate with each other and reach a consensus based on blockchain techniques. In the meanwhile, they cooperate through the execution of smart contracts. In order to reach a consensus, either on node state or smart-contract execution results, each node contains a *blockchain ledger* to sync state with other nodes. In terms of data, SecNet nodes are equipped with the *data storage* module and *access control* module for data security. SecNet nodes also have an *Operation Support System (OSS)* module which enables AI-based secure computing (ASC) for generating knowledge and secure rules from data.

A. DATA SHARING GUARANTEED BY BLOCKCHAIN

For data protection, SecNet adopts the Private Data Center (PDC) from HyperNet [1], and integrates blockchain-based protection mechanism for data sharing between untrusted entities.

PDC provides physical security for data, leveraging advanced architectural and engineering approaches to

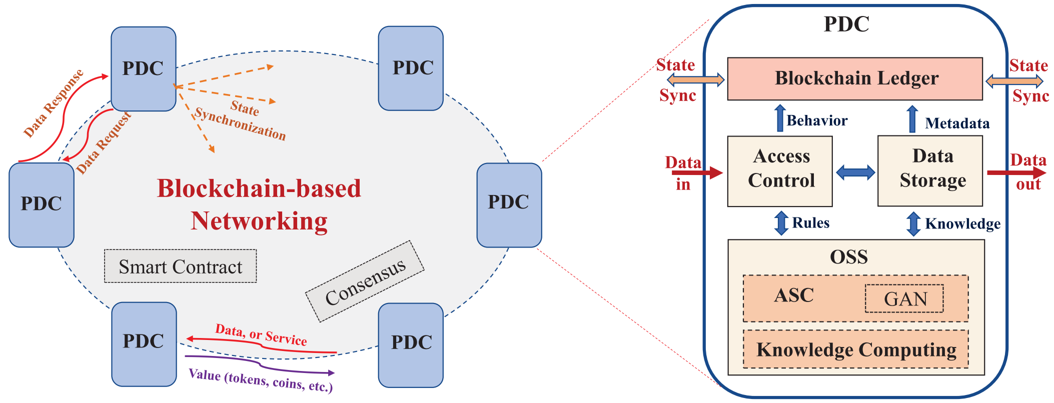

FIGURE 1. The SecNet architecture.

operating AI-based OSS. One important feature PDC provides is the uniform data access control. Uniform data access control comes from two aspects. The first one is uniform data representation (UDR). UDR helps data be represented in a standard form, in which data is self-description and can be easily parsed by applications conforming UDR standard, which makes it convenient for data sharing among entities. With UDR, various kinds of data will have a uniform representation to data consumers, which naturally mitigates data format problem in an environment where different applications have different data formats. The second one is uniform access control (UAC). UAC is very similar to access control schemes used in many file systems. It's concerned with giving access to agents (users, groups, applications and more) to perform various kinds of operations (read, write, append, etc) on data. PDC can easily decide whether a request for a data from a specific entity is legal or not with UAC. Besides the representation aspects, PDC also provides a mechanism for data identification. PDC also provides a uniform data identifier (UDI) platform for data identification and routing. With UDI, PDC is capable of identifying the source, version, ownership and many other attributes of data, makes it possible to manage and exchange data objects between different entities and applications. The UDI platform in PDC is decentralized, with which user data could be managed in a decentralized way and no service provider can control data, thus the abuse of data or data leakage is avoided.

Every PDC is housed in nondescript facilities and the physical access is strictly controlled both at the perimeter and at building ingress points by professional security staff as well as GAN-based improved rules, only providing data access for legitimate users who have such privileges.

Every entity (e.g., a user, or an institute) has a PDC to store data. All the data produced in cyberspace related to an entity is stored in a corresponding PDC, and can be merged and computed to form a knowledge system, to further improve the data security.

Before any data can be shared by Internet, this data should be registered into the DRB, to announce its availability for sharing. DRB is in charge of not only data naming, but also data validating and behavior recording of data interaction. That is, any interaction with this data would be recorded by DRB, and the authenticity and integrity of data can only be validated by DRB as well.

B. AI-BASED SECURE COMPUTING

Data is so important for its owner, and different types of data can be produced by reshaping the raw data, according to different requirements and scenarios. For example, the health information of a user stored in PDC can be extracted and reorganized to become structured medical data which is very convenient for its buyers from hospital, research institutes and heathy application developers.

All the data of an entity in cyberspace is stored in PDC, and thus its security is of great importance to its owner, as the data is in fact the digital clone of the entity in real world. To protect data, SecNet introduces ASC component into the OSS in every PDC.

AI is one of the core capabilities integrated in PDC. Various kinds of machine learning techniques have been invented for different AIs, for instance pattern matching, computer vision and self-driving. Currently there are different AI techniques being investigated for handling different data types. These data-specific AI functions can be treated as a large set of "solution islands": the academia and industry has produced numerous isolated software components and mechanisms that deal with various parts of intelligence separately. PDC works as an AI operation platform, integrating individual AI components into a coherent, intelligent system of a broader nature. Different AI functionalities collaborate with each other in PDC and act as an intelligent system.

For secure computing, at the very beginning stage, ASC can integrate Generative Adversarial Network (GAN) [22] module to generate more powerful and evolving security rules, and enable a secure and intelligent OSS for PDC.

GAN module of ASC can learn the current security rules of a PDC, and then generative malicious but "look like legitimate" access requests for some private data to confuse and confound the OSS of a PDC, aiming at making OSS lose the ability to classify the access request is illegal or not.

After extensive round of generating and classifying by GAN module, the OSS of PDC would become much more intelligent and powerful, and fake access requests for data would have little chance to compete such a secure and intelligent OSS of this PDC.

Different entities can share their computation results with each other, which is protected by blockchain, to achieve higher performance and lower energy consumption.

C. VALUE-EXCHANGE FOR SECURITY SERVICES

Except for the security concerns from every single PDC, the Internet has its own threats. For instance, various cyber attacks and computer viruses move across the Internet, and they are evolving all the time, which makes the protection from the view of each single PDC insufficient. To further improve the security of data in cyberspace, the fragmented data scattered across the Internet should be combined to produce more useful security strategies and more intelligent security rules.

Algorithm 1 Smart Contract on Data

```
Require:
    mapping (hash => struct) public Data;
    mapping (pubkey => int) public balance;
    mapping (address => hash) private reverseIndex;
    procedure RegisterData(hash, description, address, price)
        Data[hash].owner ← msg.sender
        Data[hash].address ← address
        Data[hash].description ← description
        Data[hash].price ← price
        Data[hash].subscribers ← []
        reverseIndex[address] ← hash
        return TRUE
    end procedure
    procedure WithdrawData(hash)
        require (Data[hash].owner == msg.sender)
        reverseIndex[Data[hash].address] ← NULL
        Data[hash] ← NULL
    end procedure
    procedure SubscribeData(hash)
        require (balance[msg.sender] >= Data[hash].price)
        balance[msg.sender] − = Data[hash].price
        Data[hash].subscribers + = msg.sender
    end procedure
    procedure RequestDataWithAddress(address)
        require (reverseIndex[address] ! = NULL)
        hash ← reverseIndex[address]
        require (msg.sender ∈ Data[hash].subscribers)
        return AccessToken for address with TTL
    end procedure
    procedure RequestDataWithHash(hash)
        require (msg.sender ∈ Data[hash].subscribers)
        address ← Data[hash].address
        return AccessToken for address with TTL
    end procedure
```

Blockchain together with SGX-based smart contract execution in SecNet bootstraps improvement of security rules, enabling well-defined security rules owned by one entity can be acquired by others in exchange of value. The owner of a security rule can register that rule on the blockchain with its name together with other metadata, and place a smart contract over a registered rule, hoping to get paid by those who acquired the rule through smart contract execution.

Algorithm 1 represents a sample smart contract for data sharing, and data here means user data, or security rules produced by AI in PDC. Data owners register their data on blockchain with procedure `RegisterData`, supplying hash of the data, typically derived from data content, description, address for fetching the data and desired price for subscribing this data. Owners can also withdraw a registered data with procedure `WithdrawData`. After registration, data consumers can view data descriptions from the blockchain and subscribe specific data with procedure `SubscribeData`, which will charge them at the price the owner designated and grant the subscriber with the permission to access that data. As subscribers of a specific data, one can request corresponding permission via smart contract with procedure `RequestData`, which will return an access token for that data. With the access token, the consumer can fetch desired data from the storage system at the corresponding address. SGX ensures that the process of smart contract execution cannot be meddled by users, and guarantees the value-exchange process. With SGX-based smart contract execution, blockchain ledger can guarantee the value-exchange process (e.g., proper value is paid for certain security services), if the related smart contract is employed properly.

Security rules can be categorized according to different types, and each type of rules can be treated as a specific data, and has its individual name. Thus, security rules can be shared based on their names, and blockchain can enable different PDCs to share their security strategies and rules with each other based on name, and record every behavior on this name, to replenish the security rules and improve the security of involved PDCs, as well as guarantee the traceability of every data interaction behavior. PDC can acquire security services by learning the security rules of other PDCs as described.

IV. USE SCENARIO

SecNet will enable enormous applications due to the inherent embedding of AI and blockchain. One of the typical cases for SecNet deployment and application is the trust medical data sharing among trust-less different parties, to support an intelligent and secure medical data management ecosystem, which is the key to a global health care system.

A. NECESSARIES OF IMPLEMENTING SECNET FOR MEDICAL CARE

The traditional way of medical data management is inefficient for building a global health care system. On the one hand, nowadays, the medical data is stored in diversified health care environment and controlled by different entities which

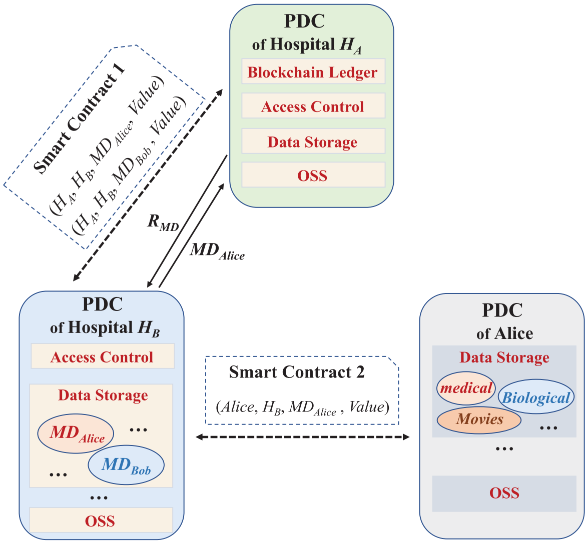

FIGURE 2. Medical data sharing using SecNet.

may have different commercial requirements. The lack of trust mechanisms for data provenance, auditing and control, makes the sharing of valuable data impossible. Moreover, in most cases, patients have to collect their medical records by themselves and then provide them to different institutions (e.g., different hospitals), although these medical records may be stored several times in other institutes before, because different institutions cannot easily share medical records due to no standard format for data or no economic incentive. On the other hand, medical data carries its owner's privacy information, but unfortunately, patients are in fact lack of authority for the usage of these data. Additionally, for better medical care services, patients have to give out their medical data without choices, due to the mismatch of the need for accurate analysis on medical data and the lack of knowledge in medical care for patients.

To solve those problems above, SecNet employs 1) blockchain-based data sharing guaranteeing, 2) smart contracts to regulate the interactions between trust-less entities, 3) AI-based secure computing for behavior analyzing, to effectively provide data provenance, auditing and control, as well as behavior tracking, via a tamper-proof way. Embedded with these characters SecNet provides, the detailed workflow to achieve trust medical data sharing is as follows.

B. MEDICAL DATA SHARING WORKFLOW USING SECNET

As shown in Figure 2, if the hospital H_A wants to use Alice's medical data MD_{Alice}, which is currently stored in another hospital H_B, to support a very important medical experiment. H_A needs to access its PDC P_A, and then send the data request R_{MD} containing the metadata/identifier ID_R to the PDC P_B belonged to H_B.

When P_B receives the R_{MD} from P_A, the Access Control module analyzes the R_{MD} with the help of ASC module in OSS, and meanwhile record this request behavior to the Blockchain Ledger, waiting for state synchronization.

After the R_{MD} is excluded from malicious access behavior according to the analyzing result from ASC as well as its submodule GAN, the Access Control module communications with the Data Storage module for the R_{MD} and then triggers the on-chain smart contract SC_1 between H_B and H_A on the requested data MD_{Alice}, and maybe necessarily triggers the smart contract SC_2 between H_B and Alice. The former regulates the value that H_A should pay for the requested data from H_B, and the latter for the value that H_B should transfer to Alice since the ownership of MD_{Alice} belongs to her.

When H_A receives the requested data MD_{Alice}, corresponding value (e.g., tokens, coins, electric cash) is transferred from H_A to H_B and from H_B to Alice, according to the smart contracts SC_1 and SC_2 respectively. That is, H_B gains rewards by providing storing service for Alice's medical data, and Alice is also paid by allowing her medical data to be shared with H_B.

To exploit the data for some further information that may be helpful to H_B, the Knowledge Computing module of P_B will merge the new received MD_{Alice} with related data storing in its Data Storage module, and may decompose the data into different types of data components (e.g., disease name, disease duration, patient name, patient age, drug-using records, *etc.*), to exploit further information and potential findings.

V. ALTERNATIVE WAY FOR SECNET

The data storage in SecNet is provided by PDC, and the security of data is the responsibility of the PDC's owner. In this way, data is under control of its owner, and any interaction with data can be monitored locally in PDC.

However, if the SecNet users wants to store their data in a secure cloud, provided by a big company which has great reputation and ability to guarantee data security, rather than storing in their own PDCs, the philosophy of InterPlanetary File System (IPFS) [24] may be a choice to replace the Data Storage module of PDC with distributed file system where data objects are exchanged within one Git repository, as shown in Figure 3.

In this way, PDC coordinates and maintains a data storing network, where all the data is treated equally, and is fragmented into data pieces and then scattered across the whole network. Thus, the privacy of data as well as the survivability can be protected better than storing all the data of a user in a single PDC. For instance, if malicious parties destroy or hack into some PDCs, they may get only some pieces of different data but cannot easily get a complete data containing valuable information, which significantly reduces the chance for privacy leakage and degrades the risk that a data is completely destroyed due to the centralized storing in a local PDC.

However, the disadvantage is that it becomes very difficult to enable personalized knowledge computing or AI-based secure computing by exploiting all the personal data for a certain user, because these data is scattered across the whole SecNet, not stored in the data-owner's PDC. One possible solution is to construct some secure computing nodes in SecNet, where data can flow in yet only answers but no

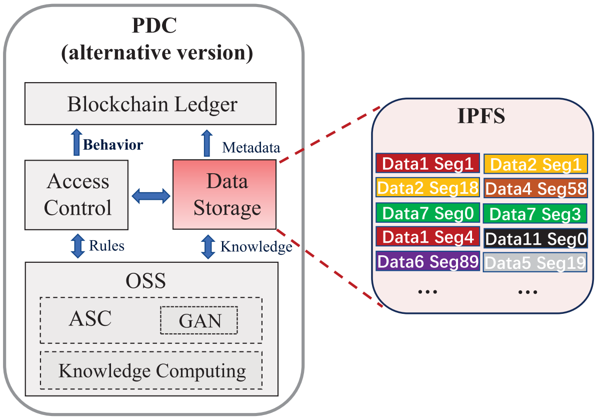

FIGURE 3. Alternative storage model of SecNet.

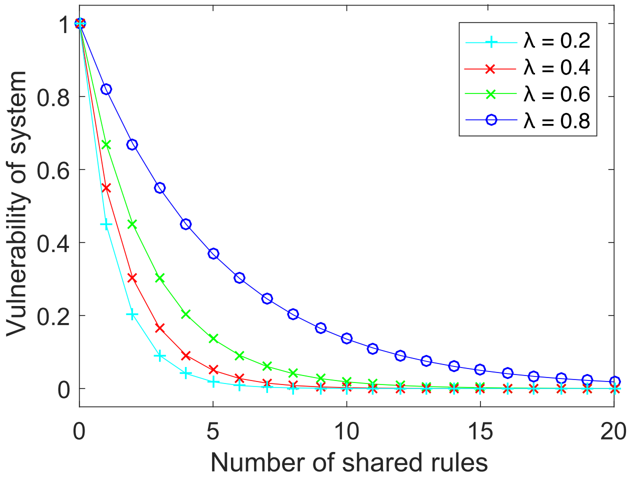

FIGURE 4. Vulnerability of SecNet when suffering DDoS security.

high-dimensional data can be flow out, to support AI-based computing and knowledge extracting for some certain users, without causing damage for data security and privacy.

VI. PERFORMANCE ANALYSIS

In this section, we evaluate the design of SecNet in two aspects: vulnerability when suffering notorious network attacks such as the Distributed Denial of Service (DDoS) Attacks, and revenue for contributors who provide the security rules on blockchain.

A. VULNERABILITY OF ARCHITECTURE

DDoS attacks continue to be one of the most serious network attacks for both the Internet infrastructure [25] and its applications [26]. Attackers can use this type of attacks to exhaust the bandwidth resource for some popular and critical Web applications, making these services unavailable to the users or even blocking Internet connectivity for a large part of a country, and thus can result in huge economic lost. For example, even a single minute of service downtime can cost up to 22000 dollars in revenue [26]. In SecNet, due to the sharing of security rules by every Internet user resulting in a more comprehensive knowledge on network security, the vulnerability that can be exploited by DDoS attackers will be decreased dramatically. That is, SecNet can greatly reduce the impact of the notorious DDoS attacks. For a scenario that DDoS attacks are happening independently and identically, the number of attacks being detected can be considered as following the Poisson distribution. In this case, we assume all the users will report their learned security rules to the blockchain once suffering DDoS attacks. Figure 4 shows the vulnerability that can be exploited by DDoS attackers (the probability of the SecNet can be attacked by DDoS attacks) varies with the sharing number of security rules, where four different security factors ($\lambda = 0.2, 0.4, 0.6, 0.8$) are considered. The security factor indicates the severity of network threats (e.g., the frequency of DDoS attacks). The results show that the vulnerability of SecNet reduces dramatically as the number of shared security rules increase. This is because the growth in the number of shared security rules leads to a more comprehensive knowledge of network security for all the participants, which makes it more difficult for attackers to launch a successful DDoS attack to avoid the detection of the growing security rules.

B. REVENUE FOR CONTRIBUTORS

The security level of SecNet will be improved continuously if every contributor shares his own security rule on blockchain with eath other, since all participants in the system have more security knowledge to protect against attacks. The revenue for each contributor is a key factor affecting contributor initiative.

Firstly, we investigate how the revenue for each participant varies when sharing security rules public for a more secure network, with different levels of rule quality control. Considering the factor in the quality effect of the real market, the revenue for every contributor will increase linearly at the very beginning stage but at different rates, yet will vary in different directions after the number of shared security rules exceeds a threshold. Accordingly, in this simulation, we reasonably set the increasing rate of the revenue of a contributor at the very beginning stage as the quality control level of the shared security rules. The quality control level represents the degree to which a rule can completely block an attack. In the evaluations, three quality control levels for the shared security rules ($\alpha_q = 0.5, 0.7, 0.95$) are investigated. After quality effect of the real market is formed, the revenue for the contributor with the highest quality will increased at a much higher rate than other ones. That is, our theoretical model fits the form of piecewise functions. Figure 5 illustrates the modeling results that the revenue for the contributor will increase at a higher rate if the shared security rules are with higher quality, especially after the quality effect of the real market is formed. This is because a high-quality security rule can characterize the network threats more accurately, and thus is more effective on countering threats than the ones with lower quality. In addition, although the whole revenue

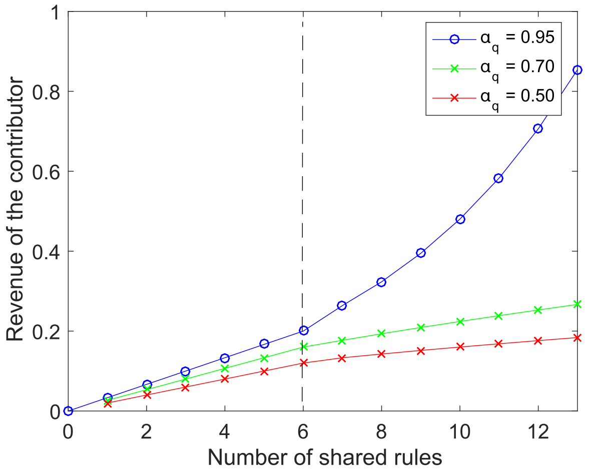

FIGURE 5. Revenue when sharing security rules with varying rule quality.

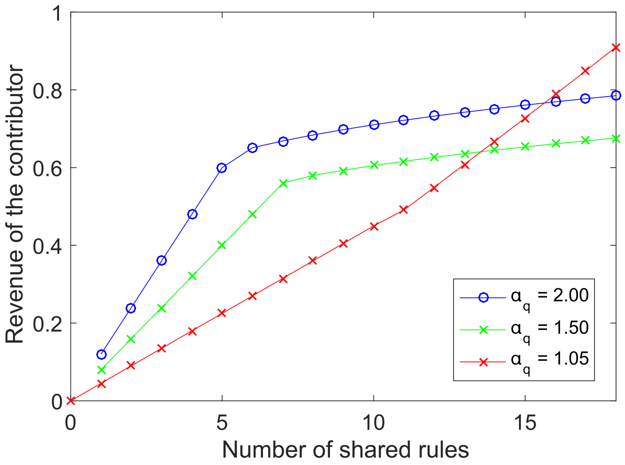

FIGURE 6. Revenue when sharing security rules with different rule price.

for all contributors is increase, the revenue for every single contributor is very different. As can be seen from the figure, when the quality effect of the real market is formed, contributors who share high-quality security rules benefit much more quickly while other contributors earn little. This is because the majority of consumers will prefer to choose high-quality security rules which are more effective on protect themselves than those with lower quality and litter effect.

Then, the effect of different rule pricing strategies on the revenue for each participant when sharing security rules is investigated. In our analyzing model, for a certain type of network threat with similar characteristics, the security rule with a higher price has more detailed description of attack characteristics and faster attack detection performance, yet may be only suitable for the security level requirements of high-end customers (e.g., commercial Banks, government data centers). For the majority of ordinary individual consumers, the security rules with similar functionality but at a lower price may be preferred. In the evaluations, three price levels for the shared security rules ($\alpha_p = 1.05, 1.5, 2$) are investigated. The price level indicates the ratio of a fixed price to its fair market value. Figure 6 indicates that if the rule price that is ready for share publicly is set unreasonably high, the revenue of the rule publisher may be decreased. This is because other participants may choose to download other security rules with similar function yet lower price. That is, every security rule has an inherent valuation. In fact, the price of a rule should be determined by the safety benefits it brings, and maybe an intelligent and fair pricing service for the shared security rules is needed to be integrated into the SecNet system in future.

VII. CONCLUSION

In order to leverage AI and blockchain to fit the problem of abusing data, as well as empower AI with the help of blockchain for trusted data management in trust-less environment, we propose the SecNet, which is a new networking paradigm focusing on secure data storing, sharing and computing instead of communicating. SecNet provides data ownership guaranteeing with the help of blockchain technologies, and AI-based secure computing platform as well as blockchain-based incentive mechanism, offering paradigm and incentives for data merging and more powerful AI to finally achieve better network security. Moreover, we discuss the typical use scenario of SecNet in medical care system, and gives alternative ways for employing the storage function of SecNet. Furthermore, we evaluate its improvement on network vulnerability when countering DDoS attacks, and analyze the inventive aspect on encouraging users to share security rules for a more secure network.

In future work, we will explore how to leverage blockchain for the access authorization on data requests, and design secure and detailed smart contracts for data sharing and AI-based computing service in SecNet. In addition, we will model SecNet and analyze its performance through extensive experiments based on advanced platforms (e.g., integrating IPFS [27] and Ethereum [28] to form a SecNet-like architecture).

REFERENCES

[1] H. Yin, D. Guo, K. Wang, Z. Jiang, Y. Lyu, and J. Xing, "Hyperconnected network: A decentralized trusted computing and networking paradigm," *IEEE Netw.*, vol. 32, no. 1, pp. 112–117, Jan./Feb. 2018.

[2] K. Fan, W. Jiang, H. Li, and Y. Yang, "Lightweight RFID protocol for medical privacy protection in IoT," *IEEE Trans Ind. Informat.*, vol. 14, no. 4, pp. 1656–1665, Apr. 2018.

[3] T. Chajed, J. Gjengset, J. Van Den Hooff, M. F. Kaashoek, J. Mickens, R. Morris, and N. Zeldovich, "Amber: Decoupling user data from Web applications," in *Proc. 15th Workshop Hot Topics Oper. Syst. (HotOS XV)*, Warth-Weiningen, Switzerland, 2015, pp. 1–6.

[4] M. Lecuyer, R. Spahn, R. Geambasu, T.-K. Huang, and S. Sen, "Enhancing selectivity in big data," *IEEE Security Privacy*, vol. 16, no. 1, pp. 34–42, Jan./Feb. 2018.

[5] Y.-A. de Montjoye, E. Shmueli, S. S. Wang, and A. S. Pentland, "openPDS: Protecting the privacy of metadata through SafeAnswers," *PLoS ONE*, vol. 9, no. 7, 2014, Art. no. e98790.

[6] C. Perera, R. Ranjan, and L. Wang, "End-to-end privacy for open big data markets," *IEEE Cloud Comput.*, vol. 2, no. 4, pp. 44–53, Apr. 2015.

[7] X. Zheng, Z. Cai, and Y. Li, "Data linkage in smart Internet of Things systems: A consideration from a privacy perspective," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 55–61, Sep. 2018.

[8] Q. Lu and X. Xu, "Adaptable blockchain-based systems: A case study for product traceability," *IEEE Softw.*, vol. 34, no. 6, pp. 21–27, Nov./Dec. 2017.

[9] Y. Liang, Z. Cai, J. Yu, Q. Han, and Y. Li, "Deep learning based inference of private information using embedded sensors in smart devices" *IEEE Netw. Mag.*, vol. 32, no. 4, pp. 8–14, Jul./Aug. 2018.

[10] Q. Xia, E. B. Sifah, K. O. Asamoah, J. Gao, X. Du, and M. Guizani, "MeDShare: Trust-less medical data sharing among cloud service providers via blockchain," *IEEE Access*, vol. 5, pp. 14757–14767, 2017.

[11] D. E. O'Leary, "Artificial intelligence and big data," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 96–99, Mar. 2013.

[12] A. Halevy, P. Norvig, and F. Pereira, "The unreasonable effectiveness of data," *IEEE Intell. Syst.*, vol. 24, no. 2, pp. 8–12, Mar. 2009.

[13] Z. Cai and X. Zheng, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Trans. Netw. Sci. Eng.*, to be published. doi: 10.1109/TNSE.2018.2830307.

[14] A. Dorri, M. Steger, S. S. Kanhere, and R. Jurdak, "BlockChain: A distributed solution to automotive security and privacy," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 119–125, Dec. 2017.

[15] J. Wang, M. Li, Y. He, H. Li, K. Xiao, and C. Wang, "A blockchain based privacy-preserving incentive mechanism in crowdsensing applications," *IEEE Access*, vol. 6, pp. 17545–17556, 2018.

[16] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 843–852.

[17] W. Meng, E. W. Tischhauser, Q. Wang, Y. Wang, and J. Han, "When intrusion detection meets blockchain technology: A review," *IEEE Access*, vol. 6, pp. 10179–10188, 2018.

[18] J.-H. Lee, "BIDaaS: Blockchain based ID as a service," *IEEE Access*, vol. 6, pp. 2274–2278, 2017.

[19] K. Wang, H. Yin, W. Quan, and G. Min, "Enabling collaborative edge computing for software defined vehicular networks," *IEEE Netw.*, vol. 32, no. 5, pp. 112–117, Sep./Oct. 2018.

[20] A. B. Kurtulmus and K. Daniel, "Trustless machine learning contracts; evaluating and exchanging machine learning models on the ethereum blockchain," 2018, *arXiv:1802.10185*. [Online]. Available: https://arxiv.org/abs/1802.10185

[21] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2nd Quart., 2016.

[22] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661*. [Online]. Available: https://arxiv.org/abs/1406.2661

[23] E. C. Ferrer, "The blockchain: A new framework for robotic swarm systems," 2017, *arXiv:1608.00695*. [Online]. Available: https://arxiv.org/abs/1608.00695

[24] *IPFS*. Accessed: Jun. 5, 2019. [Online]. Available: https://ipfs.io/

[25] S. T. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 2046–2069, 4th Quart., 2013.

[26] A. Praseed and P. S. Thilagam, "DDoS attacks at the application layer: Challenges and research perspectives for safeguarding Web applications," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 661–685, 1st Quart., 2019.

[27] J. Benet, "IPFS—Content addressed, Versioned, P2P file system," 2014, *arXiv:1407.3561*. [Online]. Available: https://arxiv.org/abs/1407.3561

[28] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," Ethereum Project Yellow Paper, 2018. Accessed: Jun. 5, 2019. [Online]. Available: https://ethereum.github.io/yellowpaper/paper.pdf

KAI WANG received the B.S. and Ph.D. degrees from Beijing Jiaotong University. He is currently a Postdoctoral with the Research Institute of Information Technology, Tsinghua University, China. He is an Assistant Professor with the School of Computer and Control Engineering, Yantai University, China. His current research interest includes cyberspace security. He has published more than 20 papers in prestigious international journals and conferences (e.g., the IEEE Network, Information Sciences), and serves as the TPC Member of IPCCC 2018/2019, the Guest Editor of *International Journal of Digital Multimedia Broadcasting*, and Technical Reviewers for many important international journals (e.g., ACM Computing Surveys).

JIAQING DONG received the B.S. degree in computer science from Peking University. He is currently pursuing the Ph.D. degreein computer science with Tsinghua University. His research interests include knowledge discovering, software-defined networking, and mobile network measurement.

YING WANG received the B.S. degree from Hunan University, and currently pursuing the master's degree in software engineering with the Wuhan University of Technology. Her research interests include networking architecture, blockchain-based applications, and data mining.

HAO YIN received the B.S., M.E., and Ph.D. degrees from Huazhong University of Science and Technology, Wuhan, China, in 1996, 1999, and 2002, respectively, all in electrical engineering. He is a Professor with the Research Institute of Information Technology (RIIT), Tsinghua University.. He was elected as the New Century Excellent Talent of the Chinese Ministry of Education in 2009, and won the Chinese National Science Foundation for Excellent Young Scholars in 2012. His research interests include multimedia communication and computer networks.