计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.201000112

面向纯文本信息隐藏的区块链隐蔽通信模型

佘 维[1,2,3] 霍丽娟[1,3] 田 钊[1,3] 刘 炜[1,2,3] 宋 轩[1,3]
1 郑州大学软件学院 郑州 450000
2 郑州大学互联网医疗与健康服务河南省协同创新中心 郑州 450000
3 郑州大学汉威物联网研究院 郑州 450000
(wshe@zzu.edu.cn)

摘 要 纯文本信息隐藏容易遭受删除、更改等主动攻击,使嵌入的秘密信息遭到破坏。区块链因具有不可篡改、不可伪造、匿名性、节点信息同步等特点,成为构建隐蔽信道的天然平台,并确保秘密信息不被破坏。文中提出了一种面向纯文本信息隐藏的区块链隐蔽通信模型。首先,根据偏序关系确定嵌入秘密信息的位置,发送方使用空格法将秘密信息嵌入到纯文本内容中;然后,构建区块链网络隐蔽通信的场景,发送方将载有纯文本内容的交易发布到区块链网络上;最后,在交易打包并形成链块后,任意节点均可作为接收方获取文件,但只有受信方可以通过嵌入算法的逆过程提取出秘密信息。实验对比及分析表明,该模型具有较好的抗检测性、鲁棒性、安全性和较高的隐藏容量。更为重要的是,以区块链作为信道的方法可使受信方身份得以隐藏,通信过程的隐蔽性得到了双重保障。

关键词:区块链;信息隐藏;隐蔽通信;纯文本
中图法分类号 TP309

Blockchain Covert Communication Model for Plain Text Information Hiding

SHE Wei[1,2,3],HUO Li-juan[1,3],TIAN Zhao[1,3],LIU Wei[1,2,3] and SONG Xuan[1,3]
1 School of Software,Zhengzhou University,Zhengzhou 450000,China
2 Henan Collaborative Innovation Center for Internet Medical and Health Services,Zhengzhou University,Zhengzhou 450000,China
3 Hanwei Internet of Things Research Institute,Zhengzhou University,Zhengzhou 450000,China

Abstract Plain text information hiding is vulnerable to active attacks such as deletion and change,which makes the embedded secret information damaged. Blockchain is characterized by non-tampering,non-forgery,anonymity and node information synchronization,making it a natural platform for building hidden channels and ensuring that secret information is not destroyed. This paper proposes the blockchain covert communication model for plain text information hiding. Firstly,the location of the embedded secret information is determined according to the partial order relation. The sender uses the space method to embed the secret information into the plain text content. Then,a scenario of hidden communication in the blockchain network is constructed,and the sender publishes the transaction containing the plain text content to the blockchain network. Finally,after the transaction is packaged and a chain block is formed,any node can obtain the file as the receiver,but only the trusted party can extract the secret information through the inverse process of the embedded algorithm. Experimental comparison and analysis show that the model has better anti-detection,robustness,security and higher hiding capacity. More importantly,the blockchain as a channel enables the identity of the trusted party to be hidden,and the concealment of the communication process is doubly guaranteed.

Keywords Blockchain,Information hiding,Covert communication,Plain text

到稿日期:2020-10-21 返修日期:2021-02-04 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家重点研发计划(2018YFB1201403);河南省高校科技创新人才支持计划(21HASTIT031);河南省重大公益专项(201300210300);河南省高等学校青年骨干教师培养计划(2019GGJS018);河南省高等学校重点科研项目(20A520035);郑州市协同创新重大专项(20XTZX06013)
This work was supported by the National Key Research and Development Project(2018YFB1201403),Program for Science&Technology Innovation Talents in Universities of Henan Province(21HASTIT031),Major Public Welfare Project of Henan Province (201300210300),Training Plan for Young Backbone Teachers of Colleges and Universities in Henan(2019GGJS018),Key Scientific Research Project of Colleges and Universities in Henan Province(20A520035) and Collaborative Innovation Major Project of Zhengzhou(20XTZX06013).
通信作者:田钊(tianzhao@zzu.edu.cn)

1 引言

信息隐藏指利用多媒体信息编码的冗余将秘密信息隐藏，并且不会引起人类视听系统的察觉[1-2]。常见的多媒体信息载体包括图像、音频、视频、文本等。其中，文本因其数据量小、编码简单、方便传输等特点成为互联网中使用最多的媒体类型[3]。然而，文本数据和其他载体数据不同，改变数据信息编码的任何一位，都可以直观地看到文本内容发生变化[4]，它需要特殊的方法将秘密信息进行隐藏。常见的文本信息隐藏有3类：广义空格法、句法结构方法以及语义方法[5]。网络隐蔽通信是信息隐藏的一个研究分支，允许通信双方在不违反系统通信规则的情况下[6]，将经过信息隐藏技术处理过的信息从发送方传递给受信方。但是现有的网络隐蔽通信过程中存在一个致命性安全缺陷，即文本信息容易遭受删除、更改等主动攻击，使嵌入的秘密信息遭到破坏。除此之外，大部分网络隐蔽通信的双方直接暴露于网络之上，通信双方易被攻击者察觉，遭到针对性地检测、阻断和干扰。

区块链本质上是一个分布式共享的数据库[7]，链上的所有节点共同维护着存储在这一数据库中的数据信息[7]。与普通的多媒体载体相比，区块链具有不可篡改性、不可伪造性、匿名性、节点信息同步等特点[8-9]，这些特点使区块链成为构建隐蔽信道的天然平台。不可篡改性使隐蔽通信的攻击者对数据的删除和修改都是无效的，即秘密信息不会遭受删除、更改等主动攻击。不可伪造性使攻击者不能冒充发送者伪造秘密消息内容。匿名性使隐蔽通信的双方可以进行匿名通信，受信方不必暴露身份。节点信息同步使发送方不用将信息直接发送给受信方，受信方也能同步获得信息。

Juha[10]首次提出一种安全嵌入秘密消息到区块链的方法，即在简化的理想区块链模型中，通过改变交易地址的最低位来嵌入1位秘密信息，并对其安全性进行了研究和验证。Tian等[11]提出一种新的区块链隐蔽信道构建方案DLchain，将真实交易数据统计分布的动态标签生成算法，以保证动态标签的隐藏。Fionov[12]探讨了比特币系统中隐蔽信道的存在，并描述这些很难或不可能检测的信道和它们的信息传输能力。Li等[13]提出区块链环境下的网络隐蔽信道模型，其具有抗干扰性、抗篡改性、多线路通信性、接收方匿名性和线路无关性，可以克服现有网络环境下的隐蔽信道特性缺陷。

考虑到纯文本为常用的数据载体，本文用纯文本作为秘密信息的载体，利用信息隐藏技术中的嵌入算法将秘密信息嵌入到纯文本中，生成含密载体。含密载体需要通过开放的环境进行直接或间接的传输，为了保障安全模型中最后一层的安全，必须使用隐蔽通信。然而在区块链网络环境下，将纯文本信息隐藏与隐蔽通信结合起来的研究较少，因此本文提出了一种面向纯文本信息隐藏的区块链隐蔽通信模型，在区块链网络中构建隐蔽信道，将嵌有秘密信息的纯文本文件在区块链网络中进行传输。本文中，一个纯文本文件可以嵌入2位及多位秘密信息，而且一个区块可以打包多个来自发送方的交易，提高了秘密信息的嵌入量，缩短了整条秘密信息的传输时间，以实现在短时间内传递大量秘密信息。本文根据偏序关系确定嵌入秘密信息的位置，即发送方提交的交易中携带的纯文本可能含有秘密信息，也可能不含秘密信息，使攻击者不易区分含秘密信息的交易和普通交易。以区块链作为信道的方法，使区块链网络中的所有节点都可能是信息的接收方，使攻击者无法准确辨别出受信方，从而不易让攻击者检测到隐蔽信道，双重保障通信过程的隐蔽性。

本文第2节描述了区块链、隐蔽通信以及信息隐藏的基本概念；第3节提出了面向纯文本信息隐藏的区块链隐蔽通信模型；第4节进行了实验并分析了实验结果；最后总结全文。

2 相关知识

2.1 区块链技术

区块链是一种按时序将数据区块以链式结构进行存储的分布式账本[7]，并通过密码学和共识机制来保证区块链中的数据不被篡改、不被伪造[8]，具有去中心化、可追溯性、不可篡改性、不可伪造性、不可抵赖性、匿名性、可编程性和节点信息同步等特点[8]。

去中心化指在区块链中没有中心节点，所有节点基于分布式系统结构彼此建立信任关系，进行数据的记录、存储、传输和验证等过程。去中心化的实现依靠的是P2P(点对点网络技术)和共识机制，是区块链最显著的优势。区块链通过与密码学相结合，可以保证交易的可追溯性、不可篡改性、不可否认性和不可伪造性。可追溯性指区块链中的每一笔交易的输入输出数据都会通过区块的数据结构存储，链式结构的存储方式使得所有交易前后关联，每一笔交易都可以追溯其源头。不可篡改性指所有的数据都是公开透明的，单个节点对数据的修改是无效的。不可伪造性指区块链结合密码学技术，使用隐蔽通信中的发送者的私钥对合法交易进行签名，以便让其他节点进行验证，即攻击者不能冒充发送者伪造秘密消息内容。同时，使用私钥进行签名也保证了不可抵赖性。匿名性指各个节点的身份信息不必公开，交易是基于地址而不是个人身份来完成。以太坊为区块链提供了一个可编程的数据共享平台，实现了区块链的可编程性。节点信息同步指交易数据通过验证达成共识并写入区块链中，所有节点共同记录这些交易数据，即所有节点各自的数据保持一致。

一般说来，区块链系统由数据层、网络层、共识层、激励层、合约层和应用层组成[14-16]，如图1所示。数据层封装了底层数据区块以及相关的数据加密和时间戳等技术，如图2所示，是实现其他5层功能的基础；网络层包括分布式组网机制、数据传播机制和数据验证机制等，其核心是确保区块链节点的合法加入和有效通信；共识层是区块链系统的核心，主要封装网络节点的各类共识算法；激励层将经济因素集成到区块链技术体系中，主要包括经济激励的发行机制和分配机制等；合约层主要封装各类脚本、算法和智能合约，是区块链可编程特性的基础；应用层封装了区块链的各种应用场景和案例[17]。

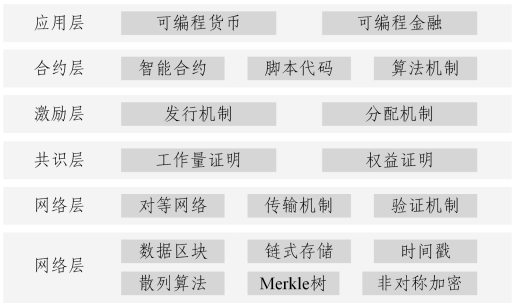


图 1　区块链的基础架构

Fig. 1　Blockchain infrastructure

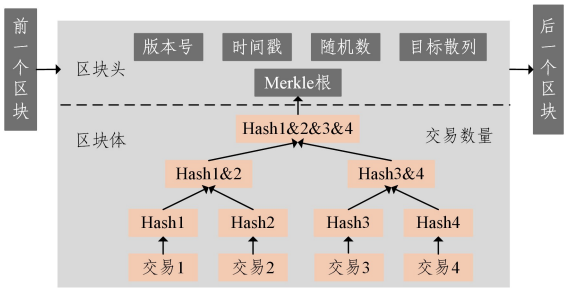


图 2　区块结构

Fig. 2　Block structure

2.2　隐蔽通信

隐蔽通信最早由 Lampson 提出[18]，是信息隐藏的一个重要研究分支。隐蔽通信是允许通信双方在不违反系统通信规则的情况下，将经过信息隐藏技术处理过的信息从发送方传递给受信方，网络隐蔽通信过程隐藏了秘密信息正在传递的事实。隐蔽通信中的隐蔽信道可分类两大类：一类是存储型隐蔽信道[19]，通过将秘密信息隐藏在网络数据包中进行传递；另一类是时分型隐蔽信道[19]，通过将秘密信息调制到网络数据包中的时间间隔进行传递。本文中的接收方指能接收到信息的节点，受信方指能接收到信息并且可以通过嵌入算法的逆过程提取出秘密信息的节点。

区块链网络隐蔽通信模型如图 3 所示，本文假设事先通过安全通道传输密钥 K。发送方 Alice 使用特定的嵌入算法将秘密信息嵌入到公开的载体中，含密载体经区块链隐蔽信道进行传递，区块链网络中所有节点都可能是信息的接收方，但只有受信方 Bob 可以使用特定的提取算法从含密载体中提取出秘密信息。在嵌入秘密信息过程中，发送方使用密钥 K 对秘密信息进行加密，受信方只有拥有和嵌入过程相同或者相关的密钥 K 才能提取出秘密信息。在秘密信息的传输过程中，受信方身份并不公开，攻击者不仅无法准确辨别出受信方，而且也无法提取出秘密信息，使通信过程的隐蔽性得到了双重保障。

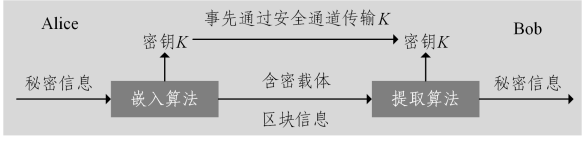


图 3　区块链隐蔽通信模型

Fig. 3　Blockchain covert communication model

2.3　纯文本信息隐藏

信息隐藏和信息加密不同，信息隐藏指监听者不知道秘密信息的存在，在嵌入秘密信息之后监听者看到的是与原始数据一样的数据，而信息加密则是把秘密信息处理成不易看懂的数据。

文本数据类型很多，根据内容表现的样式可分为格式化文本和无格式文本。在格式化文本中，可以为有相同编码的字符设置不同的样式，如不同的大小、颜色等，这给格式化文本提供了格式冗余[20]。除此之外，格式化文本还存在语法冗余和语义冗余[21-22]。常见的 PDF，DOC，WPF 等就是格式化文本。在无格式文本中，有相同编码的字符只有一种表现形式，常见的纯文本(txt)就是一种无格式文本。这种文本数据不存在格式冗余，改变数据信息编码的任何一位，都会让文本本身发生错误。常见的纯文本信息隐藏方法是基于空格法的信息隐藏。

本文采用基于空格法的信息隐藏，即在文本每一行行尾嵌入空格，用嵌入空格的个数来对应秘密信息的二进制编码。嵌入空格的个数为 1 时，代表嵌入的二进制信息为“1”；嵌入空格的个数为 2 时，代表嵌入的二进制信息为“0”。因为空格在文本的正常显示和阅读都不会显示，所以不会引起人类感知系统的察觉。结合区块链不可篡改性、不可伪造性、匿名性等特点，文本信息可抵御主动攻击的破坏。

3　面向纯文本信息隐藏的区块链隐蔽通信模型

本文定义了面向纯文本信息隐藏的区块链隐蔽通信模型，其形式化描述为 $\Pi=(Embed, blockchainCC, Extract)$。其中，$Embed$ 代表嵌入模块，功能是将秘密信息嵌入到纯文本文件中；$BlockchainCC$ 代表区块链网络隐蔽通信模块，功能是进行纯文本文件的传输；$Extract$ 代表提取模块，功能是将秘密信息从纯文本文件中提取出来。

3.1　嵌入模块

(1)发送方将秘密信息 m 加密得到密文 c。在本文中，采用对称加密方案 $AES=(K, Encrypt, Decrypt)$ 对秘密信息进行加解密，其中 K 是密钥；$Encrypt$ 是一种加密函数，$c \leftarrow Encrypt(K, m)$，通过输入明文消息 m 和密钥 K，得到密文 c；$Decrypt$ 是一种解密函数，$m \leftarrow Decrypt(K, c)$，使用密钥 K 将密文 c 解密得到明文消息 m。

(2)将密文 c 转成二进制序列 BS，BS 的形式化描述为 $BS=(BS_1, BS_2, \cdots, BS_n)$，其中，$BS_i=0$ 或 $BS_i=1$，$i=1, 2, \cdots, n$。

(3)将二进制序列 BS 嵌入到纯文本文件中。偏序关系的基本定义为：给定集合 A，“$\leqslant$”是 A 上的二元关系，若“$\leqslant$”满足自反性($\forall a \in A$，有 $a \leqslant a$)、反对称性($\forall a, b \in A, a \leqslant b$ 且 $b \leqslant a$，则 $a=b$)、传递性($\forall a, b, c \in A, a \leqslant b$ 且 $b \leqslant c$，则 $a \leqslant c$)，则称“$\leqslant$”是 A 上的偏序关系。假设由集合 A 中元素构成的二元组用 $\langle m, n \rangle$ 表示，则称集合 A 上满足偏序关系的二元组集合为偏序集 G，其中，二元组 $\langle m, n \rangle$ 中的 m 和 n 是集合 A 中的元素，即 $G=\{\langle m, n \rangle \mid m, n \in A \bigcap m \leqslant n\}$。

假设集合 A 为一周内日最高气温数据，每天的日最高气温数据都会记录在一个 txt 文件中，全部的 txt 文件用 T

表示，$T=(T_1,T_2,\cdots,T_n)$，其中，n 为天数，T_i 为记录当天日最高气温数据的 txt 文件，如表 1 所列，则 $A=\{1.7,3,3.9,4.6,5.9,7.9,8\}$，$G=\{\langle 1.7,1.7\rangle,\langle 1.7,3\rangle,\langle 1.7,3.9\rangle,\langle 1.7,4.6\rangle,\langle 1.7,5.9\rangle,\langle 1.7,7.9\rangle,\langle 1.7,8\rangle,\langle 3,3\rangle,\langle 3,3.9\rangle,\langle 3,4.6\rangle,\langle 3,5.9\rangle,\langle 3,7.9\rangle,\langle 3,8\rangle,\langle 3.9,3.9\rangle,\langle 3.9,4.6\rangle,\langle 3.9,5.9\rangle,\langle 3.9,7.9\rangle,\langle 3.9,8\rangle,\langle 4.6,4.6\rangle,\langle 4.6,5.9\rangle,\langle 4.6,7.9\rangle,\langle 4.6,8\rangle,\langle 5.9,5.9\rangle,\langle 5.9,7.9\rangle,\langle 5.9,8\rangle,\langle 7.9,7.9\rangle,\langle 7.9,8\rangle,\langle 8,8\rangle\}$。

表 1 中二元组 $\langle x,y\rangle$ 由相邻两天的日最高气温数据组成，即 x 为 T_{i-1} 携带的日最高气温数据，y 为 T_i 携带的日最高气温数据。若满足 $\langle x,y\rangle\in G$，则在 T_i 中嵌入二进制序列。根据表 1，在 T_3，T_5，T_6，T_7 中嵌入二进制序列。

表 1 一周内的日最高气温数据

Table 1 Daily maximum temperature data for a week

日期	txt 文件	气温/℃	二元组 $\langle x,y\rangle$
周一	T_1	4.6	
周二	T_2	3	$\langle 4.6,3\rangle$
周三	T_3	3.9	$\langle 3,3.9\rangle$
周四	T_4	1.7	$\langle 3.9,1.7\rangle$
周五	T_5	5.9	$\langle 1.7,5.9\rangle$
周六	T_6	7.9	$\langle 5.9,7.9\rangle$
周日	T_7	8	$\langle 7.9,8\rangle$

本文中采用基于空格法的信息隐藏技术将秘密信息嵌入 txt 文件中，即采用在行尾嵌入空格的方法，从而可以在一个 txt 文件中嵌入 2 位及多位秘密信息，本模型以嵌入 2 位为例。如果待嵌入的二进制比特 BS_i 为 1，则在纯文本文件 T_i 行尾加一个空格(空格的十六进制为 0x20)；如果待嵌入的二进制比特 BS_i 为 0，则在纯文本文件 T_i 行尾加两个空格。其形式化描述如式(1)所示：

$$T_i\leftarrow\begin{cases}0x20, & BS_i=1\\ 0x200x20, & BS_i=0\end{cases}\tag{1}$$

其中，$i=1,2,\cdots,n$。假设秘密信息二进制序列为 1001001101，则在 T_3 中嵌入二进制序列 10，在 T_5 中嵌入二进制序列 01，在 T_6 中嵌入二进制序列 00，在 T_7 中嵌入二进制序列 11，如图 4 所示。

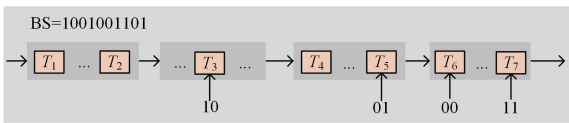


图 4 嵌入过程

Fig. 4 Embedding process

3.2 区块链网络隐蔽通信模块

在区块链网络隐蔽通信过程中，发送方 Alice 向区块链网络中发起多笔交易，这些交易的 data 字段存放的可能是嵌入秘密信息的 txt 文件，也可能是不含秘密信息的 txt 文件。发送方 Alice 不直接选择受信方 Bob 的账户作为接收地址，而是随机选择一些其他账户作为接收地址，用来提交这些交易，并将它们广播到网络中；经过合法性验证，将它们打包到一个区块中。受信方 Bob 不断地遍历新生成的区块，寻找其中 Alice 提交的交易，并按照提取算法提取嵌入的秘密信息。区块链的交易过程如图 5 所示。

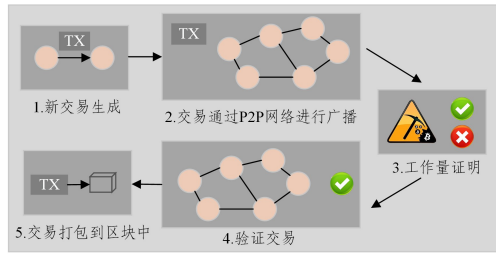


图 5 区块链交易过程

Fig. 5 Blockchain transaction process

(1)交易的生成

假设区块链网络中交易的发送方 Alice 用 P_s 表示，Alice 的公私钥对用 (P_k^s,S_k^s) 表示，并用公钥得到账户地址 $a^s=H(P_k^s)$。接收方用 P_r 表示，接收方的公私钥对用 (P_k^r,S_k^r) 表示，$a^r=H(P_k^r)$ 表示接收方的账户地址。

Alice 提交的单个交易 P_i 定义为：$P_i=(a^s,a^r,t,\delta,T)$，其中，$i=1,2,\cdots,n$，a^s 是发送方的地址，a^r 表示接收方地址，t 表示时间戳，$\delta=Sign(S_k^s,(P_k^s,a^r,t,T))$ 表示数据签名，T 表示嵌有秘密信息的 txt 文件或者不含秘密信息的 txt 文件。数字签名用于身份验证，发送方 Alice 使用自己的私钥对消息进行签名 $\delta\leftarrow Sign(S_k^s,m)$，以便接收方能验证交易是 Alice 提交的，最后通过判断是否能输出 $1\leftarrow Verify(P_k^s,m,\delta)$ 来验证签名 δ。

(2)交易通过 P2P 网络进行广播

在区块链网络中，每个节点都有一个分布式的数据库，用来管理交易的信息。当一个节点发起一笔交易以后，这个节点要立即向附近的节点进行广播，附近的节点会检查该交易是否有效，如果有效，表示它们同意这次交易。在同意的基础上，这些节点又会将这笔交易再向附近的节点进行广播，很快整个网络就会确认这笔交易。

因为交易在区块链网络中进行广播，区块链网络中任意节点均可作为接收方获取文件，最终受信方 Bob 也能收到 Alice 提交的交易，如图 6 所示。但只有受信方 Bob 可以通过密文嵌入的逆过程提取出秘密信息，区块链网络隐蔽通信可以隐藏受信方 Bob 身份，使攻击者无法准确辨别出受信方 Bob，从而不易让攻击者检测到隐蔽信道。

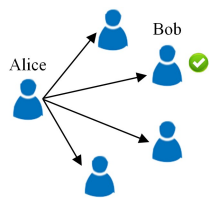


图 6 交易广播

Fig. 6 Transaction broadcast

(3)工作量证明

每个节点通过工作量证明机制来获取记账的权利。该过程相当于解一道数学题，最先解出的节点用于验证交易，然后向全网广播该区块所有的交易。

(4)验证交易

其他节点验证最先“解出数学题”的区块内的所有交易是否合法。若合法，则节点接受该区块加入到链中；若不合法，则验证失败，节点不会接受该区块加入到链中。

(5)交易打包到区块中

其他节点验证通过后，将交易打包到区块中，节点在该区块之后竞争下一个区块。

3.3　提取模块

提取是嵌入的逆过程。实际接收者 Bob 不断地遍历新生成的区块，根据 Alice 的签名查找区块中 Alice 提交的交易。

(1)实际受信方 Bob 遍历新生成的区块，查找 Alice 提交的全部交易 $P=(P_1,P_2,\cdots,P_n)$。

(2)在 Alice 提交的所有交易中，根据交易提交的时间戳的先后顺序依次提取交易中 data 字段的 txt 文件 $T'=(T_1{}',T_2{}',\cdots,T_n{}')$，$T_i{}'$为交易 P_i 中携带的 txt 文件。

(3)根据 $T_1{}',T_2{}',\cdots,T_n{}'$中的日最高气温数据构成集合 A'，集合 A'中相邻两天的日最高气温数据构成二元组$\langle x',y'\rangle$，其中，x'为 T'_{i-1}携带的日最高气温数据，y'为 $T_i{}'$携带的日最高气温数据。根据集合 A'和偏序关系“≤”构成偏序集 G'。若满足$\langle x',y'\rangle\in G'$，则 $T_i{}'$是嵌入秘密信息二进制序列的文件。

(4)在嵌入秘密信息的 txt 文件中，查找行尾的空格个数 sp。如果行尾的空格个数为 1，则嵌入的二进制比特 $BS_i{}'$为 1；如果行尾的空格个数为 2，则嵌入的二进制比特 $BS_i{}'$为 0。其形式化描述如下式所示：

$$BS_i{}'=\begin{cases}1, & sp=1\\0, & sp=2\end{cases}\tag{2}$$

其中，$i=1,2,\cdots,n$。拼接提取的二进制字符 $BS_i{}'$形成二进制序列 BS'，再将二进制序列 BS'转成密文 c'，经 AES 解密得到秘密信息 $m'\leftarrow Decrypt(K,c')$。

4　实验与分析

4.1　实验设计

本文中使用 ganache 来模拟以太坊节点，ganache 启动后可以在 8545 端口监听 http 请求，使用 http 协议将 JSON RPC 调用请求发送到 8545 端口。

实验目标是在区块链网络中构建隐蔽信道，将嵌有秘密信息的纯文本文件在区块链网络中进行传输。假设本次实验需要被嵌入的秘密信息 m，经 AES 加密后得密文 c，将密文转成二进制序列为 bs，则上述符号代表的含义和值如表 2 所列。

表 2　嵌入过程中相关符号的说明

Table 2　Description of related symbols in the embedded process

符号	含义	值
m	秘密信息	secret
c	密文	U2FsdGVkX18Y792kU0xvK10OIm6rIlJd7Hz1h73jro0=
bs	二进制序列	01010101001100100100011001110011011001000100011101010110011010110101100000110001001110000101100100 11011100111001001100100110101101010101001100000111100001110110010010110011000100110000010011110100 100101101101001101100111001001001001011011000100101001100100001101110100100001111010001100010110100 0001101110011001101101010011100100110111100110000000111101

本次实验选取 2019 年全年(364 天)日最高气温数据为原始数据，发送方 Alice 将每天的日期、地点、日最高气温数据和日最低气温数据写入 txt 文本中，文本中的原始数据信息如表 3 所列。

表 3　原始数据信息

Table 3　Raw data information

日期	2019 年 1 月 2 日
地点	河南，郑州
日最高气温	4℃
日最低气温	−7.9℃

2019 年全年日最高气温数据如表 4 所列，相邻两天的日最高气温数据构成二元组$\langle x,y\rangle$。从原始数据信息中提取日最高气温数据，把不重复的日最高气温数据放到集合 A 中，得到 $A=\{-1.1,-0.8,-0.3,0.1,0.5,1.4,1.6,1.7,1.8,2.2,2.5,2.8,2.9,3,3.3,3.9,4,\cdots,40\}$。根据集合 A 和偏序关系“≤”构成偏序集 $G=\{\langle-1.1,-1.1\rangle,\langle-1.1,-0.8\rangle,\langle-1.1,-0.3\rangle,\langle-1.1,0.1\rangle,\langle-1.1,0.5\rangle,\langle-1.1,1.4\rangle,\langle-1.1,1.6\rangle,\langle-1.1,1.7\rangle,\langle-1.1,1.8\rangle,\langle-1.1,2.2\rangle,\langle-1.1,2.5\rangle,\langle-1.1,2.8\rangle,\langle-1.1,2.9\rangle,\langle-1.1,3\rangle,\langle-1.1,3.3\rangle,\langle-1.1,3.9\rangle,\langle-1.1,4\rangle,\langle-1.1,4.4\rangle,\langle-1.1,4.6\rangle,\langle-1.1,4.7\rangle,\langle-1.1,5.1\rangle,\langle-1.1,5.2\rangle,\langle-1.1,5.2\rangle,\langle-1.1,5.3\rangle,\langle-1.1,5.4\rangle,\langle-1.1,5.9\rangle,\langle-1.1,6\rangle,\langle-1.1,6.2\rangle,\langle-1.1,6.3\rangle,\cdots,\langle 39.1,39.1\rangle,\langle 39.1,39.2\rangle,\langle 39.1,40\rangle,\langle 39.2,39.2\rangle,\langle 39.2,40\rangle,\langle 40,40\rangle\}$。

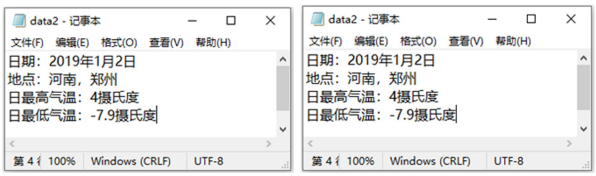
表 4　2019 年全年日最高气温数据

Table 4　2019 annual maximum daily temperature data

序号	txt 文件	气温/℃	二元组$\langle x,y\rangle$	二进制序列的嵌入
1	T_1	4		
2	T_2	4	〈4,4〉	01
3	T_3	4.7	〈4,4.7〉	01
4	T_4	2.9	〈4.7,2.9〉	
5	T_5	1.8	〈2.9,1.8〉	
6	T_6	4.6	〈1.8,4.6〉	01
7	T_7	3	〈4.6,3〉	01
8	T_8	3.9	〈3,3.9〉	
9	T_9	1.7	〈3.9,1.7〉	
10	T_{10}	5.9	〈1.7,5.9〉	00
11	T_{11}	7.9	〈5.9,7.9〉	11
12	T_{12}	8	〈7.9,8〉	00
13	T_{13}	6.8	〈8,6.8〉	
14	T_{14}	7.2	〈6.8,7.2〉	10
15	T_{15}	4.4	〈7.2,4.4〉	
16	T_{16}	5.2	〈4.4,5.2〉	01
17	T_{17}	10.3	〈5.2,10.3〉	00
18	T_{18}	10.6	〈10.3,10.6〉	01
19	T_{19}	6.2	〈10.6,6.2〉	
20	T_{20}	6	〈6.2,6〉	
21	T_{21}	8.9	〈6,8.9〉	10
…	…	…	…	
324	T_{324}	15.9	〈13.5,15.9〉	11
325	T_{325}	16.7	〈15.9,16.7〉	01
326	T_{326}	16.1	〈16.7,16.1〉	
327	T_{327}	12.8	〈16.1,12.8〉	
328	T_{328}	9	〈12.8,9〉	
329	T_{329}	6.8	〈9,6.8〉	
330	T_{330}	6.2	〈6.8,6.2〉	
331	T_{331}	7.3	〈6.2,7〉	
…	…	…	…	
360	T_{360}	14.5	〈7.7,14.5〉	
361	T_{361}	12.6	〈14.5,12.6〉	
362	T_{362}	13.4	〈12.6,13.4〉	
363	T_{363}	6.5	〈13.4,6.5〉	
364	T_{364}	1.4	〈6.5,1.4〉	

表 4 中若二元组$\langle x,y\rangle \in G$，则 y 所在的 txt 文件被选中来嵌入秘密信息，秘密信息的二进制序列嵌入情况如表 4 第 5 列所示。T_{331}，T_{333}，T_{334}，T_{336}，T_{337}，T_{340}，T_{341}，T_{342}，T_{344}，T_{346}，T_{349}，T_{350}，T_{352}，T_{354}，T_{355}，T_{357}，T_{358}，T_{359}，T_{360}，T_{362} 虽然被选中，但是秘密信息已经嵌入完成，所以不必再进行处理。

嵌入结果以 T_2 为例，在第一行的行尾嵌入 2 个空格，第二行的行尾嵌入 1 个空格，嵌入秘密信息前后的数据信息对比如图 7 所示，可以看出，在视觉上不会发现含秘密信息的文本和原始文本的不同。



(a) Before embedding secret information (b) After embedding secret information

图 7 嵌入秘密信息前后对比

Fig. 7 Comparison before and after embedding secret information

使用命令 ganache-cli-a 500 创建 500 个账户，如表 5 所列。假设账户 1 为发送方 Alice 的账户，在其他账户中随机选择 364 个账户作为接收方。发送方创建交易 payload，交易的 data 字段存放的是嵌入秘密信息的 txt 文件，用发送方的私钥对交易进行签名 signed，之后将交易发布到区块链网络中，经过合法性验证后，将交易打包到一个区块中。

表 5 500 个账户

Table 5 500 accounts

账户	账户地址
1	0x535f9aafb28Ad66f623B2E76Fa5e78a09282F5C7
2	0xCe8584A4f0D14638b005d61E5441B6b3CcD13b97
3	0x416063b017c7486B8bA98Dcea2E8661621aC0326
4	0x469BC4817E816C10Ed43eDDdEa690372aA706eeD
…	…
500	0xC4cE5ef277408cA72245a8ac91E56b6e85d7A950

受信方 Bob 遍历新产生的区块，根据 Alice 的签名查找 Alice 提交的交易。找到 Alice 提交的全部交易后，根据交易提交时的时间戳顺序依次从交易中的 data 字段中提取出 txt 文件 $T'=(T_1',T_2',\cdots,T_n')$。在全部 txt 文件中，把文本中第三行的日最高气温数据截取出来，如将“日最高气温为：4℃”中的数据“4”放到数组 datas 中，得 $datas=[4,4,4.7,2.9,1.8,4.6,3,3.9,1.7,5.9,6.8,7.2,4.4,5.2,10.3,10.6,6.2,6,8.9,15.2,13.1,13,8.8,5.6,9.7,7.7,9.4,5.3,2.8,9.2,12.1,\cdots,14.5,12.6,13.4,6.5,1.4]$，数组 datas 中相邻两天的日最高气温数据构成二元组$\langle x'y'\rangle$，其中，x'为 T'_{i-1}携带的日最高气温数据，y'为 T_i'携带的日最高气温数据，则二元组集合为$\{\langle x',y'\rangle\}=\{\langle 4,4\rangle,\langle 4,4.7\rangle,\langle 4.7,2.9\rangle,\langle 2.9,1.8\rangle,\langle 1.8,4.6\rangle,\langle 4.6,3\rangle,\langle 3,3.9\rangle,\langle 3.9,1.7\rangle,\langle 1.7,5.9\rangle,\langle 5.9,7.9\rangle,\langle 7.9,8\rangle,\langle 8,6.8\rangle,\langle 6.8,7.2\rangle,\langle 7.2,4.4\rangle,\langle 4.4,5.2\rangle,\langle 5.2,10.3\rangle,\langle 10.3,10.6\rangle,\langle 10.6,6.2\rangle,\langle 6.2,6\rangle,\langle 6,8.9\rangle,\langle 8.9,15.2\rangle,\langle 15.2,13.1\rangle,\langle 13.1,13\rangle,\langle 13,8.8\rangle,\langle 8.8,5.6\rangle,\langle 5.6,9.7\rangle,\langle 9.7,7.7\rangle,\cdots,\langle 4.6,4.7\rangle,\langle 4.7,7.7\rangle,\langle 7.7,14.5\rangle,\langle 14.5,12.6\rangle,\langle 12.6,13.4\rangle,\langle 13.4,6.5\rangle,\langle 6.5,1.4\rangle\}$。将数组 datas 中重复的数据去除得到集合 $A'=\{-1.1,-0.8,0.3,0.1,0.5,1.4,1.6,1.7,1.8,2.2,2.5,2.8,2.9,3,3.3,3.9,4,\cdots,40\}$。

根据集合 A'和偏序关系“$\leqslant$”构成的偏序集为 $G'=\{\langle -1.1,-1.1\rangle,\langle -1.1,-0.8\rangle,\langle -1.1,-0.3\rangle,\langle -1.1,0.1\rangle,\langle -1.1,0.5\rangle,\langle -1.1,1.4\rangle,\langle -1.1,1.6\rangle,\langle -1.1,1.7\rangle,\langle -1.1,1.8\rangle,\langle -1.1,2.2\rangle,\langle -1.1,2.5\rangle,\langle -1.1,2.8\rangle,\langle -1.1,2.9\rangle,\langle -1.1,3\rangle,\langle -1.1,3.3\rangle,\langle -1.1,3.9\rangle,\langle -1.1,4\rangle,\langle -1.1,4.4\rangle,\langle -1.1,4.6\rangle,\langle -1.1,4.7\rangle,\langle -1.1,5.1\rangle,\langle -1.1,5.2\rangle,\langle -1.1,5.2\rangle,\langle -1.1,5.3\rangle,\langle -1.1,5.4\rangle,\langle -1.1,5.9\rangle,\langle -1.1,6\rangle,\langle -1.1,6.2\rangle,\langle -1.1,6.3\rangle,\cdots,\langle 39.1,39.1\rangle,\langle 39.1,39.2\rangle,\langle 39.1,40\rangle,\langle 39.2,39.2\rangle,\langle 39.2,40\rangle,\langle 40,40\rangle\}$。

若上述二元组集合中的元素满足$\langle x',y'\rangle \in G'$，则 y'所在的 txt 文件，即是嵌入秘密信息的文件。由表 4 可知，在 T_2，T_3，T_6，T_7，T_8，T_{10}，T_{11}，T_{12}，T_{14}，T_{16}，T_{17}，T_{18}，T_{21}，T_{22}，T_{27}，$\cdots$，T_{322}，T_{324}，T_{325}中嵌有秘密信息。

以 T_2 为例，把文件中的内容转成十六进制，查找 0x20 的个数，若有 1 个 0x20，则代表嵌入秘密信息的二进制为“1”；若有 2 个相邻的 0x20，即“0x200x20”，则代表嵌入秘密信息的二进制为“0”。用相同的方法处理完所有嵌入秘密信息的文件，拼接获得的二进制，最终得到二进制序列 bs'。把二进制序列转成密文 c'，最后密文经 AES 解密得到秘密信息为 m'，上述符号代表的含义和值如表 6 所列。

表 6 提取过程中相关符号的说明

Table 6 Description of related symbols in the extraction process

符号	含义	值
bs'	二进制序列	0101010100110010010001100111001101100100010001110101011001101011010110000011000100111000010110001011011100111001001100100111010110101010100110000011110000111011001001011001100010011000001001111010010011011010011011001110010010010010110110001001010011001000011011101001000011110100011000101110100000110111001100110110101001100100111011100110000000111101
c'	密文	U2FsdGVkX18Y792kU0xvK10OIm6rIlJd7Hz1h73jro0=
m'	秘密信息	secret

4.2 实验分析

(1)抗检测性。抗检测性指秘密信息和载体信息在统计特征上的一致性，使攻击者在检测秘密信息时十分困难。设 Σ 为一个信息隐藏系统，P 表示嵌入秘密信息后的载体信息概率分布，Q 表示没有嵌入秘密信息时的载体信息概率分布。用相对熵 $D(P\|Q)$衡量两个概率分布的相似性，即相对熵越小，P 和 Q 两个概率分布越相似，P 和 Q 在统计特征上的一致性就越高。相对熵 $D(P\|Q)$的公式如下：

$$D(P \| Q)=\sum_i p_i \log \frac{p_i}{q_i} \tag{3}$$

若存在一个 ε，使得 $D(P \| Q)\leqslant\varepsilon$，则说明该系统是抗检测的，$\varepsilon$ 越小，抗检测能力就越强。

以实验中的 T_2 为例，嵌入秘密信息前后载体信息的十六进制对比如表 7 所列。本实验中假设 $\varepsilon=0.01$，用某个字符出现的次数与总字符数的比值表示某个字符的概率，则 P 和 Q 的概率分布如表 8 所列。计算得 $D(P \| Q)=0.00341$，即存在一个 ε，满足 $D(P \| Q)\leqslant\varepsilon$，因此本文模型是抗检测的，攻击者无法将普通交易和含有秘密信息的交易区分出来。

表 7　嵌入秘密信息前后载体信息的十六进制对比

Table 7　Hexadecimal comparison of carrier information before and after embedding secret information

嵌入秘密信息之前	嵌入秘密信息之后
e697a5e69c9fefbc9a32303139e5b9b431e69c8832e697a50d0ae59cb0e782b9efbc9ae6b2b3e58d97efbc8ce98391e5b79e0d0ae697a5e69c80e9ab98e6b094e6b8a9efbc9a34e69184e6b08fe5baa60d0ae697a5e69c80e4bd8ee6b094e6b8a9efbc9a2d372e39e69184e6b08fe5baa6	e697a5e69c9fefbc9a32303139e5b9b431e69c8832e697a5**2020**0d0ae59cb0e782b9efbc9ae6b2b3e58d97efbc8ce98391e5b79e**20**0d0ae697a5e69c80e9ab98e6b094e6b8a9efbc9a34e69184e6b08fe5baa60d0ae697a5e69c80e4bd8ee6b094e6b8a9efbc9a2d372e39e69184e6b08fe5baa6

表 8　P 和 Q 的概率分布

Table 8　Probability distributions of P and Q

i	P_i	q_i
0	0.0619	0.0733
1	0.0221	0.0215
2	0.0265	0.0388
3	0.0487	0.0474
4	0.0309	0.0302
5	0.0442	0.0431
6	0.0841	0.0819
7	0.0354	0.0345
8	0.0708	0.0689
9	0.1327	0.1293
a	0.0796	0.0776
b	0.0973	0.0948
c	0.0487	0.0474
d	0.0265	0.0259
e	0.1549	0.1508
f	0.0354	0.0345

(2)鲁棒性。鲁棒性指在对载体信息进行操作时，若载体信息发生失真，不会造成秘密信息的丢失，秘密信息保持一定的完整性，并且能以一定的准确率获取到，因此可以用秘密信息的丢失率来代表鲁棒性[23]。假设 L 代表可嵌入秘密信息的位数，TC 代表载体信息的长度，LP 表示丢失的秘密信息位数在载体信息中所占比例，则鲁棒性用 $R=1-LP$ 表示，其中 $LP=L/TC$。

文献[10]通过修改交易地址的最低有效位来嵌入 1 位秘密信息，交易地址长度为 34 位，则 $L=1$，$TC=34$，鲁棒性 $R=1-1/34=97.05\%$。

本文通过在 txt 文件中的行尾插入空格来嵌入秘密信息，以表 5 的十六进制为例，嵌入的空格为 6 位，实验中 txt 文件的十六进制为 226 位，则 $L=6$，$TC=226$，鲁棒性 $R=1-6/(226+6)=97.41\%$。不同模型的鲁棒性对比如表 9 所列。

表 9　不同模型的鲁棒性对比

Table 9　Comparison of robustness of different models

模型	鲁棒性 R/%
文献[10]	97.05
本文	97.41

(3)安全性。安全性指嵌入算法必须可以抵御一定程度的人为攻击，使秘密信息不被破坏。因为区块链隐蔽通信具有不可篡改性、不可伪造性、匿名性等特点，可以抵御一定的删除或更改等人为攻击，使秘密信息不被破坏，所以攻击者很难在不知道对称密钥和偏序关系的情况下解码原始的秘密信息。

(4)隐藏容量。隐藏容量指在不被人类感知系统发现的前提下，载体信息可以隐藏秘密信息的最大比特数。一方面，本文使用基于空格法的信息隐藏技术将秘密信息嵌入到纯文本文件中每一行的行尾，因为文本中有很多行，所以在一个纯文本文件中可以嵌入多位秘密信息。另一方面，发送方 Alice 可以在生成一个块的时间内多次提交交易，与在一个块内仅有一笔 Alice 提交的交易相比，在相同的时间提高了隐藏容量以及传输效率。不同模型的隐藏容量对比如表 10 所列。

表 10　不同模型的隐藏容量对比

Table 10　Hidden capacity comparison of different models

模型	1 笔交易携带的秘密信息比特数	一个区块中含有秘密信息的交易数
文献[10]	1	1
本文	$\geqslant 2$	$\geqslant 2$

结束语　本文提出一种面向纯文本信息隐藏的区块链隐蔽通信模型，在区块链网络中构建隐蔽信道，发送方将秘密信息采用基于空格法的信息隐藏技术嵌入到纯文本中，然后将纯文本文件在区块链网络中进行传输。受信方接收文件并通过嵌入算法的逆过程来提取秘密信息，实现了区块链网络下的隐蔽通信。实验结果表明，该模型具有较好的抗检测性、鲁棒性、安全性和较高的隐藏容量。

参 考 文 献

[1] NAHARUDDIN A, WIBAWA A D, SUMPENO S. A High Capacity and Imperceptible Text Steganography Using Binary Digit Mapping on ASCII Characters[C]// 2018 International Seminar on Intelligent Technology and Its Applications (ISITIA). Bali, Indonesia, 2018: 287-292.

[2] KATARIA S, KUMAR T, SINGH K, et al. ECR (encryption with cover text and reordering) based text steganography[C]// 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013). Shimla, 2013: 612-616.

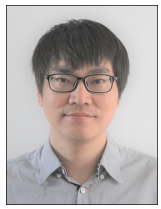
[3] WANG K, GAO Q. A Coverless Plain Text Steganography Based on Character Features[C]// IEEE Access. 2019: 95665-95676.

[4] WU G H, GONG L C, YUAN L F, et al. Review of information hiding on Chinese text[J]. Journal on Communications, 2019, 40(9): 145-156.

[5] ZHOU J J, YANG Z, NIU X X. Research on the detecting algorithm of text document information hiding[J]. Journal on Communications, 2004, 25(12): 97-101.

[6] WANG Y J, WU J Z, ZENG H T, et al. Covert channel research [J]. Journal of Software, 2010, 21(9): 2262-2288.
[7] NAKAMOTO S. Bitcoin: a peer-to-peer electronic cash system [EB/OL]. https://bitcoin.org/bitcoin.pdf.
[8] ZHENG Z, XIE S, DAI H, et al. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends[C]// IEEE International Congress on Big Data. IEEE, 2017.
[9] SASSON E B, CHIESA A, GARMAN C, et al. Zerocash: Decentralized Anonymous Payments from Bitcoin[C]// 2014 IEEE Symposium on Security and Privacy. IEEE, 2014: 459-474.
[10] PARTALA J. Provably Secure Covert Communication on Blockchain[J]. Cryptography, 2018.
[11] TIAN J, GOU G, LIU C, et al. DLchain: A Covert Channel over Blockchain Based on Dynamic Labels[C]// Information and Communications Security. Cham: Springer, 2019: 814-830.
[12] FIONOV A. Exploring Covert Channels in Bitcoin Transactions [C]// 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). Novosibirsk, Russia, 2019: 59-64.
[13] LI Y F, DING L P, WU J Z, et al. Research on a new network covert channel model in blockchain environment[J]. Journal on Communications, 2019, 40(5): 67-78.
[14] CHRISTIDIS K, DEVETSIKIOTIS M. Blockchains and Smart Contracts for the Internet of Things[J]. IEEE Access, 2016, 4: 2292-2303.
[15] HALPIN H, PIEKARSKA M. Introduction to Security and Privacy on the Blockchain[C]// IEEE European Symposium on Security and Privacy Workshops. IEEE, 2017: 1-3.
[16] SHEN X, PEI Q Q, LIU X F. Survey of blockchain[J]. Chinese Journal of Network and Information Security, 2016, 2(11): 11-20.
[17] YUAN Y, WANG F Y. Blockchain: the state of the art and future trends[J]. ACTA Automatica Sinica, 2016, 42(4): 481-494.
[18] LAMPSON B W. A note on the confinement problem[M]. Communications of the ACM, 1973, 16(10): 613-615.
[19] WENDZEL S, ZANDER S, FECHNER B, et al. Pattern-based survey and categorization of network covert channel techniques [J]. ACM Computing Surveys, 2015, 47(3): 1-26.
[20] KAMARUDDIN N S, KAMSIN A, POR L Y, et al. A review of text watermarking: theory, methods and applications[J]. IEEE Access, 2018, 6(1): 8011-8028.
[21] QI W F, LI X L, YANG B, et al. Document watermarking scheme for information tracking[J]. Journal on Communications, 2008, 29(10): 183-190.
[22] JIANG C X, CHEN X W, LI Z. Robust text watermarking based on significant components[J]. Acta Automatica Sinica, 2010, 36(9): 1250-1256.
[23] CHEN Y N, LI Q M, LV C X, et al. Robust Text Watermarking Based on Significant Components[J]. Cyberspace Security, 2019, 10(5): 88-96.

SHE Wei, born in 1977, Ph. D, professor, doctorial supervisor, is a member of China Computer Federation. His main research interests include blockchain technology, information security and trusted distributed system.



TIAN Zhao, born in 1985, Ph. D, lecturer, is a member of China Computer Federation. His main interests include blockchain technology and trusted distributed system.

(责任编辑:喻藜)