基于比特币系统的隐蔽通信技术

吕婧淑, 操晓春

(中国科学院信息工程研究所 信息安全国家重点实验室, 北京 100093)

摘要 为了满足日益增多且机密性要求很高的情报传输需求, 急需提出新的隐蔽通信方式。隐蔽通信技术需要满足的性能包括安全性和隐藏率。与此同时, 区块链技术具有去中心化、匿名性、可追溯且分布式记账等特点, 比特币是最为经典且普及的区块链货币应用。因此, 本文提出了利用比特币的 P2P 网络广播机制和交易机制构造了 BDTX(Broadcast-Transaction, 广播-交易)隐蔽信道; 利用比特币的节点连接机制构造 ADDR(ADDR, 地址广播)隐蔽信道。对两种隐蔽信道分别进行了详细介绍, 分析了两种隐蔽信道的安全性和隐藏率, 并将其与传统 IP 隐蔽信道进行对比。两种隐蔽信道的安全性相比传统 IP 网络隐蔽信道都有提高。

关键词 隐蔽信道; 区块链; P2P 网络; 比特币; 情报传输
中图法分类号 TP399 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2021.03.10

Covert Communication Technology Based on Bitcoin

LV Jingshu, CAO Xiaochun

(State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China)

Abstract In order to meet the increasing demands for information transmission with highconfidentiality requirements, it is urgent to propose new covert anonymous communication methods. Covert Anonymous communication technology needs to meet the performances including security and hiding rate. At the same time, blockchain technology is characterized by decentralization, anonymity, traceability and distributed accounting. Bitcoin is the most well-known and popular blockchain currency application. Therefore, this paper proposes to use Bitcoin's P2P network broadcasting and transaction mechanism to construct BDTX (Broadcast-Transaction) secret channel, use Bitcoin node connection mechanism to construct the ADDR (Version-ADDR) secret channel. Two kinds of secret channels are introduced in detail, and the security, hiding rate and real-time performance of the two hidden channels are analyzed. The security of the two hidden channels has been improved compared to the traditional IP network covert channels.

Key words covert channel; Blockchain; Bitcoin; P2P network; secret information transmission

1 引言

近年来, 区块链[1](blockchain)作为一种去中心化、不可篡改、可追溯、多方共同维护分布式数据库的技术, 被广泛地应用于金融、身份认证、公示公证等领域[2]。区块链源自化名为“中本聪”的学者提出的比特币[3]系统的底层技术。

区块链有去中心化、匿名性、可追溯和安全可信的特点[4]。1)去中心化: 区块链中数据的新建、验证、更新等操作都是通过 P2P 网络的传播和数学方法的计算完成, 而无需中心化机构的管理和参与; 2)匿名性: 在区块链上进行交易的节点是通过一个公钥地址完成的, 并不知道对方在现实世界的真实身份, 因此区块链具有很强的匿名性; 3)可追溯: 数据一旦被验证正确, 所有区块都会记录该数据及其时间戳, 因此区块链具有很强的可追溯和不可抵赖性; 4)安全可信: 采用哈希算法对区块链应用的交易数据、区块数据等加密以保证数据的安全性、采用数字签名技术对交易数据进行签名以保证其他节点能够验证该数据的真实性; 同时借助分布式系统各节点的工作量证明等共识算法形成的强大的计算能力来抵御外部攻击, 从而保证区块链数据不可篡改和不可伪造。

基于区块链开发的比特币是区块链技术最广为

通讯作者: 操晓春, 博士, 教授, 是计算机学会(CCF)会员(12395D), 主要研究领域为多媒体内容安全、计算机视觉. E-mail: caoxiaochun@iie.ac.cn.

本课题得到国家重点研发计划(No. 2016YFB0800603)资助。
收稿日期: 2019-04-02; 修改日期: 2019-04-26; 定稿日期: 2020-12-21

人知的应用, 以下简称为比特币。比特币本质上是由点对点(P2P)网络系统生成的数字货币, 其发行过程不依赖特定的中心化机构, 而是依赖于 P2P 网络中所有节点共同参与一种称为工作量证明(Proof of work, POW)的共识过程以完成交易的验证与记录。POW 共识过程(俗称“挖矿”, 每个节点称为“矿工”)是各节点为了获得新区块的记账权使用自己的计算资源来竞争解决一个难度可动态调整的数学问题。难度值通常为含有很多个前置零的哈希值, 各节点不断地对区块头部的数据进行哈希计算, 若某节点计算出的哈希值小于等于难度值则成功解决该数学问题, 获得新区块的记账权并向全网广播。其余所有节点验证通过后将当前时段所有比特币的交易打包记入新区块、按照时间顺序链接到比特币主链上。比特币的第一个区块(称为创世区块)诞生于 2009 年 1 月 4 日, 由创始人“中本聪”持有。一周后, “中本聪”向密码学专家哈尔芬尼发送了 10 个比特币(此处的比特币为货币单位), 成为历史上第一次比特币交易。

信息隐藏模型[5]最早源自 Simmons 提出的囚徒问题: Alice 和 Bob 是狱友, 他们想谋划一个越狱计划, 但又不能让监视他们的典狱官 Wendy 发现。这个问题可以被抽象为: A 想向 B 秘密传递一些消息, 因此 A 需要选择一个看似平常的消息 R(当 R 在网络中公开传输时不会引起怀疑, 称消息 R 为载体), 将秘密信息 m 加密后嵌入载体 R 中, 此时载体 R 变成隐秘对象 S, 并且 S 尽可能地保持原有载体 S 的特征不变, 使得攻击者在仅知 R 的表面消息时无法检测到秘密消息 m。这样, 就实现了信息的隐蔽传输。

结合囚犯问题和上述模型的分析, 信息隐藏的概念可以定义为将秘密信息隐藏于某个公开的载体中传输并不被接收者以外的人发现。根据信息隐藏技术的目的和载体不同, 其可以分为四个分支: 隐写术[6]、匿名通信[7]、隐蔽信道[8]、数字水印[9]。其中, 隐写术是一种将秘密信息隐藏于某种不被怀疑的载体中实现隐蔽通信的技术; 匿名通信指采取一定的措施来隐藏通信流中的通信关系, 使窃听者难以获取或辨别通信双方的关系及内容。近年来, 基于信息隐藏中隐写术和匿名通信两个分支发展出的隐蔽匿名通信[10]技术在多媒体、网络、信道等载体中都有应用。隐蔽通信技术具有安全、匿名、隐藏容量的特点: 1)安全性: 对普通用户来说, 隐藏过程是无法“感受”到的, 对攻击者来说, 载体的特征在统计角度上不可分辨; 2)匿名性: 隐蔽传输的发送方和接收方的真实身份对普通用户和攻击者来说都是未知的或很难推断的; 3)隐藏容量: 是指载体所能承载的秘密信息的数量, 即秘密信息与载体信息大小的比例, 一般希望这个比例越大越好。隐蔽信道的传输介质分为存储类型和时间类型, 由此衍生出存储型隐蔽信道和时分型隐蔽信道。操作系统、数据库和网络系统中都可以构造隐蔽信道, 本文主要讨论基于网络的隐蔽信道。

然而, 现有的隐蔽通信技术存在一定的问题: 1)存储型隐蔽信道通常使用网络协议包的冗余字段进行信息隐藏, 其安全性受所嵌协议包通信频率的影响, 若两个进程通过该协议通信的频率很低, 那么隐藏在网络协议中的秘密信息很容易被发现; 2)通信双方的匿名性较难保证, 一旦被检测到利用网络协议隐藏信息, 便能够通过双方 IP 地址收集与其在同一地址内移动端的定位数据从挖掘双方现实生活中的实际地址, 从而造成隐私泄露; 3)时分型隐蔽信道受当前网络环境和设定的时间间隔影响很大: 对基于分组交换的 IP 网络而言, 数据分组因为选取不同的路由而引入不同的处理延时, 若延时超过了设置的时间间隔, 就会出现误码, 即本应为“1”的消息被接收方误认为是“0”。

可以看出, 区块链技术在一定程度上满足隐蔽通信的特点, 1)安全可信: 采用哈希算法对区块链应用的交易数据、区块数据等加密以保证数据传输及存储的安全性。区块链天然的加密机制保障了存储型隐蔽信道所隐藏内容的安全性; 2)去中心化: 新区快、新交易的产生不需要经过中心化机构的审核, 所有数据保存在分布于全球的完整节点。新区快、新交易产生时会向全网广播消息, 而不会通过中心化机构转发, 因此去中心化特点提高了隐蔽信道的不可感知性; 3)假名性: 以比特币为例, 每个节点在网络中交易使用的是其公钥生成的地址, 因此无法获得节点的真实身份, 保护了隐蔽信道通信双方的隐私。这些特点使得区块链技术很适合隐蔽通信的场景。因此本文要解决的问题是将区块链某些技术层作为隐蔽匿名通信的载体传输秘密消息。具体来说, 是在区块链的网络层和应用层构建隐蔽信道。

然而, 正是由于区块链的可追溯、不可抵赖等特性, 使得隐蔽传输的难度相对普通载体变大——1)新区块的产生和所有交易信息都会在区块链 P2P 网络中广播, 因此所有节点都能够查看到新区块和新交易的内容; 2)上述信息不仅会在网络中广播, 而且还会被记录在区块体中, 而区块链是分布式记账的系统, 因为这些信息会永久存储在所有节点的账本中。

那么隐蔽匿名通信的双方若要使用基于存储的信息隐藏算法, 隐藏的信息会随着新区块和新交易实时地被全网节点获知, 并且任何节点都能随时查看区块记录的之前的信息, 但通信双方并不希望其他节点发现, 因此这对隐藏算法的要求很高。

2 相关工作

2.1 网络隐蔽通信

2.1.1 网络隐蔽通信的背景

隐蔽通信[12]的概念首先由 Lampson 提出。隐蔽通信是允许一方以不违反系统的安全策略的方式传送信息到另一方, 是一种能抵抗审查的通信方式。在互联网开放环境下, 隐蔽通信也能使用网络中数据流等载体来实现。因为所有通过网络的信息都以数据流的形式传输, 数据流传输会通过不同网络拓扑, 数据流由这些网络节点共享, 如果以网络数据流为载体, 采用隐写术是可以实现隐蔽通信的。隐写术是一门关于信息隐藏的科学, 信息隐藏指的是不让除预期的接收者之外的任何人知晓信息的传递事件或者信息的内容。现代隐写术利用信号或协议的冗余作为载体, 将秘密信息隐藏于其中, 以达到在不影响使用的前提下掩盖传输秘密信息及其存在的目的。

2.1.2 网络隐蔽通信的类型

网络隐蔽通信[13]的基本方式是以网络数据流作为载体, 采用隐写术构建隐蔽信道, 是相对公开信道而言的。隐蔽信道是在公开信道中建立起来的一种进行隐蔽通信的信道, 该信道的存在仅为确定的收方所知。

网络隐蔽通信中的隐蔽信道可分为两种模式: 基于存储的隐蔽信道和基于时间的隐蔽信道。1)基于存储的隐蔽信道又称为存储型隐蔽信道, 秘密信息被嵌入到网络数据包的某些未用位或者载荷中, 随着网络包一起发送出去。2)基于时间的隐蔽信道又称为时分型隐蔽信道, 此种隐蔽信道利用数据包到达的顺序, 或者利用单位时间内是否有数据包到达传递信息。

其中, 存储型隐蔽信道主要在网络层、传输层和应用层采用隐写术建立, 即在各层的常用协议中的某些冗余字段嵌入信息, 既不会影响正常的网络通信, 又能达到发送秘密信息的目的。例如, 网络层经常使用的 IP 协议中的服务类型、标识 ID、标志字段等字段在特定情况下都是空置字段, 可以被用来构造隐蔽信道; 传输层常使用的 TCP 协议中的序号、确认号、标志和紧急指针等字段可以被用户构造隐蔽信道。应用层中的 HTTP 协议利用请求报文的请求首部 Accept 域中选项参数之间的排列位置来编码隐藏信息。

2.2 区块链系统

从技术角度看, 区块链系统由数据层、网络层、共识层、激励层和应用层组成, 如图 1 所示。其中, 数据层包括了区块, 区块头部包含了时间戳、难度值, 区块体包含了交易数据, 区块之间通过哈希算法计算后的链式结构相连; 从网络层可以看出, 区块链系统是建立在对等网络(P2P 网络)和 TCP/IP 协议上的分布式系统, 新区块和新交易在这样的网络上有其独特的广播机制; 在共识层上, 节点通过使用数据层中的哈希算法和动态变化的难度值来确保新区块产生的公平性和链条的一致性。一旦数据被记录在区块内, 若要篡改需要重新按照共识机制计算一遍所有的区块内容, 因此只要诚实节点占比超过 50%, 篡改的速度就会比最长链的增长速度慢。在应用层: 1)区块链系统中每个节点共享同一份账本数据的特点使其很适合重要数据的存储; 2)区块链的不可篡改性和可追溯性使其很适合用户身份、企业各类许可证和执照的验证; 3)区块链去中心化的特点使其能够应用于股权众筹、P2P 网络借贷等商业模式, 而不需要第三方中介机构, 无需交易双方的信用以节省成本; 4)区块链的共识机制和不可篡改性保证了其在公共事务中的公平和透明, 因此也可以应用于政治选举、股东投票等。

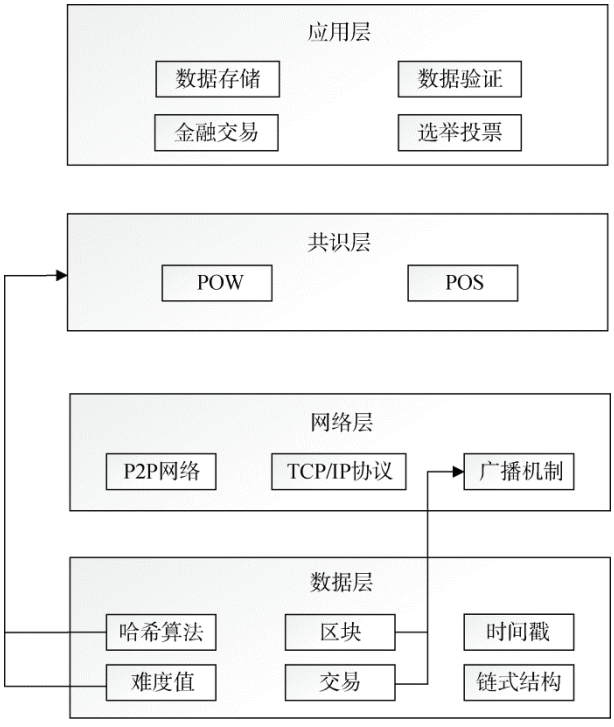

图 1 区块链技术框架图

Figure 1 Blockchain Technology Framework

2.3 比特币网络

比特币是基于区块链技术的虚拟货币应用, 属于图 1 的应用层。它本质上不依赖中心机构, 而是由去中心化的点对点网络生成的虚拟货币, 并由该网络中所有节点共同记账、维护账本。比特币发行的方式就是由上述 POW[11]机制决定的。中本聪的白皮书定义的 POW: P2P 分布式网络中所有节点搜索一个随机数能够在特定公式计算下满足难度值。一旦某个节点花费 CPU 资源找到一个随机数满足了 POW 的公式, 便成为新区块的拥有者, 并创建该区块的初始交易(包含 coinbase 字段的交易)。随后, 该节点将新区块和初始交易打包通过比特币特有的广播机制向全网节点广播。本文提出的隐蔽信道就是基于比特币 P2P 网络的广播机制、交易机制和节点间通信机制构造的。因此, 本节将详细介绍比特币网络中的广播机制、交易机制和节点的通信机制。

2.3.1 广播机制

比特币网络中, 每一个新区块、新交易产生后都需要向周围的节点广播, 得到足够数量的验证后新区块才能被加到已有区块链上、新交易才能被打包加入到现有区块中。为了避免重复发送节点已经收到的区块消息或交易消息, 这些消息的广播过程并不是直接发送, 而是如图 2 所示:

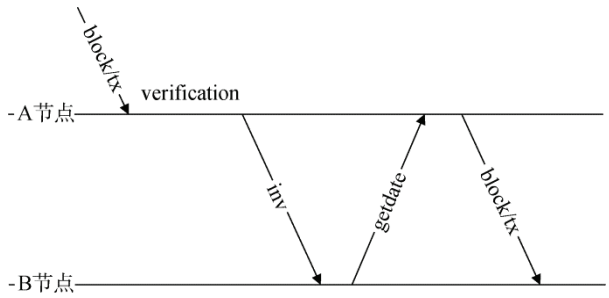

图 2 比特币网络广播机制示意图

Figure 2 Schematic Diagram of Bitcoin Network Broadcasting Mechanism

节点 A 和节点 B 是邻居节点。节点 A 收到了一个新区块或新交易产生的消息, 首先 A 要对新区块或交易进行验证, 完全验证通过后向节点 B 发送一个 inv 消息(inv 消息包含了一系列 A 收到的交易散列值和区块散列值, 它们现在都是可被请求的状态); 节点 B 收到的 inv 消息中若有交易和区块散列值还未存储在 B 本地, 那么节点 B 将会向节点 A 发送一个 getdata 消息(getdata 消息包含了节点 B 未存储的区块或交易的散列值)。最终, 节点 A 发送给节点 B 请求的区块或交易整体数据。

节点通过 inv 消息可以广播(advertise)它所拥有的对象信息。inv 消息可以主动发送, 也可以用于应答 getblocks 消息。其具体的数据格式如下表所示。其中, inv 消息负载的最大长度为 50000 字节。inventory 数据为区块或交易的散列值。

表 1 inv 消息的数据格式

Table 1 Data Structure of inv Message

字段尺寸	描述	数据类型	说明
?	count	var_int	清单(inventory)数量
36*?	inventory	inv_vect[]	清单(inventory)数据

getdata 用于应答 inv 消息来获取指定对象, 它通常在接收到 inv 包后滤去 inventory 数据中的已知元素, 并将其发送给 inv 消息的发送者。它的数据格式与 inv 消息相同。并且, getdata 消息负载的最大长度为 50000 字节。

表 2 getdata 消息的数据格式

Table 2 Data Structure of getdata Message

字段尺寸	描述	数据类型	说明
?	count	var_int	清单(inventory)数量
36*?	inventory	inv_vect[]	清单(inventory)数据

inv 消息和 getdata 消息数据格式中共有的清单向量(Inventory vector)是用于告知其他节点本节点拥有的对象或请求的数据。

表 3 清单向量的数据格式

Table 3 Data Structure of inventory vector

字段尺寸	描述	数据类型	说明
4	type	uint32_t	对象类型标识
32	hash	char[32]	对象散列值

清单向量数据格式中第一个字段, 即对象类型(type)标识已经定义如下 3 个值。第二个字段, 即对象散列值(hash)为区块或交易的散列值。而这里的 hash 为区块或交易 ID 经过 SHA256 算法的结果, 共 256 位, 占存储 32 字节。其中, 类型值为 0 清单向量 hash 字段为可忽略的错误数据; 类型值为 1, hash 字段存储的是关于交易的哈希值; 类型值为 2 时, hash 字段存储的是关于区块的哈希值。清单向量的类型字段说明如下表所示:

表 4 清单向量的类型值

Table 4 Type Value of inventory vector

值	名称	说明
0	ERROR	数据可忽略
1	MSG_TX	散列是关于交易的
2	MSG_BLOCK	散列是关于数据块的

2.3.2 交易机制

比特币交易数据的格式如下表所示。一个完整的交易数据包括了 version(协议版本号)、输入数量、tx_in(输入列表)、输出数量、tx_out(输出列表)和锁定时间。其中, 输入列表包括了 prev output(引用交易的散列值)、index(前向交易的索引)、script length(解锁脚本长度)、scriptsig(解锁脚本)、sequence(交易序列号)。输出列表包括了 value(每笔交易输出的比特币量)、pk_script length(锁定脚本的长度)、pk_script(锁定脚本)和 lock time(锁定时间)。

表 5 交易消息的数据格式

Table 5 Data Structure of tx Message

字段名称	大小	数组构成	含义
version	4 字节	无	交易参照的协议版本号
tx_in count	1+	无	被包含的输入交易的数量
tx_in	41+	prev output index script length scriptsig sequence	一个或多个输入交易构成的数组
tx_out_count	1+	无	被包含的输出交易的数量
tx_out	8+	value pk_script length pk_script	一个或多个输出交易构成的数组
lock_time	4 字节	无	是否立即执行/区块高度/Unix 时间戳

其中, 有两种情况可以嵌入数据: 1)比特币每 10min 产生一个新区快, 此时会创建该区块的 coinbase 交易(初始交易)。此时, tx_in(交易输入列表)中的 coinbase 字段可以嵌入自定义数据。2)tx_out(交易输出列表)中的 value 为 8 个字节的零值, 即交易输出比特币数量为零, 则表明这不是普通的转账交易, 而是追加的备注信息, 可嵌入自定义内容。

2.3.3 通信机制

比特币网络中节点的通信分为两个阶段: 第一个阶段通过 TCP 三次握手协议建立连接, 第二个阶段为建立连接后广播消息的互传及每天向邻居节点发送 ADDR 消息。本文使用 ADDR 消息构造隐蔽信道, 因此本节重点介绍节点通信机制中的 ADDR 消息。ADDR 消息包含最多 1000 个 IP 地址及其时间戳, 用于从邻居节点中获取已知的活动节点。只有在节点建立与其邻居节点的传出连接后, 才能发起 ADDR 消息的请求, 即 getaddr 消息, 但这个消息没有附加数据。ADDR 消息在两种情况下会发送: 1)每天, 每个节点会向其邻居节点发送一条 ADDR 消息; 2)当一个节点接收到一条包含地址数量不超过 10 个的 ADDR 消息时, 它将 ADDR 消息转发给两个随机选择的已连接的邻居节点。每个 ADDR 消息包含节点从其路由表中随机选择的最多 1000 个地址。该消息的数据结构如表 6 所示。

表 6 ADDR 消息的数据结构

Table 6 Data Structure of ADDR Message

字段尺寸	描述	数据类型	说明
24	message header	char[24]	消息头部
1000	payload	uint64_t	消息包含的地址数量
30*?	addr_list	(uint32_t + net_addr)	其他节点的地址列表: 时间戳+service 类型+地址+端口号

3 隐蔽传输框架

本文以比特币广播和交易消息作为隐蔽传输的载体, 在比特币网络上搭建隐蔽传输框架。需要说明的是, 网络的广播、交易和通信机制只有比特币全节点才会参与, 因此本文的框架是搭建于全节点参与的 P2P 网络。该框架可以分为数据层、网络层、共识层和应用层四层。其中, 网络层的节点通信机制和广播机制传输的消息中有可嵌入数据的字段; 应用层的交易信息也有可嵌入数据的字段。因此, 本文的隐蔽传输框架包含两个: (1)利用应用层交易机制、网络层广播机制两层的消息及其之间的关系设计的 BDTX(broadcast-transaction, 广播-交易)隐蔽信道; (2)利用网络层节点通信机制设计了 ADDR(地址广播)隐蔽信道。隐蔽传输的整体框架如图 3 所示。

4 隐蔽传输信道

本节将分别详细地阐述两种隐蔽信道的细节并分析其原理。

4.1 BDTX 隐蔽信道

新交易或新区块产生后, 网络层会通过 inv 消息和 getdata 消息来进行有效的广播(广播过程如图 2 所

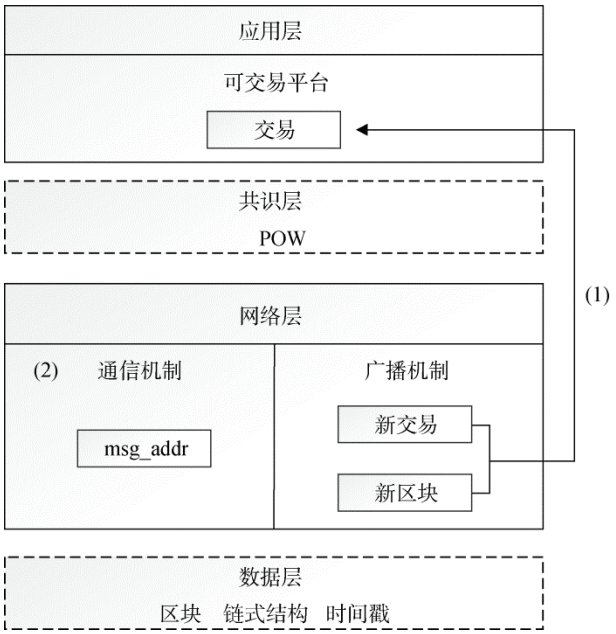

图 3 隐蔽传输框架图

Figure 3 Covert Transmission Frame Diagram

示)。两个消息都包含清单向量(结构如表 3 所示), 清单向量中 type 字段为 1 时, hash 字段为新产生的交易哈希值。清单向量中的交易哈希值与接收节点已验证过的交易有索引关系。因此, 本节通过两者的索引关系构建如图 4 所示的 BDTX 隐蔽信道。

从图 4 可以看出, BDTX 隐蔽信道是由比特币广播模块、交易模块、嵌入模块和提取模块四个部分构成。其中, 广播模块和交易模块分别已在 2.3.1 和 2.3.2 节介绍过。嵌入模块包括加密、插入索引和密文分片三个功能。加密是将秘密信息转换成二进制信息后使用 AES 算法[14]进行加密; 插入索引是将嵌入密文的交易哈希值作为索引插入广播模块的清单向量中; 密文分片将密文按照双方预先约定好的协议分片并插入交易模块的 tx_in 中。分片协议为将密文平均分成 x 片, 每片 y 比特, 将 x 片密文以相同长度间隔地插入 tx_in 的 coinbase 字段, 即每间隔 z 比特插入一个分片。需要特别说明的是, 含有 coinbase 字段的交易仅在新区块产生时会创建, 即每 10min 才会广播一个含有 coinbase 字段的交易哈希值。提取模块的作用是将密文按照上述分片协议复原出完整密文、解密、恢复二进制秘密消息。

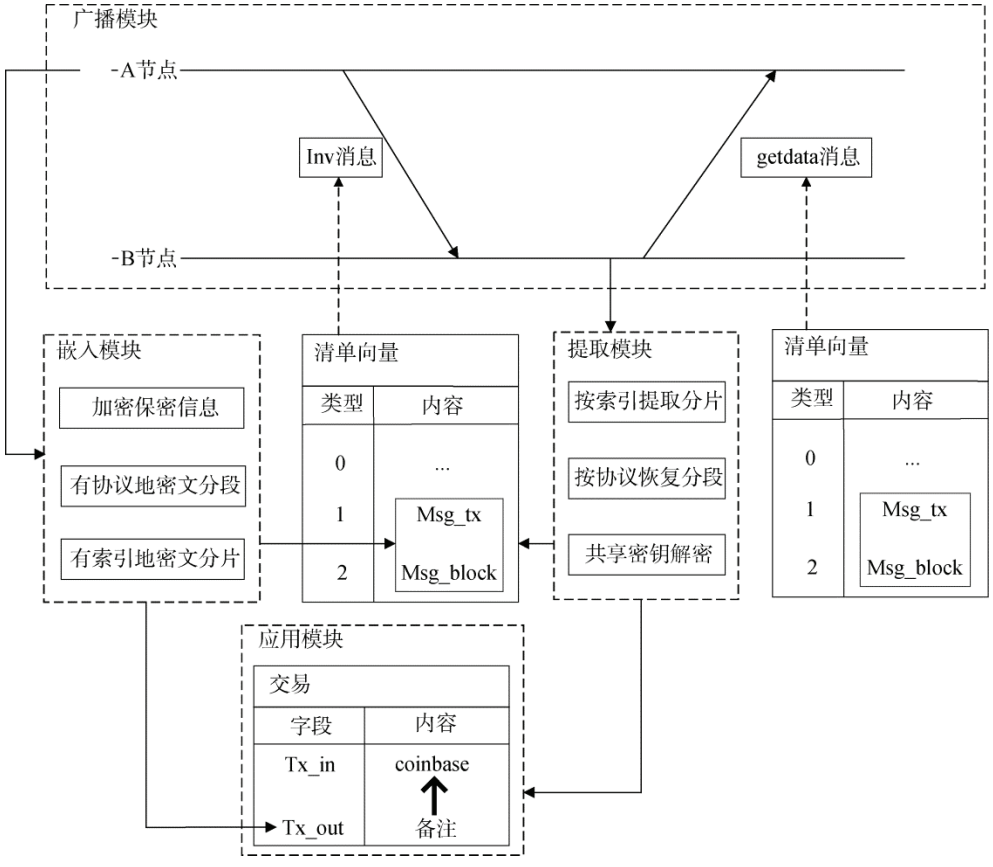

图 4 BDTX 隐蔽信道示意图

Figure 4 BDTX Covert Channel Schematic

对于两个已经建立连接的节点 A、B 来说, 由于使用 AES 加解密算法, 因此假设双方已共享该算法的密钥。一次秘密信息的传输流程如下: 1)A 节点使用网络层广播机制向 B 节点发送 inv 消息, 该消息的清单向量的类型值为 1 或 2, 即清单向量的内容分别为一系列交易哈希值或区块哈希值; 2)A 节点使用嵌入模块将加密后的秘密信息按照分片协议分片并嵌入交易消息的 tx_in(交易输入列表)的 coinbase 字段; 3)B 节点收到 inv 消息后, 读取清单向量中的类型值, 使用提取模块恢复秘密信息: 以清单向量中协议位置的交易哈希值为索引在交易消息中搜索是否存在相等的交易哈希值; 若存在, 找到该交易 tx_in 的 coinbase 字段被分片的密文、按照分片协议恢复密文、用约定好的密钥解密; 4)若 B 节点成功完成上一步的提取, 则返回给 A 节点正常的 getdata 消息, 即滤去索引哈希值; 若并未成功提取出秘密信息, 则返回给 A 节点仍包含索引哈希值的 getdata 消息。

4.2 ADDR 隐蔽信道

由于 ADDR 消息能够隐藏信息的存储空间相对有限, 但其相对其他消息具有较强的时分性——在节点传出连接数量未达到上限时, 每天每个节点都要向其邻居节点广播 ADDR 消息。因此, 基于该特点设计了 ADDR 隐蔽信道, 如图 5 所示。

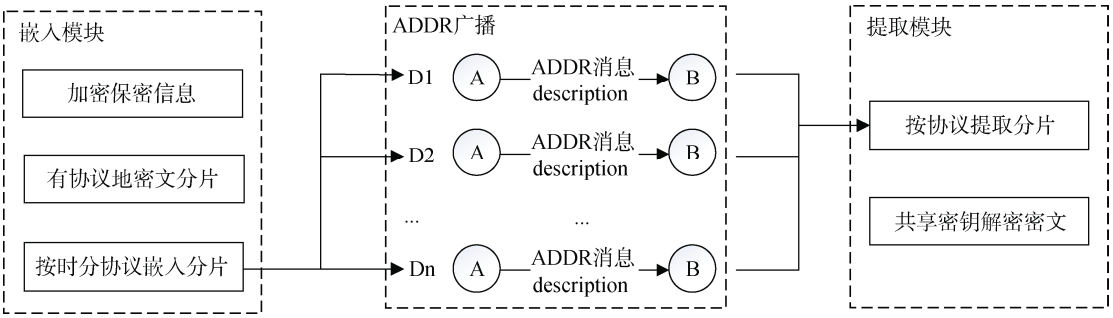

图 5 ADDR 隐蔽信道示意图

Figure 5 ADDR Covert Channel Schematic

如图 5 所示, ADDR 隐蔽信道分为三个模块: ADDR 广播模块、嵌入模块和提取模块。其中, ADDR 广播模块的原理已在 2.3.3 节中阐述过。嵌入模块的作用是将秘密信息转换成二进制消息后使用 AES 算法加密、按照分片协议分割密文、按照时分协议将分割后的密文片段嵌入 ADDR 消息中。分片协议为将密文分为 x 片,每片 y 比特。时分协议为通信双方提前约定 y 个地址作为标志位, 并约定好 y 个标志地址的顺序。若第y_i比特为“1”, 则在当日 ADDR 消息中嵌入对应顺序的标志地址; 若第y_i比特为“0”, 则在当日 ADDR 消息中不嵌入对应顺序的标志地址。提取模块的作用是将密文按照时分协议复原完整的密文、解密恢复二进制消息。

对于已建立网络连接的两个节点 A、B 来说, A 节点是秘密信息的发送方, B 节点是秘密信息的接收方。一次秘密信息的传输过程如下: 1)A、B 提前约定分片数量与每片的比特数, 及与每片比特数相等数量的特定地址作为判断分片内容的标志位; 2)A 每天都向包括 B 在内的邻居节点发送 ADDR 消息, 按照时分协议根据分片的每个比特决定 ADDR 消息中是否包含对应的特定地址, 即分片内容为“1”时 ADDR 消息包含对应顺序的特定地址、分片内容为“0”时 ADDR 消息不包含对应顺序的特定地址; 3)B 节点收到 ADDR 消息后根据是否包含特定地址复原密文、用提前约定的密钥解密恢复秘密信息。

5 隐蔽信道算法分析

隐蔽信道需要满足两个性能要求: 安全性和有一定的隐藏容量。首先, 安全性是传输秘密信息最重要的特性, 指的是对普通节点来说隐藏过程是无法感知到的; 对攻击者来说, 载体的特征在统计上不可分辨, 即如何保证不被攻击者发现且不影响其他节点通信的情况下传输秘密信息。其次, 隐藏容量指的是载体所能承载的秘密信息的数量, 它决定了隐秘传输的效率。本节将从以上两个性能方面分别分析本文提出的 BDTX 和 ADDR 隐蔽信道。

5.1 安全性

隐蔽信道的安全性分为不可感知性和不可检测性。对普通节点来说, BDTX 隐蔽信道利用的广播消息和交易消息分别是由比特币网络协议、分布式记账规则规定的, 即新区块或新交易产生后的广播机制、任何一笔交易都同步记录在所有节点的账本中。

ADDR 隐蔽信道利用的 ADDR 消息是网络广播协议规定的, 即每当节点收到连接请求时都要立即宣告其版本和每天节点都要向邻居节点广播其地址消息。因此, BDTX 和 BDTX 道并未修改 inv 消息、getdata 消息的数据格式和内容本身, 只是利用其内容作为索引和反馈, 并且新区快初始交易信息中的 coinbase 字段是可以允许节点写入数据的。因此, 对普通节点来说, BDTX 和 ADDR 隐蔽信道都是不可感知的。

在 BDTX 隐蔽信道中使用了 inv 消息中清单向量的 hash 值作为索引, 由于 inv 消息的负载最大长度为 50000 字节, 除去 count 字段外, 最多存储 1387 个清单向量(每个清单向量长度为 36 字节), 因此攻击者需要从 1387 个 inventory 向量中遍历找到索引(约定好某个位置的交易哈希值)。因此, 攻击者在清单向量找到索引需要尝试的次数为 1387, 在嵌入模块中将每条密文都分为 x 片, 每片有 y 位, 再按固定间隔 z 比特嵌入, 同时假设攻击者破解加密密文的难度为 α。交易信息的 coinbase 字段占 69 字节, 即 552 比特, 密文长度为$x \times y$比特, 可选择嵌入分片的位置个数为$\left(\frac{552-x\times y}{z}-1\right)$, 因此攻击者在这些位置中找到正确分片需要遍历的次数为$C_{\frac{552-x\times y}{z}}^{x}-1$。

对 ADDR 隐蔽信道来说, 假设密文分为 x 片每片为 y 比特, 需要耗时x天进行传输。每条 ADDR 消息可能包含 0~1000 条地址, 因此假设实际传输过程中每条 ADDR 消息平均包含 n 条地址。假设攻击者持续监控发送节点的网络消息, 需要在 X 条 ADDR 消息中遍历找出特定的 x 条消息, 在每条 ADDR 消息遍历 n 条地址找出特定的 y 条地址。则共需要遍历 $C_X^x \times C_n^y$。

同时, 为了对比本文提出的三种隐蔽信道和传统隐蔽信道的安全性, 此处再对同样为网络层的 IP 协议内构造隐蔽信道的安全性做分析。IP 包能构造隐蔽信道的字段都集中在其 20 字节的首部, 如: 服务类型 ToS 的最后一位、数据包长度小于当前网络 MTU 时的 16 位标识 ID 和 3 位标志字段、数据包不分片时的 13 位分片偏移量。因此, 攻击者从首部 160 位中找出x片y位的密文需要尝试的次数如表格最后一列所示。

从表 7 可以看出, BDTX 相对 ADDR 和 IP 隐蔽信道的攻击难度是最大的, 安全性最高, 其中 IP 隐蔽信道的攻击难度最低。这是由于 BDTX 隐蔽信道以清单向量的交易 hash 值作为索引, 破解难度更大。并且本文提出的两种隐蔽信道的安全性都比传统的 IP 隐蔽信道高。

表 7　三种隐蔽信道与传统隐蔽信道的安全性对比

Table 7　Comparison of Security between Three Hidden Channels and Traditional Covert Channels

隐蔽信道	BDTX	ADDR	IP
攻击难度	$1387\times\prod_{i=1}^{x}\prod_{j=1}^{y}\prod_{k=0}^{z-1}C_{\frac{552-i\times j}{k}-1}^{i}$	$C_x^x\times C_n^y\times\prod_{i=1}^{x}\prod_{j=1}^{y}$	$C_{160-i\times j}^{j}\times\prod_{i=1}^{x}\prod_{j=1}^{y}$

5.2　隐藏容量

隐藏率的计算方式为秘密信息所占比特数与载体所占比特数之比。BDTX 隐蔽信道的载体为交易信息的 coinbase 字段, 占 69 字节, 即 552 比特; IP 隐蔽信道的载体为 IP 包内服务类型 ToS 字段的最后一位、16 位标识 ID、3 位标志字段和 13 位分片偏移量, 共计 33 比特。ADDR 的载体为 ADDR 消息的 addr_list 字段, 占 30 字节, 即 240 比特。

对于 BDTX 和 IP 两种隐蔽信道来说, 由于每种载体容量大小不同, 因此假设每种信道每片密文都为 y 比特, 分片间隔都为 z 比特, 而分片数量分别为 x_1、x_2。每个载体能存储的秘密信息容量为分片数乘以每片所占比特数, 即$x \times y$。两种信道的分片数量 x_1、x_2分别可以由公式(1)、(2)表示。

$$x_1=\frac{552-x_1\times y}{z}+1 \tag{1}$$

$$x_2=\frac{33-x_2\times y}{z}+1 \tag{2}$$

对于以上两个方程, 可以求解得出公式(3)、(4):

$$x_1=\frac{552+z}{z+y} \tag{3}$$

$$x_2=\frac{33+z}{z+y} \tag{4}$$

对于 ADDR 隐蔽信道来说, ADDR 消息最多包含 1000 个地址, 载体容量为 240bit, 而分片的每比特仅仅使用一个标志位地址, 即 0.24 比特。密文分片协议为将密文分为x_3片, 每片 y 比特(与以上两种信道每片比特数相同), 但其使用时分协议进行嵌入, 因此该信道的分片间隔为不是存储维度, 而是时间维度。因此, 可直接计算该隐蔽信道的隐藏率, 如公式(5)所示:

$$\frac{0.24\times y\times x_3}{240\times x_3}=0.001\times y \tag{5}$$

将 BDTX 和 IP 隐蔽信道的分片数量分别代入隐藏率的公式, 即分片量乘以每片比特数与载体比特数的比例; 并与已计算出的 ADDR 隐蔽信道的隐藏率进行对比。得到了如表 8 所示的隐藏率对比结果。

表 8 三种隐蔽信道与传统隐蔽信道的隐藏率对比

Table 8 Comparison of Ratio between Three Hidden Channels and Traditional Covert Channels

隐蔽信道	BDTX	ADDR	IP
隐藏率	$\frac{\frac{552+z}{z+y}\times y}{552}$	0.001	$\frac{\frac{33+z}{z+y}\times y}{33}$

将三种隐蔽信道的隐藏率两两相比, IP 隐蔽信道的隐藏率是最高的, 然后依次是 BDTX、ADDR 隐蔽信道。而在安全性分析中, 即 BDTX 隐蔽信道的安全性相对更高, 但是其隐藏率就更低。IP 隐蔽信道的隐藏率更高, 但是其安全性更低。这也符合安全性和隐藏率的相互制约的特点, 即提高安全性会因导致算法复杂而降低隐藏率。

6 总结与展望

本文提出了两种隐蔽信道, 分别是 BDTX 和 ADDR 隐蔽信道。两种隐蔽信道都有各自的优缺点。BDTX 隐蔽信道利用了比特币网络层的广播机制和应用层的交易机制之间的关系建立密文的索引, 再于新区块初始交易的 coinbase 字段嵌入密文, 提高了该隐蔽信道的安全性, 但也牺牲了一部分隐藏率; 同时, 含有 coinbase 字段的交易每次新区块产生才会被创建一次, 因此秘密信息的传输会受新区块产生速度的限制(一般 10 分钟产生一个新区快)。ADDR 隐蔽信道利用了 ADDR 消息每天广播一次的时分特性, 没有索引机制。因此, 其安全性和隐藏率都居中; 然而, ADDR 消息是在通信双方建立连接的基础上才能互相广播, 一旦传输过程的某一天断开连接, 那么会出现误码情况。

本文提出了利用比特币网络的广播机制、交易机制和通信机制构造隐蔽信道。除此之外, 还有其他两种可以考虑的隐蔽信道: 1)可以通过两节点之间三次握手建立连接时 TCP 包内没有数据的特点, 发送者在报文的序号字段(32 位)嵌入自定义数据以构造隐蔽信道; 2)利用数据传输协议头中一些必须填充的位(如 TCP 数据包协议头的源端口、IP 协议数据包协议头中的源 IP 地址等)来隐藏信息; 3)全球的大部分比特币节点为轻节点, 即只使用类似 SPV 客户端进行交易而不存储历史账本的节点, 但轻节点之间发生的交易也会被全节点验证并写入账本, 并且轻节点可以查看某笔交易是否被验证机器次数。因此可以利用轻节点之间的通信协议构造隐蔽信道。

参考文献

[1] Cai W D, Yu L, Wang R, et al. Blockchain Application Development Techniques[J]. *Journal of Software*, 2017, 28(6): 1474-1487. (蔡维德, 郁莲, 王荣, 等. 基于区块链的应用系统开发方法研究[J]. *软件学报*, 2017, 28(6): 1474-1487.)

[2] Pilkington M. 11 Blockchain technology: principles and applications[J]. Research handbook on digital transformations, 2016: 225.

[3] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system[J]. 2008.

[4] Yuan Y, Wang F Y. Blockchain: The State of the Art and Future Trends[J]. *Acta Automatica Sinica*, 2016, 42(4): 481-494. (袁勇, 王飞跃. 区块链技术发展现状与展望[J]. *自动化学报*, 2016, 42(4): 481-494.)

[5] Petitcolas F A P, Anderson R J, Kuhn M G. Information Hiding-a Survey[J]. *The IEEE*, 1999, 87(7): 1062-1078.

[6] Johnson N F, Jajodia S. Exploring Steganography: Seeing the Unseen[J]. *Computer*, 1998, 31(2): 26-34.

[7] Joinson A N. Self-disclosure in Computer-mediated Communication: The Role of Self-awareness and Visual Anonymity[J]. *European Journal of Social Psychology*, 2001, 31(2): 177-192.

[8] Millen J K. Covert channel capacity[C]. *Security and Privacy*, 1987 IEEE Symposium on. IEEE, 1987: 60-60.

[9] Petitcolas F A P, Anderson R J. Evaluation of copyright marking systems[C]. *Multimedia Computing and Systems*, 1999. IEEE International Conference on. IEEE, 1999, 1: 574-579.

[10] Moskowitz I S, Newman R E, Crepeau D P, et al. Covert Channels and Anonymizing Networks[C].*The ACM workshop on Privacy in the electronic society - WPES '03*, 2003: 79-88.

[11] Bentov I, Lee C, Mizrahi A, et al. Proof of Activity: Extending Bitcoin's Proof of Work via Proof of Stake [Extended Abstract] y[J]. *ACM SIGMETRICS Performance Evaluation Review*, 2014, 42(3): 34-37.

[12] Epishkina, Anna, and Konstantin Kogos. Covert channels parameters evaluation using the information theory statements[C]. *International Conference on IT Convergence and Security. IEEE*, 2015: 1-5.

[13] Zander S, Armitage G, Branch P. A Survey of Covert Channels and Countermeasures in Computer Network Protocols[J]. *IEEE Communications Surveys & Tutorials*, 2007, 9(3): 44-57.

[14] Daemen J, Rijmen V. The Design of Rijndael[M]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.

吕婧淑 于 2017 年在北京大学软件工程专业获得硕士学位。现任中国科学院信息工程研究所信息安全国家重点实验室中级工程师。研究领域为区块链、数据挖掘等。研究兴趣包括: 数字货币、社交挖掘等。Email: lvjingshu78@163.com

操晓春 于 2006 年在美国佛罗里达大学获得计算机科学专业博士学位，现为中国科学院信息工程研究所信息安全实验室研究员。研究领域为计算机视觉、网络空间安全等。Email：caoxiaochun@iie.ac.cn