区块链技术与应用前瞻综述

何 蒲 于 戈 张岩峰 鲍玉斌
(东北大学计算机科学与工程学院 沈阳 110819)

摘 要 区块链技术是一种去中心化、去信任化的分布式数据库技术方案。该数据库由参与系统的所有节点集体维护,具有去中心化、不可篡改、透明、安全等特性。区块链技术归功于比特币应用,它作为比特币的底层技术支持,是比特币系统的核心支撑。区块链技术具有广阔的发展前景,从关键技术、内容、原理、瓶颈、应用和前景几个方面进行介绍,对相关研究问题进行探讨。

关键词 区块链,比特币,Merkle树,POW共识,智能合约
中图法分类号 TP315 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.04.001

Survey on Blockchain Technology and Its Application Prospect

HE Pu YU Ge ZHANG Yan-feng BAO Yu-bin
(School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China)

Abstract Blockchain is a decentralized, untrustworthy distributed database technology. The database is collectively maintained by all the nodes involved in the system, and has the features of decentralization, tamper-resistance, transparentness and security. The emergence of blockchain technology thanks to the Bitcoin applications. Blockchain is the underlying technique support of Bitcoin system and is the fundament of the Bitcoin system. Since Blockchain technology is a promising technology, the key technologies, components, principles, limitations, applications and prospect were introduced, and the related research issues was discussed.

Keywords Blockchain, Bitcoin, Merkle tree, POW consensus, Smart contract

2015年是区块链元年,Fintech上近期出现最频繁的一个词语就是"区块链"。区块链的出现始于2008年末一个自称为中本聪(Satoshi Nakamoto)的人或者团体发表在比特币论坛的一篇论文《Bitcoin: A Peer-to-Peer Electronic Cash System》[1]。该文指出区块链技术是构建比特币系统的基础技术,区块链记录着所有元数据和加密交易信息,从而建立了一个完全通过点对点(P2P)技术实现的电子现金系统,此系统使得在线支付的双方不用通过第三方金融机构而直接进行交易。随后比特币系统大行其道,得到越来越多人的关注和研究,区块链技术作为比特币系统的底层技术也得以重视,由此拉开了研究区块链技术的序幕。

区块链技术是一项新技术,但它不是一种创新技术。它是将许多已有的跨领域的学科整合到一起,从而形成的一种技术,涉及数学、密码学、计算机科学等领域。比特币系统是第一个采用区块链技术作为底层技术构建的系统,它是一个实现了去中心化、去信任化、安全、可靠的电子现金系统。

区块链技术的迅猛发展引起了政府、金融机构的广泛关注。2016年初,中国央行表态:积极推进官方发表的数字货币。随后中国越来越多的金融以及科研机构开始关注数字货币背后的创新技术,即区块链技术[2];几乎同一时间,英国政府发布区块链专题研究报告,即《分布式账本技术:超越区块链》[3],大力发展区块链在政府中的应用;此外,为在区块链技术发展进程上占得先机,平安集团、招商银行、中国外汇交易中心及民生银行更是加入了R3国际联盟[4](一家为全球金融市场设计和提供先进的分布式账本技术的金融创新公司)。一切迹象表明区块链技术正受到政府和金融机构的青睐。

据麦肯锡研究报告[5]指出:区块链技术,是继蒸汽机、电力、信息和互联网科技之后,目前最有潜力触发第五轮颠覆性革命浪潮的核心技术。

目前区块链技术仍然处于萌芽期,特别是在学术方面的研究相对滞后,以谷歌学术和Web of Science为英文数据源的检索结果显示,关于区块链(Blockchain)的相关论文仅有30多篇,本文列举其中的14篇[6-19]。

本文将详细介绍和分析区块链的运行原理和关键技术,并探讨区块链的应用和发展趋势。

1 区块链的基础技术

本节简要介绍与构建区块链相关的基础技术。

到稿日期:2016-11-27 返修日期:2017-02-20 本文受国家自然科学基金重点课题(61433008,61672141)资助。
何 蒲(1992-),男,硕士,主要研究方向为大数据与区块链,E-mail:872896913@qq.com;于 戈(1962-),男,博士,教授,主要研究方向为数据库理论与技术、分布与并行式系统等,E-mail:yuge@mail.neu.edu.cn(通信作者);张岩峰(1982-),男,博士,副教授,主要研究方向为大数据处理与挖掘、云计算等;鲍玉斌(1968-),男,博士,教授,主要研究方向为商务智能、数据挖掘等。

1.1 哈希算法

哈希(也称为散列)算法将任意长度的输入值映射为较短的固定长度的二进制值。例如,SHA256 算法[20]就是将任意长度的输入映射为长度为 256 位的固定长度输出,这个二进制值称为哈希值(也称为散列值)。数据的哈希值可以检验数据的完整性,一般用于快速查找和加密算法。

哈希算法广泛应用于区块链中,区块链通常不保存原始数据,而是保存该数据的哈希值,Merkle 树中的节点信息是两次 SHA256 哈希运算得到的。以太坊账户地址,是用 Keccak-256 哈希运算一个公钥得到的;而比特币地址,则是通过 SHA256 和 RIPEMD160 哈希运算一个公钥而得到的。此外,签名频繁应用于区块链中,它由私钥和需要被签名的数据经哈希运算而成。著名的工作量证明算法、Merkle 树都是哈希算法的应用。

1.2 Merkle 树

1.2.1 Merkle 树

Merkle 树[21-23]是由 Ralph Merkle 发明的一种基于数据哈希构建的树:1)其数据结构是一棵树,一般为二叉树,也可以为多叉树;2)其叶子节点是数据块(如文件或文件集合)的哈希值;3)非叶子节点是其所有子节点的哈希值。

Merkle 树在验证、文件对比中应用较多,特别是在分布式环境下,Merkle 树会大大减小数据的传输量和计算的复杂度。

1.2.2 区块链中的 Merkle 树

区块链中的每个区块都包含了记录于该区块的所有交易,区块链系统采用二叉树型的 Merkle 树对这些交易进行归纳表示,同时生成该交易集合的数字签名,如图 1 所示[1]。Merkle 树支持快速地归纳和校验区块中交易的完整性与存在性。

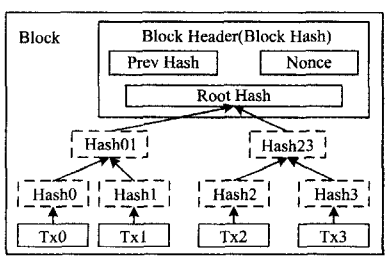

图 1 Merkle 树中的交易哈希

1.3 时间戳服务

区块链技术的发展受到比特币应用需求的推动。比特币作为数字货币,首先需要解决"重复支付(doublespending)"问题,即一笔货币不能被花费两次或者一笔资金不能出现在两个交易中。中心化的信用系统(例如银行)依靠国家机器的强制力来防止伪钞,而区块链系统完全依靠技术来解决"重复支付"问题。系统给每一笔交易盖上正确的时间戳[24-25],以此证明在这个时刻这笔交易确实发生,交易中资金的所属权已经转移,之前资金所有者再次使用这笔资金时就会报错,从而解决重复支付问题。另外,每一个区块也会盖上正确的时间戳,从而形成一个按时间顺序发展的正确链表。

1.4 工作量证明机制

工作证明(Proof Of Work,POW),也称为工作量的证明[1]。比特币系统利用 POW 机制使系统各节点最终达成共识,进而得到最终区块。这里的工作是指找到一个合理的区块哈希值,它需要不断地进行大量的计算,计算时间取决于当前目标的难度和机器的运算速度。当一个节点找到这个值之后,就说明该节点确实经过了大量的计算,这就是工作量证明。由于验证只需对结果值进行一次哈希运算,因此 POW 的验证效率很高。

1.5 权益证明机制

相比 POW 浪费大量的算力,点权益证明(Proof Of Stake,POS)仅仅需要少量的计算就能维持区块链的正常运转。这种机制根据货币持有量和时间来分配相应的利息。但是这种机制存在一点不足,即区块的产生没有消耗大量算力,导致这种机制下的货币价值来源难以确定,因为任何区块链系统都可以实现。

1.6 P2P 网络技术

P2P 网络技术又称为点对点技术,它是一个没有中心服务器、依靠用户群交换信息的互联网体系。P2P 网络由于没有中心化服务器,使得它天生具有耐攻击、高容错的优点;并且各个节点地位平等,服务分散在各个节点上进行,因此部分节点或网络遭到攻击对整个系统几乎没有影响。比特币系统应用 P2P 技术,使各个节点独立地参与系统,每个节点都是一个独立的个体,单独节点宕机或者遭到攻击都不会对系统造成影响。

1.7 非对称加密技术

非对称加密需要密钥对即公钥和私钥成对出现。公钥公开、私钥保密,私钥加密的信息只有对应的公钥才能解开,公钥加密的信息只有对应的私钥才能解密,即公钥加密,私钥解密;私钥签名,公钥验证。在比特币系统中,公钥由私钥通过椭圆曲线加密算法生成;交易信息中必须要有正确的数字签名才能验证交易有效。

2 区块链的概念与结构

2.1 定义

区块链这一概念最早是在中本聪的比特币白皮书[1]中提出的,但它不是以区块链出现的,而是以工作量证明链(proof-of-work chain)的形式存在。下面是中本聪对区块链概念的叙述:时间戳服务就是通过对区块中数据项加上时间戳进行哈希,并把这一哈希值广泛地传播出去,就像是新闻或者在世界性新闻网络(Usenet)上的发帖一样。显然,要得到这个哈希值,就需证明在过去的某个时刻加上时间戳的数据必然存在。每个时间戳包含了先前的时间戳,这样就形成了一条链,并且后面的时间戳都对前一个时间戳进行了增强。

关于区块链的定义,各个机构、权威都给予了不同的定义。

维基百科中文[26]定义:区块链是一种分布式数据库,起源自比特币。区块链是一串使用密码学方法相关联产生的数

据块，每一个数据块中包含了一次比特币网络交易的信息，用于验证其信息的有效性（防伪）和生成下一个区块。中本聪创建的第一个区块，即为“创世区块”。

维基百科英文[27]定义：区块链由包含一系列加盖了时间戳的有效交易的区块组成。每个区块都包含了前一个区块的哈希值，这样就把区块连接在了一起。连接在一起的区块形成区块链，并且每一个随后的区块都是对之前一个区块的增强，因此给它取了一个数据库类型的名字。

巴比特网站[28]：区块链是由一串使用密码学方法产生的数据库组成的，每一个区块都包含了上一个区块的哈希值，从创世区块开始连接到当前区块，形成区块链。每一个区块都确保按照时间顺序在上一个区块之后产生，否则前一个区块的哈希值是未知的。这些特征使得比特币的重复支付（double-spending）变得困难。

以上是不同机构、权威对区块链的定义，虽然有不同，但本质上都一样，即区块链拥有去中心化、去信任化、开放、信息不可更改、匿名、自治的特性。下面总结概括区块链的概念：区块链是指一种电子记录形式的账簿，其中每一个区块是账簿的一页，从第一页“链接”到最新一页。这些区块一旦被确认，几乎不能做修改操作，每个区块包含了当前一段时间内的所有交易信息和区块元数据，如图 2 所示。

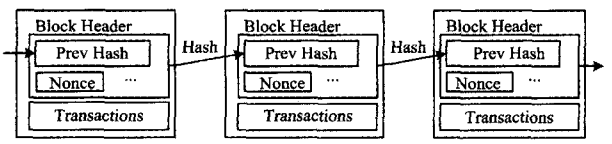

图 2　区块链示意图

2.2　区块链的发展

自 2009 年起，各种各样基于区块链的类比特币的数字货币相继出现，常见的有：bitcoin，litecoin，eth，etc，dogecoin 及最新的堪比比特币的零币 Zcash。区块链除了在货币上的应用之外，还有各种衍生应用，如文件存储系统 Storj、预测市场系统 Augur、电子商务系统 OpenBazaar、智能合约系统等。

区块链的进化过程[29]如下：

1）区块链 1.0——数字货币；

2）区块链 2.0——数字资产与智能合约；

3）区块链 3.0——从 DAO，DAC（区块链自治组织、区块链自治公司）到区块链社会（科学、医疗、教育等）。

2.3　区块链类型

区块链分为 3 类[30]：公有、私有及行业区块链。私有链和行业链是广义的私链。

公有区块链（简称公有链）是指任何个体或者团体都共用一条区块链，只要接入此链都可以在上面发送交易，并且交易能够获得该区块链的有效确认，任何团体或个人都可以参与其共识过程。公有区块链是最先出现的区块链，也是目前应用最为广泛的区块链，这类区块链被认为是“完全去中心化”的。

行业区块链（简称行业链）指共识过程受到某些预选节点控制的区块链。由该行业集体内部首先指定多个预选节点为记账人，每个区块的生成是由所有的预选节点共同决定的（预选节点决定区块链的共识），其他节点只能接入区块链负责交易，但不参与共识过程，任何人都可以通过此区块链对外开放的 API 进行有限查询。这类区块链被认为是“部分去中心化”的。

私有区块链（简称私有链）指仅仅使用区块链这一技术进行记账操作，但它不对外公开。它的对象可以是一个公司也可以是个人，单独拥有此区块链的写入权限，或许会对外开放有高度限制的读取权限。目前金融巨头都在探索自己的私有区块链，既应用到区块链的特性，又能保证安全。

行业链结合了公有链的完全开放和私有链的高度集中，提供了一种混合折中的模式；而私有链由于完全限制的写入权限和高度受限的读取权限，对于保护个人隐私非常合适。

2.4　区块链的数据结构

2.4.1　区块结构

区块链技术中，区块是指一种数据结构，它包含两部分：区块元数据和区块体。其中区块元数据记录的是区块的元数据信息，区块体记录的是从上一区块产生到此区块创建之间所发生的所有交易，如表 1 所列。区块元数据包含区块大小、区块头和交易计数器 3 部分，如表 2 所列。

表 1　区块结构

字段名	大小	描述
区块元数据	变长的	该区块相关元数据信息
交易	变长的	该区块记录的交易

表 2　区块元数据

字段名	大小/Byte	描述
区块大小	4	该字段之后的区块大小
区块头	80	组成区块头的几个字段
交易计数器	1～9（可变整数）	交易的数量

2.4.2　区块头结构

区块头由两组元数据组成，一组与挖矿有关，包括时间戳、难度目标及 Nonce 值；另一组则与区块本身有关，包括链接父区块的字段、版本号及 Merkle 树的根。表 3 列出了区块头的数据结构[30]。

表 3　区块头结构

字段名	大小/Byte	描述
版本	4	用于监测软件/协议更新的版本号
父区块哈希值	32	区块链中父区块哈希值的引用
Merkle 根	32	区块中所有交易构成的 Merkle 树根的哈希值
时间戳	4	该区块创建的近似时间（Unix 纪元秒）
难度目标	4	该区块工作量证明算法的难度目标
Nonce	4	计数器，用于工作量证明算法

将区块中所有交易记录都进行两次哈希运算之后，将结果作为 Merkle 树的叶子节点，然后递归两个相邻节点的哈希值，直到得到最后一个哈希结果，此哈希值就是 Merkle 根。

难度目标 Bits 是一种特殊的浮点编码类型，占 4Byte，首字节是指数，仅用其中的最低 5 位，后 3 个字节是尾数，它能够表示 256 位的数。一个区块头的 SHA256 哈希值必须小于或等于 Bits 难度目标才能被整个网络认可，难度目标 Bits 值越小，产生一个新区块的难度就越大，即目标值越小，得到正

确结果的区间就越小，难度就越大。

Nonce 字段是指随机数，各个区块头的值往往不同，但它却是从 0 开始严格按照线性方式增长的随机数，每次计算都会增长。挖矿就是来寻找一个满足条件的 Nonce 值。

2.4.3 区块标识符

每个区块有两种标识符：区块头哈希值和区块高度。这两个字段都没有被真正记录下来，因为这两个字段可以直接被计算出来。

区块的主标识符是它的区块头哈希值，通过对区块头 6 个字段进行两次 SHA256 哈希计算得到数字签名，产生的 256 位的值被称为区块头哈希值，简称为区块哈希值。例如 000000000019d6689c085ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f是比特币创世区块的区块哈希值。区块哈希值可以唯一、确切地标识一个区块，并且任何节点都可以对区块头进行独立计算从而得到该区块的哈希值。但是为了方便从磁盘检索区块，可能会把哈希值作为元数据存储在一个独立的数据库表中。

区块的第 2 个标识是区块高度，创世区块的高度为 0，通过识别该区块在区块链中的“深度”来确定一个区块。每一个之后产生的区块都比当前最新的区块高出一个位置。2016 年 11 月 1 日的区块高度大约是 436867，这说明已经有 436867 个区块被堆叠在 2009 年 1 月创建的这个区块上。

与区块哈希值不同，区块高度并不是唯一的标识符。虽然一个区块对应着一个确定的高度，反之却不成立，即一个高度并不总能确定一个区块，当区块分叉发生时，两个或者多个区块竞争同一高度。

总之，一个区块的区块哈希值总是能唯一地识别出一个确切的区块。一个区块也有唯一的区块高度。但是，一个特定的区块高度并不是总能唯一地识别出一个特定区块。

2.4.4 创世区块

创世区块即区块链系统中的第一个区块，例如比特币区块链，它的第一个区块创建于 2009 年，称之为创世区块。它是比特币系统中所有区块的共同祖先，这意味着比特币区块链中的任一区块都可以回溯到第一个区块。

比特币创世区块中包含一个隐藏的信息。在其 Coinbase 中包含这样一句话“The Times 03/Jan/2009 Chancellor on brink of second bailout forbanks”，且是泰晤士报当天的头版头条标题，且被留在创世区块里永远不可修改。引用这句话说明区块链的时间戳服务可以用来进行存在性证明，正如上面所说，创世区块记录下了它产生的时间和当时正发生的事件，并且不能被修改。由此可以得到启发，区块链技术将可以用来存储信息，并且是一个存在性证明的最佳技术。

2.4.5 区块的连接

区块链将数据库的架构进行了创新，它把数据信息分成块，每个块通过特定的字段链接到上一块，并按顺序连接起来，整个链囊括了所有完整的数据，并且不可更改。从创世区块开始，所有区块逐个顺接串联起来就形成了区块链，图 3 是一个由 3 个比特币区块连接起来的链。

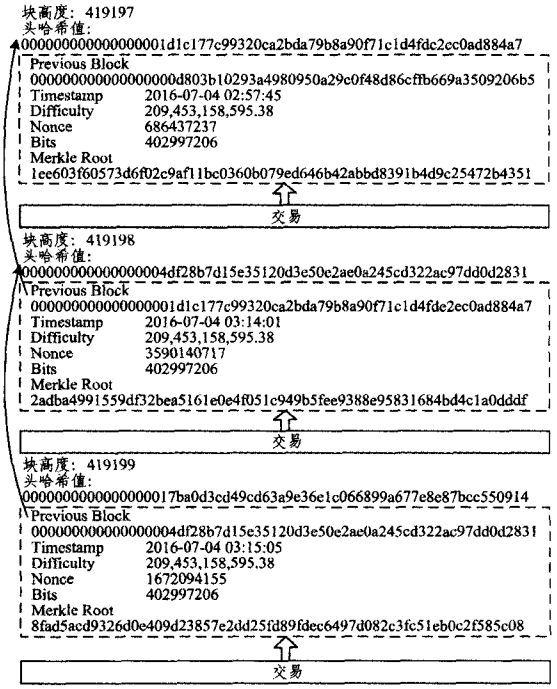

图 3 区块连接

2.5 区块链的特性

(1)可靠开放性：区块链的设计使它能够有效预防故障与攻击，51%攻击除外(如果攻击者拥有网络中 51%的算力，他就可以对区块进行伪造，然后自己又最快地计算出正确的解，造成区块链分叉，达到攻击的目的)。所有参与系统的用户共享一个公共区块链，不会存在因为单点失效而导致系统故障的情况，从而保证了系统的可靠性和数据的可获得性。

(2)信息透明性：网络上的任何节点都可以查看整个账本。由于记录的交易信息不包含任何隐私，因此任何记录在册的信息都可以被查看，保证了数据的透明性。

(3)不可更改性：区块链系统采取的是完全冗余的策略，所有完整节点都有一份完整数据，要想更改某一区块的数据，必须保证所有完整节点数据被修改，这个情况几乎不可能发生，因此降低了欺诈的风险。

(4)不可逆转性：交易不存在撤销操作，交易一旦被验证认可，就不可再逆转。

2.6 区块链技术的瓶颈

区块链技术存在如下瓶颈。

(1)过大的完整账本的存储空间。现在，比特币区块链的容量达到了 80GB，并且还在逐日增加，对于想要运行完整区块链的节点来说这无疑是一个很大的挑战。

(2)信息分发采用全网广播。这是一种洪泛式，要求网络性能好。

(3)交易效率低下。比特币交易效率主要受到两个方面的影响：1)区块产生时间；2)区块大小。现在的比特币交易速度为 7 笔/秒，不适合处理实时交易。

(4)算力浪费。整个系统中，只有成功得到区块合法记账权的那部分算力得到了奖励的回报，其他算力都是做无用功，浪费巨大。

(5)单个节点独自升级，升级成本高。

3　区块链的运行原理

3.1　区块链系统的运行原理

区块链系统运行原理的基本算法如下。

1. 节点发送交易
2. 输出
- 生成 $\{Block_i\}$
- 广播 $Block_i$
3. 刷新节点

所有节点接收和验证交易：
- 验证交易
- If *true*

  加入交易至缓冲区
4. 创建区块

所有节点计算一个哈希值：
- 从缓冲区获取交易
- 构建新区块
- 计算区块头的哈希值 hash{block header}
5. 循环

If hash{block header}<Bits

转至步骤 2(输出)

中本聪在其比特币白皮书中非常详细地介绍了区块链系统的建立过程[1]。

第 1 步　新的交易向全网所有节点广播；

第 2 步　每个节点把收到的交易都写入到一个区块中；

第 3 步　每个节点都在新的区块上进行计算，寻找一个工作量证明解；

第 4 步　某个节点找到工作量证明解时，就把其所在区块向全网进行广播；

第 5 步　其他节点收到广播的这个区块后对其进行验证，只有所有交易都被验证是有效的且未被使用的之后，该区块才能被认可；

第 6 步　每个节点通过将此区块的哈希值作为父哈希值来进行下一个区块的计算，表示节点认可了此区块有效。

一般情况下，一笔交易必须经过至少 6 次确认(在此区块之后每产生一个区块就是 1 次确认)，才能最终在区块链上被承认是合法交易。若达到 6 次确认后要想修改记录，花费代价太大，得不偿失。

3.2　挖矿

所谓“挖矿”，就是竞争区块的合法记账权，在区块链中，得到区块的合法记账权只有唯一一个方式，即“工作量证明”。用计算机反复计算区块头，直到得到满足要求的“解”，即可得到这个区块的记账权，也就是挖到了“矿”。

具体方法：用 SHA256 算法不断地对区块头和一个随机数字 nonce 进行计算，直到计算出一个与预设值 bits 相匹配的解，匹配的解指找到一个哈希值要小于或者等于目标哈希值 bits。第一个找到这个解的矿工会获得此区块的记账权，并将此区块发布到系统中让其他节点进行验证。

挖矿难度由比特币网络自动调整，实现平均每十分钟产生一个区块的平衡。平衡情况下每两周应该产生 2016 个区块(2016=2 * 7 * 24 * 6)，两周进行一次统计，统计出实际区块数量和 2016 的差值，然后按照一定的百分比调整难度目标值来增大或减小难度目标 bits，以达到每 10 分钟一个区块的平衡。

3.3　区块分叉

由于区块链是去中心化的，没有协调机构，加之网络带宽等不同，使得不同副本之间不能总是保持一致，存在着一定的时延。因此区块有可能在不同时间到达不同节点，导致节点存在不同的区块链视图。区块链为了达到共识，让每一个节点总是选择并尝试延长最长的区块链，最长指其计算难度累计最大。区块链节点通过计算所有区块头中的难度总和，来得到建立这条链所要付出的工作量证明的总量。只要所有节点都选择累计难度最长的区块链，比特币网络最终就会达到共识，收敛到一致。

一般情况，分叉发生在两名矿工几乎在同一时间内各自都找到了工作量证明解的时候。它们都向网络发送各自的区块，由于节点连接以及时延等原因，总是一部分节点先收到这个区块，另一部分先收到另一个区块，此时分叉就产生了。所有节点都用它们首先收到的区块的哈希值进行下一个区块的计算，同时保留另外一个区块，因为它有可能成为”正确“的区块。当下一个区块产生时，各个节点选择难度最长的链作为正确链，放弃难度较短的链，此时分叉就会消失。

4　区块链的主要技术平台与应用

4.1　以太坊

Ethereum(以太坊)[31]是一个平台和一种编程语言，使开发人员能够建立和发布下一代分布式应用。其在 2013 年由 Vitalik Buterin 提出，它的目的是为建立去中心化的应用创建一种可替代的协议，给一大类的去中心化的应用程序提供一组不同的平衡机制，这对需要快速开发、安全性要求低、很少使用的应用程序以及在不同应用之间能有效互动很重要。

以太坊是一个编程平台，它提供了各种模板，用户只需要把以太坊提供的各种模板链接到一起就能搭建自己的应用。因此，在以太坊上创建应用的成本大大减少、速度大大提高，这也造就了以太坊成为区块链中最好的项目之一。具体来说，以太坊通过一种图灵完备的脚本语言(Ethereum Virtual Machine Code，EVM 语言)来创建应用，类似于汇编语言，但编写以太坊应用并不需要直接使用 EVM，而是使用 Solidity，Serpent，LLL 类编程语言，再通过编译器转换成 EVM 语言供以太坊平台使用。

开发者可以通过以太坊这一平台创建自己的区块链应用。一般来讲，以太坊上有 3 种应用：1)金融应用，包括电子货币、金融衍生品、对冲交易合约、存储钱包、遗嘱，甚至是一些最终的完善就业合同；2)半金融应用，这类应用涉及金钱，但不是完全看重金钱，也有重要的非金钱的应用，一个典型的例子就是为解决计算问题的自实施奖励；3)非金融应用，例如在线投票和去中心化管理。

4.2　Hyperledger

Hyperledger[32]是 Linux 基金会于 2015 年发起的旨在推进区块链技术发展的开源项目。它领导全球各行业，包括金融、银行、物联网、供应链、制造以及技术领域等，在区块链技术上进行合作，建立一个开放平台，以满足来自各行各业的不同需求，并简化业务流程。

目前 Hyperledger 中主要有 4 个比较成熟的项目：Blockchain Explorer，Fabric，Iroha 和 Sawtooth Lake[33]，本节只以

Blockchain Explorer 为例。

Blockchain Explorer 是由 Christopher Ferris(IBM),Dan Middleton(Intel)和 Pardha Vishnumolakala(DTCC)提出的孵化项目,旨在为 Hyperledger 创建一个用户友好型的 Web 应用程序,用以查看/查询区块、事务和关联数据、网络信息(名称、状态、节点列表)、链码/事务族(查看/调用/部署/查询)以及存储在账本中的任何其他相关信息。

4.3 基于区块链的智能合约

智能合约[34]这一理念最早是在 1994 年出现的,几乎和互联网同时出现。这个术语是由密码学家 Nick Szabo 提出的,定义如下[35]:一个智能合约是计算机协议,它促进、验证或者执行合约的协商或履行,或使合约条款不必要。核心上,这些智能合约的工作原理与任何其他编程语言中的 if-then 语句类似,当满足一个预先编好的条件时,智能合约就被触发执行相应的条款。

区块链技术的出现,使得智能合约再次活跃起来,并被认为是应用在区块链技术上的又一热门技术;并且重新定义了智能合约[36]:智能合约是由事件驱动的、具有状态的、运行在一个可复制和共享的账本上、且能够保管账本上资产的程序。

智能合约与以太坊结合得最好,以太坊中有自己的虚拟机供智能合约使用;Siacoin 文件存储系统采用智能合约来规范用户主机之间的存储协议;而在 Hyperledger 中则称其为 chaincode。

智能合约可以应用在很多方面,例如:1)证券。智能合约状态可以记录证券所有权的所有信息,因此可以用来进行证券的登记和清算。2)智能遗嘱。现今网络带来便利的同时也存在弊端,例如个人在网上存有很多钱,但是亲属并不知道密码,某天这个人去世了,那么他存在网络账户上的钱就不能被提取出来。这时就可以应用智能合约,可以在合约中设定每过一定的时间就必须登录自己的账户确认一次,否则账户中的资金就会按照合约中设定的方式转移出去,这样就能避免前面所说的问题。3)投票、金融衍生品、博彩、预付款等其他方面。

4.4 基于区块链的数据存储

区块链能够保存所有完整信息,这是它的优势,并且任何人、任何节点在任何情况下都可以用加密哈希验证数据块。区块链利用序列化链路,按照时间顺序把所有数据区块串联起来,每个区块包含父区块的哈希值,由此形成了一个去中心化的数据账本。

作为公有链使用,这个账本可以是公开的,因此可以清楚地了解到整个事件的发展顺序;也可以作为私有链或行业链,许可型的,只对外开放一定的接口,严格限制其访问权限。

但无论这个账本是公有化的还是私有化的,它都能够实现连贯一致的数据存储;而且去中心化的特性可以防范外部或内部的恶意破坏。

更重要的是,区块链采取高度冗余的存储策略,可以在大多数节点备份数据;即使在网络连接不稳定或者网络不安全的情况下也不受影响。至于安全性,区块链技术本身十分安全,不存在安全缺陷。

Storj 云存储[37],利用区块链搭建的去中心化云存储系统。Storj 认为,基于区块链的云存储在安全性能和速度方面不输于现有大企业的技术,并且成本更低,因为它不需要购买大型存储设备以及数据中心机架,可利用现成的空闲存储空间。

另外,以太坊合约也考虑到了对去中心化文件存储生态系统的开发,在这种系统里个人用户可以通过出租他们的硬件设备和未用的存储空间获得小批量的金钱收益,从而降低文件存储成本。

5 区块链的应用前景

5.1 当前的区块链应用领域

区块链技术,作为数字货币的背后支撑,已经引起了金融巨头们的高度重视,包括花旗银行、摩根大通、高盛、纽约梅隆银行、汇丰银行、巴克莱银行在内的众多金融巨头,均与区块链公司取得合作,研究区块链技术在金融世界的应用。

麦肯锡研究报告[5]指出了区块链在金融业应用的 5 大场景。

1)数字货币:提高货币发行便利性;

2)跨境支付与结算:实现点到点交易,减少中间费用;

3)票据与供应链金融业务:减少人为介入,降低成本及操作风险;

4)证券发行与交易:实现实时资产转移,加速交易清算速度;

5)客户征信与反欺诈:降低法律合规成本,防止金融犯罪。

除了金融业,其他很多地方也能用到区块链技术。区块链现有的应用场景基本如下:

1)存在性证明。由于区块链技术的不可篡改性,可把区块链技术应用于存在性证明,把过去的某一状态存在区块链上,未来就可以证明在过去确实存在。

2)智能合约。以太坊上的智能合约就是一个最好的例子,把智能合约部署在区块链上,合约内容事先定好,达到合约中的某个条件时合约自动触发,执行合约中的内容可以免去现实生活中合约执行的一些苛刻条件,能在不信任的环境下执行合约。

3)身份验证。智能合约可以存储个人的身份信息,可以保存现有的身份状态,一旦身份信息被篡改就会触发一定的条款,身份所有者就会知晓。

4)预测市场。例如 Augur,它是一个基于区块链技术的去中心化的预测市场的平台,任何人都可以随时随地地访问和使用 Augur,利用这种技术可以消除中心化服务器的风险。

5)电子商务。把比特币无监管模式应用到电商,应用这一模式不仅能免去中间冗杂的环节,还能达到市场和谐。例如 OpenBazaar,直接用比特币进行交易,类似于一个去中心化的淘宝平台。

6)社交通讯。例如去中心化通信平台 Gems,它试图打破现有的社交媒体的模式,不仅社交公司可以赚钱,用户也能从中获利。

7)文件存储。基于区块链的存储技术将直接冲击甚至颠覆传统的云计算架构。Storj,Siacoin,Filecoin 等基于区块链技术的文件存储系统越来越成熟。

5.2 区块链的应用前景

麦肯锡研究报告[5]指出,区块链技术是继蒸汽机、电力、

信息和互联网科技之后，目前最有潜力触发第五轮颠覆性革命浪潮的核心技术；该技术正在悄然改变金融领域，并且可能完全改变现有的交易流程和信息保存方式，从而降低成本，提高效率。在过去的两年中，区块链技术越来越受到各界人士的青睐，俨然已经成为最受关注的话题，各种建立于区块链上的应用应运而生。国际巨头尤其是金融巨头纷纷活跃起来，组建了 R3 CEV 和 Hyperledger 这样的超级联盟来探索未知的区块链的奥秘。

就区块链技术带来的机遇与挑战，各大领先金融机构已经抢先布局，采取的策略不尽相同，大致可分为 3 类[5]。

1)组建区块链大联盟，制订行业标准。如 R3 CEV 集结超过 40 家国际领先银行建立了行业监管及相应的技术标准。

2)携手金融科技公司，发展核心业务区块链应用。如 Capital One 及 Visa 通过战略投资金融科技公司，紧抓区块链技术的突破口。

3)银行内部推进局部领域的应用，快速实施试点。如 UBS、花旗、德意志及巴克莱都已经成立区块链实验室，自行研发或与金融科技公司合作，针对不同的应用场景进行测试。

基于区块链技术的应用有可能颠覆现有的游戏规则，并且可能会重塑行业标准，不仅能降低交易的成本、提高效率，还可能会衍生出许多的商业模式。

结束语　区块链技术的发展被业内人士广泛看好，这个越来越受青睐的趋势会一直持续，这些趋势不一定以区块链的形式直接出现，可能会作为区块链技术的衍生品兴起。区块链应用已经从最初单纯的数字货币过渡到更广泛的金融业，并且渗透到社会中的很多领域，比如身份验证、跨境支付、文件存储、物联网等，其中金融领域是目前最成熟、应用最广的领域。区块链技术作为当下最热门的技术之一，值得我们投入更多的时间与精力进行学习和研究。

参 考 文 献

[1] NAKAMOTO S. Bitcoin: a peer-to-peer electronic cash system [EB/OL]. https://bitcoin.org/bitcoin.pdf.

[2] PILKINGTON M. Blockchain technology: Priciplesand applications [OL]. http://ideas.repec.org/p/hal/journl/halshs-01231205.html.

[3] HANCOCK M, VAIZEY E. Distributed ledger technology: beyond block chain [R]. UK: Government Office for Science, 2016.

[4] R3. About R3 [OL]. http://www.r3cev.com/about.

[5] MCKINSEY COMPANY. Blockchain in insurance —opportunity or threat? [EB/OL]. [2016-07]. http://www.mckinsey.com/industries/financial-services/our-insights/blockchain-in-insurance-opportunity-or-threat.

[6] SWAN M. Blockchain thinking: the brain as a decentralized autonomous corporation[J]. IEEE Technology and Society Magazine, 2015, 34(4): 41-52.

[7] LEE L. New kids on the blockchain: How bitcoin's technology could reinvent the stock market[EB/OL]. (2015-09-18). [2016-09-22]. https://www.finextra.com/resources/feature.aspx?featureid=2030.

[8] GODSIFF P. Bitcoin: bubble or blockchain[J]. Springer International Publishing, 2015, 38: 191-203.

[9] KRAFT D. Difficulty control for blockchain-based consensus systems[J]. Peer-to-Peer Networking and Applications, 2016, 9(2): 397-413.

[10] WILSON D, ATENIESE G. From pretty good to great: enhancing PGP using Bitcoin and the blockchain[C]// Proceedings of the 9th International Conference on Network and System Security. New York: Springer International Publishing, 2015, 9408: 368-375.

[11] ZYSKIND G, NATHAN O, PENTLAND A S. Decentralizing privacy: using blockchain to protect personal data [C]// Proceedings of the 2015 IEEE Security and Privacy Workshops (SPW 2015). San Jose, CA: IEEE, 2015: 180-184.

[12] KYPRIOTAKI K N, ZAMANI E D, GIAGLIS G M. From Bitcoin to decentralized autonomous corporations: extending the application scope of decentralized peer-to-peer networks and blockchains[C]// Proceedings of the 17th International Conference on Enterprise Information Systems (ICEIS2015). Barcelona, 2015: 284-290.

[13] LEWENBERG Y, SOMPOLINSKY Y, ZOHAR A. Inclusve block chain protocols[J]. Financial Cryptography and Data Security, 2015, 8975: 528-547.

[14] BACK A, Corallo M, Dashjr L, et al. Enabling blockchain innovations with pegged sidechains[EB/OL]. [2014-10-22]. http://blockstream.com/sidechains.pdf.

[15] PETERS G W, PANAYI E, CHAPELLE A. Trends in cryptocurrencies and blockchian technologies: A monetary theory and regulation perspective[DB/OL]. [2015-08-19]. https://arxiv.org/pdf/1508.04364v1.pdf.

[16] ATZORI M. Blockchain technology and decentralized governance: is the state still necessary? [OL]. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2709713.

[17] VUKOLIC M. The Quest for Scalable Blockchain Fabric: Proof-of-Work vs. BFT Replication[C]// International Federation for Information Processing. New York: Springer International Publishing, 2016: 112-125.

[18] WRIGHT A, FILIPPI P D. Decetralized Blockchain Technology and the Rise of Lex Cryptographia[OL]. http://www.intgovforum.org/cms/wks2015/uploads/proposal_background_paper/SSRU_ID2580664.pdf.

[19] KOSBA A, MILLER A, SHI E. Hawk: The blockchain model of cryptography and privacy-preserving smart contracts[C]// Symposium on Security and Privacy. New York: IEEE Computer Society, 2016: 839-858.

[20] WOLRICH G M, YAP K S, GuilfordJ D, et al. Instruction set for message scheduling of SHA256 algorithm: US, 8838997B2 [P]. 2012-09-28.

[21] SZYDLO M. Merkle tree traversal in log space and time [J]. Lecture Notes in Computer Science, 2004, 3027: 541-554.

[22] MERKLE R C. Protocols for public key cryptosystems [C]// Proc. 1980 Symposium on Security and Privacy. New York: IEEE Computer Society, 1980: 122-133.

[23] MERKLE R C. A digital signature based on a conventional encryption function[J]. Conference on Advances in Cryptology-crypto, 1987, 293(1): 369-378.

（下转第 15 页）

从表 2 中的数据可知，对原技术的改进是成功的。改进方法略微牺牲了时间开销，在保证克隆检测快速高效进行的前提下，误报率得到了显著的下降。需要说明的是，原技术找到的确认的克隆程序对，改进后的方法也都找到了，不存在原技术找到而改进后的方法漏报的情况。

5　相关工作

基于文本的检测[4-6]是从源代码的文本结构及词法的角度进行克隆检测。这种检测不需要使用重量级的语法分析器，具有良好的时空效率，但是其准确率不是很理想。

基于语法树的检测[7-8]是用静态分析工具把源代码构造成抽象语法树，然后在语法树上寻找相似子树进行克隆检测。基于语法树的检测有很大的时空开销。

基于图的检测[9-10]是分析语法结构、数据流以及调用关系等，构造出整个程序依赖关系图，然后寻找相似的子图。这个过程时空开销很高且效率不高。

基于度量值的检测[3,11]需先设定源代码的比较单元（类、函数等），然后提取比较单元中的一些特征并计算一系列的度量值，最后通过比较度量值进行克隆检测。该类检测方法普遍具有良好的时空效率。

结束语　本文提出了一种面向功能类似程序的高效克隆检测技术。我们根据功能类似克隆检测的特点，对克隆检测技术进行了改进，从而使得克隆检测更加快速，检测结果更加准确。本文实验样本取自真实的课程编程实践，实验结果表明，改进后的克隆检测技术可以快速、高效地检测功能类似程序，结合功能类似程序克隆检测的特点可以有效地控制误报率。

参 考 文 献

[1] ROY C K, CORDY J R, KOSCHKE R. Comparison and evaluation of code clone detection techniques and tools: A qualitative approach[J]. Science of Computer Programming, 2009, 74(7): 470-495.

[2] GU T X, CAO C, XU C, et al. Low-disruptive dynamic updating of Java applications[J]. Information and Software Technology, 2014, 56(9): 1086-1098.

[3] CHEN K, LIU P, ZHANG Y J. Achieving accuracy and scalability simultaneously in detecting application clones on Android markets[C]// Proceedings of the 36th International Conference on Software Engineering. 2014: 175-186.

[4] KAMIYA T, KUSUMOTO S, INOUE K. CCFinder: a multilinguistic token-based code clone detection system for large scale source code[J]. IEEE Transactions on Software Engineering, 2002, 28(7): 654-670.

[5] LI Z M, LU S, MYAGMAR S, ZHOU Y Y. CP-Miner: finding copy-paste and related bugs in large-scale software code[J]. IEEE Transactions on Software Engineering, 2006, 32(3): 176-192.

[6] HITESH S, VAIBHAV S, CRISTINA L. A parallel and efficient approach to large scale clone detection[J]. Journal of Software: Evolution and Process, 2015, 27(6): 402-429.

[7] JIANG L X, MISHERGHI G, SU Z D, et al. DECKARD: Scalable and accurate tree-based detection of code clones[C]// Proceedings of the 29th International Conference on Software Engineering. 2007: 96-105.

[8] RAINER K. Large-scale inter-system clone detection using suffix trees and hashing[J]. Journal of Software: Evolution and Process, 2014, 26(8): 747-769.

[9] LIU C, CHEN C, HAN J W. GPLAG: detection of software plagiarism by program dependence graph analysis[C]// Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2006: 872-881.

[10] QIU J, SU X H, MA P J. Library functions identification in binary code by using graph isomorphism testings[C]// IEEE 22nd International Conference onSoftware Analysis, Evolution and Reengineering. 2015: 261-270.

[11] KAUR M, LAL M. Code clone detection using function based similarities and metrics[J]. International Journal of Emerging Research in Management and Technology, 2015, 4(7): 156-159.

（上接第 7 页）

[24] HABER S, STORNETTA W S. How to time-stamp a digital document[J]. Journal of Cryptology, 1991, 3(2): 99-111.

[25] BAYER D, HABER D, STORNETTA W S. Improving the efficiency and reliability of digital time-stamping[M]. New York: Springer New York. 1993: 329-334.

[26] Wikipedia. 区块链[EB/OL]. [2016-10-21]. https://zh.wikipedia.org/wiki/区块链.

[27] Wikipedia. Blockchain (database) [EB/OL]. [2016-9-15]. https://en.wikipedia.org/wiki/Blockchain_(database).

[28] 巴比特. 区块链是什么[EB/OL]. www.8btc.com/what-is-blockchain.

[29] SWANM. Blockchain: Blueprint for New Economy[M]. USA: O'Reilly Media Inc, 2015.

[30] ANTONOPOULOS A M. Mastering Bitcoin[M]. USA: O'Reilly Media, 2014.

[31] BUTERIN V. Ethereum: A next generation smart contractand decentralized application platform [EB/OL]. https://github.com/ethereum/wiki/wiki/White-Paper, 2013.

[32] HYPERLEDGER. About the hyperledger project[EB/OL]. https://www.hyperledger.org/about.

[33] HYPERLEDGER. Projects[EB/OL]. https://www.hyperledger.org/community/projects.

[34] WIKIPEDIA. Smart contract[EB/OL]. https://en.wikipedia.org/wiki/Smart_contract.

[35] CASSANO J. What are smart contracts?cryptocurrency's killer app[N/OL]. Fastcompany, 2014-09-17 [2016-10-23]. https://www.fastcompany.com/3035723/app-economy/smart-con-tracts-could-be-cryptocurrencys-killer-app.

[36] BROWN A R. A simple model for smart contracts[EB/OL]. https://gendal.me/2015/02/10/a-simple-model-for-smart-contracts.

[37] WILKINSON S, BOSHEVSKI T, BRANDOFF J, et al. Storj: A peer-to-peer cloud storage network(V0.2)[EB/OL]. [2016-11-15]. https://storj.io/storj.pdf.