HZ BOOKS
华章IT

以太坊创始人、首席科学家Vitalik Buterin倾力推荐！
工业界与学术界区块链专家联合撰写，权威性和实用性毋庸置疑

深入剖析以太坊架构、核心部件、智能合约编写与开发案例等关键技术，并涵盖以太坊数据分析、性能优化、隐私与数据安全等前沿实践与进展

区块链技术丛书

以太坊
技术详解与实战

闫莺 郑凯 郭众鑫 ◎编著

机械工业出版社
China Machine Press

作者简介

闫莺（博士），微软亚洲研究院主管研究员，区块链领域负责人，微软 Coco 区块链平台中国负责人。中国软件协会区块链创业学院及区块链专委会专家、中国电子学会区块链专家委员。专注于区块链技术、大数据分析、数据库以及云计算的研究。在区块链领域获得多项国际专利，并在数据库和云计算领域国际顶级会议和期刊发表论文 30 余篇。参与翻译《区块链项目开发指南》。

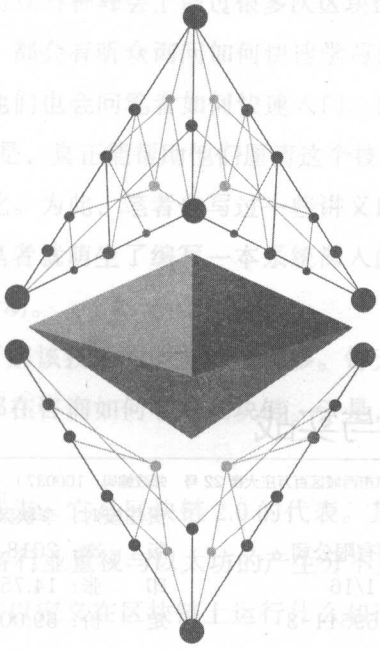
郑凯（博士），电子科技大学教授，博士生导师，中组部“千人计划”专家，澳大利亚昆士兰大学计算机科学博士。主要研究领域为区块链数据管理，以及时空数据挖掘、不确定数据库、内存数据库、图数据库等。在数据库、数据挖掘等领域的重要会议和期刊发表论文 100 余篇，被累计引用 1500 余次。2013 年获澳大利亚优秀青年基金，2015 年获数据库顶级会议 ICDE 最佳论文奖。担任数据库领域知名国际会议的程序主席和联合执行主席，国际 SCI 期刊客座编委，以及数十个国际顶级会议的程序委员。

郭众鑫 微软亚洲研究院研发工程师，微软 Coco 区块链平台核心开发者。专注于区块链技术、大数据分析、分布式系统等方面的研究和开发。

区块链
技术丛书

以太坊
技术详解与实战

闫莺 郑凯 郭众鑫 ◎编著



机械工业出版社
China Machine Press

图书在版编目（CIP）数据

以太坊技术详解与实战 / 闫莺，郑凯，郭众鑫编著 . —北京：机械工业出版社，2018.6
（区块链技术丛书）

ISBN 978-7-111-59511-3

I. 以… II. ①闫… ②郑… ③郭… III. 分布式数据库 - 数据库系统 IV. TP311.133.1

中国版本图书馆 CIP 数据核字（2018）第 055354 号

以太坊技术详解与实战

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）
责任编辑：高婧雅 责任校对：李秋荣
印 刷：北京市荣盛彩色印刷有限公司 版 次：2018 年 4 月第 1 版第 1 次印刷
开 本：186mm × 240mm 1/16 印 张：14.75
书 号：ISBN 978-7-111-59511-3 定 价：59.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88379426 88361066 投稿热线：（010）88379604
购书热线：（010）68326294 88379649 68995259 读者信箱：hzit@hzbook.com

Preface 前　言

为什么要写这本书

随着区块链技术近两年迅速“走红”，身边越来越多的朋友想了解区块链技术及其应用场景。2017 年一整年，笔者也在各种峰会上做过很多次区块链的演讲，约 80% 会议的听众是入门级别的，每次演讲完，都会有听众询问如何快速学习区块链技术。每当有新的学生加入我们的实习生团队时，他们也会问笔者如何快速入门。通常笔者会回答他们“从以太坊白皮书、黄皮书看起”。但是，真正能帮助他们厘清这个技术的背景、原理、关键知识点和实战要领的资料尚未系统化。为此，笔者也写过一些讲义以帮助大家理解，但是仍难以做到全面和系统。从那时起笔者就萌生了编写一本系统深入的区块链书籍的想法。但是由于工作繁忙，一直没有付诸行动。

随着区块链的升温，想了解该技术的朋友持续增多。每天笔者的微信、信箱都会有来自同事、朋友的信息，他们都在咨询如何学习区块链。于是，笔者觉得是时候写一本探索技术、指导开发的书了。

为什么选择以太坊呢？首先，它是区块链 2.0 的代表。其实“区块链”这个词脱离比特币（区块链 1.0）而单独被各行业重视与以太坊的产生分不开。以太坊是第一个通用的区块链平台，换句话说，用户可以定义在区块链上运行什么和记录什么。以太坊的公有链已经运行两年多，整个社区不断修补出现的问题，积极寻求优化的途径。尽管它不是完美的，但它是目前经得起时间和应用验证的最稳定的系统。其他很多区块链项目都或多或少受到以太坊的启发。因此，系统学习以太坊可认为学习其他系统打下非常好的基础。其次，以太坊社区的建设比较完善和活跃，各个版本的代码质量较高，开发工具相对完善，应用也有一定规模，这使得大家易于上手学习。再次，笔者团队的工作也是以以太坊为主。比如

笔者团队在开发微软的 Coco 区块链平台时，就以集成和优化以太坊为 coco 第一版本的目标。通过项目开发，笔者更加熟悉以太坊源码，这样也自然使得本书更加具体化。笔者曾在 2017 年翻译了《区块链项目开发指南》[⊖]一书，该书介绍了以太坊开发相关知识，特点是覆盖面比较广，而本书会在深度上下工夫。因此，读者可以“搭配”着学习。

本书将展现给读者一个系统、全面的以太坊知识体系，以通俗易懂的语言结合直观的图示介绍每一个原理和工作流程，相信读者通过本书的学习可轻松快速地入门以太坊开发。

本书特色

首先，为了增强知识结构的凝聚性，本书没有泛泛而谈整个区块链，而是更加专注于以太坊公有链本身的技术。通过阅读本书，读者可以全面、深入地了解以太坊的顶层设计、实现原理、重要模块的技术细节，以及智能合约的编写与部署等重要概念和技术。这是本书与目前介绍区块链相关技术的书籍最大的不同。

其次，从技术深度上讲，本书所涉及的内容具有很好的层次性，既涵盖初学者所需的基本概念，也包括以太坊 DApp 开发工程师感兴趣的编程指南和代码解析，此外对以太坊在性能和安全性方面所尝试的改进技术进行了前瞻性介绍，以供资深工程师和研究人员参考、探讨。

再次，本书不仅介绍以太坊本身技术细节，还加入笔者在开发中的经验和技巧。比如在部署以太坊的时候可以手工操作，也可以用脚本在“云”上操作，其中脚本也分享给大家借鉴。

最后，本书的文字力求简洁、朴实且准确，可读性较强。

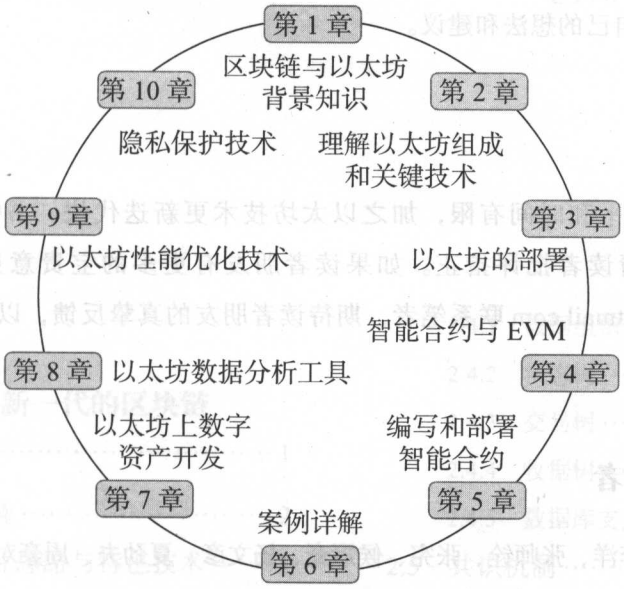
读者对象

- ❑ 区块链开发初学者
- ❑ 区块链应用架构师
- ❑ 开发应用架构师
- ❑ 区块链产品经理
- ❑ 其他对区块链技术感兴趣的人员

⊖ 该书已由机械工业出版社出版，ISBN：978-7-111-58400-1。

如何阅读本书

本书分为 10 章，下图比较清晰地展示了各章的主题。



第 1 章从区块链背景知识讲起，包括区块链基本原理及应用，使得初学者和开发者都能对区块链有整体性了解。然后引出为什么需要以太坊以及以太坊的基本知识，这为后面章节的阅读提供整体形象的铺垫。

第 2 章介绍以太坊的组成、关键概念和技术。本章比较重要，其后介绍的内容都将以本章的概念为基础。因此，必须仔细阅读。

第 3 章介绍不同区块链网络类型，以及如何部署不同类型的区块链。建议读者在阅读本章时也能同时跟着书中介绍的部署步骤进行操作，以更好地理解以太坊网络。根据实际经验，本章将介绍一些部署的窍门及脚本样例，相信一定能为大家的学习提供帮助。

第 4 章介绍智能合约和以太坊虚拟机的原理。了解该原理，可为接下来第 5 章学习开发智能合约打好基础。

第 5 章和第 6 章详细地介绍具体编写智能合约的方法以及案例详解。建议读者在阅读这两章时能同步操作，一起编写、编译、部署合约，达到最佳的学习效率和理解深度。

第 7 章介绍以太坊上数字资产定义的原理和方法，其中包括近期火爆的 CryptoKitties（养猫游戏）的 ERC 721 代币合约标准的介绍。到这里为止，读者可以开始编写自己的以太坊应用了。

第 8 章将进一步对查看、分析以太坊公有链数据的工具和方法进行介绍。

第 9 章和第 10 章探讨区块链和以太坊的前沿技术。这两章会对以太坊在性能优化和隐私保护方面的技术进行介绍和讨论。这些技术尚处于比较初级的阶段，读者可以一边阅读一边思考，并提出自己的想法和建议。

勘误和支持

由于笔者的水平和时间有限，加之以太坊技术更新迭代快，书中难免存在一些不准确的叙述，恳请读者批评指正。如果读者朋友有更多的宝贵意见，欢迎通过邮箱 EthereumDetail@hotmail.com 联系笔者，期待读者朋友的真挚反馈，以在技术之路上互勉共进。

本书的其他贡献者

感谢我们团队李洋、张师铨、张宪、候冠豪、杨文彦、夏劲夫、周豪对本书内容的贡献！

致谢

笔者要特别感谢微软亚洲研究院的周礼栋和洪小文院长对笔者团队区块链项目的指导和支持。感谢陈洋博士过去一年多在区块链方面的共同探讨。感谢杨懋、伍鸣、熊一远、黎强、周沛源、Thomas Moscibroda、张益肇、殷秋丰、田江森、程鼎、黎江、梁戈碧、宋青见、桂柯里、石朝阳、张蓉等同事、领导和朋友的支持与鼓励。感谢导师周傲英教授和周晓方教授指引方向。感谢同行的共同努力，感谢家人的支持！还要感谢 V 神 Vitalik 的支持与肯定。

最后还要感谢机械工业出版社华章公司的高婧雅编辑对本书的全程支持和指导。她在本书的内容组织和阅读体验方面给我们提出十分宝贵的意见和设计方案，正是她的兢兢业业、一丝不苟的负责态度，保证了本书内容的质量和可读性。

闫莺

Contents 目　录

第 1 章 Chapter 1

以太坊：新一代的区块链平台

区块链是近期大家关注和讨论的热点，几乎每个行业都在积极地探索区块链技术，渴望从中挖掘出新的运营模式和商机。区块链的魅力究竟在哪里呢？若说今天的互联网是信息通过 TCP/IP 进行点对点的传递，是信息互联网，那么价值（比如电子货币、电子资产、设备访问权限等）怎样才能脱离第三方直接进行点对点的转移呢？这就需要一个价值互联网（见图 1-1）。与信息的复制和粘贴不同，价值的转移涉及所有权的变更，如我把我的资产转给你，意味着这份资产需要从我的账户里面扣除，而在你的账户里面添加。因此，在价值转移过程中，我们需要一份账本来记录资产的变更。该账本需要安全、稳定可靠，以及具有一定的覆盖面和可用性（如全球资产需要全球覆盖，在任何地方都可以查询到当前的资产状态）。如何构建这样一个账本呢？区块链提供了这样一种可能的技术手段。

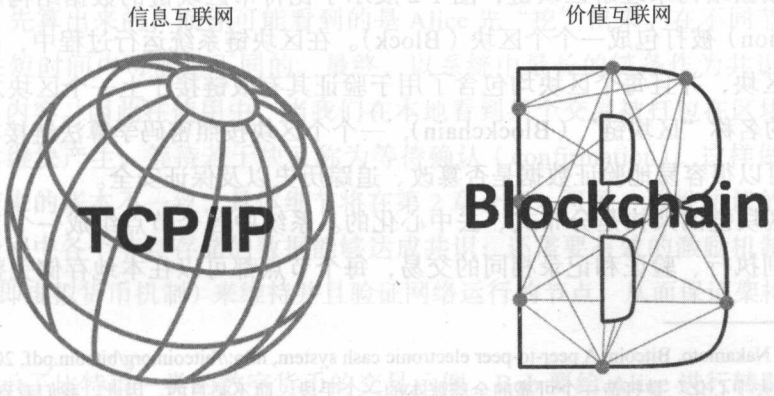


图 1-1　信息互联网与价值互联网

1.1 理解区块链

区块链通常被定义为去中心的分布式记账系统，该系统中的节点无需互相信任，通过统一的共识机制共同维护一份账本。比特币可以说是第一个区块链应用。在金融危机爆发的 2008 年，一位名叫中本聪（Satoshi Nakamoto）的神秘人物在《比特币：一个点对点电子现金系统》[⊖]中首次提出了“比特币”这一概念。比特币的底层记账系统就是现在我们说的区块链技术，而中本聪身份之谜也为比特币和区块链技术带来了更加神秘的色彩。在前几年，大家会关注比特币而不会单独谈论区块链这个技术。直到 2015 年，区块链这一概念才被单独提出来为更多人所了解，且向着更广泛的应用场景发展。发生在这个时间点的主要原因之一是以太坊的出现和日益成熟。

区块链是一种**分布式、去中心化的计算与存储架构**。在详细了解区块链每个技术组成部分之前，先来理解为什么需要这种架构。

区块链要解决的是如何用一种可信的方式记录数据，使得用户可以信任区块链系统记录的数据，而无须假设记账节点的可信性。怎么实现呢？“无须信任”技术上的解决办法就是假设互相不信任。因此，每个节点都存有一份完整的数据记录，每条新的交易都要被重新验证。当一个节点重新加入网络并需要同步数据的时候，也是从其他节点同步交易历史，然后重新计算验证——这就决定了其第一个特点，即分布式存储（不能完全信任他人的存储）。也正是为了高效可靠的验证需要，才有了区块链现在的数据结构：区块链由成块的交易通过密码学算法连接在一起，使得整个账本公开透明、可追踪、不可篡改（数据被篡改时很容易被验证发现）[⊖]。这么多记账节点为什么愿意按照一致性协议记账呢？依靠的就是巧妙的记账激励机制——诚实的记账节点会得到相应的奖赏，且诚实的记录比恶意篡改记录的收益更大——这就是一致性协议设计中的要点。下面就对区块链的数据结构、分布式存储和一致性协议进行详细介绍。

首先从数据结构来理解区块链，图 1-2 展示了比特币区块链的数据结构。系统中的交易（Transaction）被打包成一个个区块（Block）。在区块链系统运行过程中，区块链每次只能添加一个区块，并且每个区块均包含了用于验证其有效链接于上一个区块之后的数学凭证。正如它的名称“区块链”（Blockchain），一个个区块按照密码学算法链接在一起。这样的组织设计可以很容易地验证数据是否篡改、追踪历史以及保证安全。

其次，区块链的架构是分布式、去中心化的。系统中各个节点组成一个 P2P 网络，每个节点均分别执行、验证和记录相同的交易，每个节点都可以在本地存储完整的区块链数

⊖ Satoshi Nakamoto, Bitcoin: A peer-to-peer electronic cash system, http://bitcoin.org/bitcoin.pdf, 2009。

⊖ 这里“去中心化”是打造一个可靠的全球账本的一个手段，而不是目的。因此，我们看到具体实施的区块链解决方案中也有多中心的设计。

据。没有一个中心机构能够干预交易的执行顺序和结果。因此，该架构具有很强的鲁棒性。这里要说明一点，我们看到的公有链的平台是去中心的，因为其设计假设以没有任何信任作为前提，即都不可信。在实际的应用中，如果有一些可信的元素，是完全可以利用的。因此我们也看到很多系统设计是多中心或者弱中心的模式。“去中心”在这里不是目的，而是一种达到可信的手段。

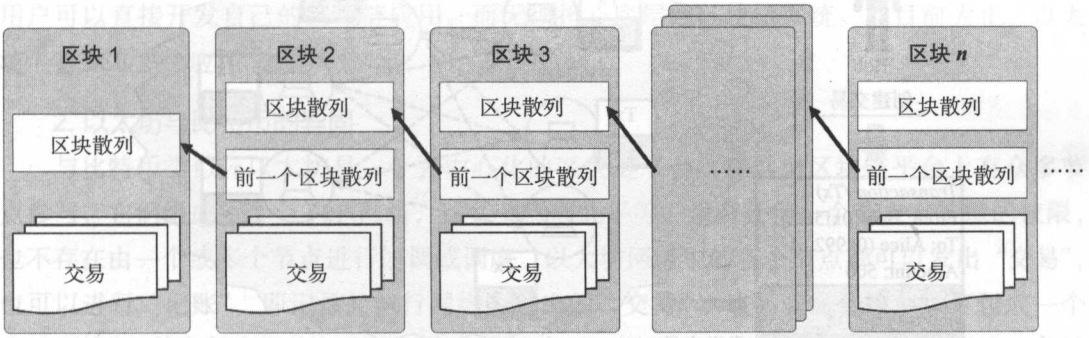


图 1-2　比特币区块链的数据结构

最后，为了保证各节点状态的一致性，还需要共识机制，即一致性协议（如 PoW、PoS、PoA 等）。以 PoW 为例，为了使得各个节点记录的结果是一致的，在每一时刻系统要选择一个记账节点来计算下一个区块。其他节点对该记账节点的区块结果进行验证，通过后则接受这个区块。为了激励大家高效正确地记账，系统对记账节点有相应的奖赏，这样一来大家会贡献计算和存储资源来争夺记账权。由于可以互相验证，也保证了记账的可靠性。接下来，我们要解决的问题就是：如何公平地选取这个记账节点，以及如何设计激励机制。PoW 中采用的是“猜散列值”这个公平的、依靠消耗算力的方式，也被称作“挖矿”。谁先算出给定要求的散列值，谁就以大概率争夺到这个记账权。为什么说是概率呢，因为在分布式网络中，由于延迟，消息传递到其他各个节点的时间是不一样的。比如，我看到的是 Bob 先算出来的，而你可能看到的是 Alice 先“挖”出来。在不同节点上对下一区块的认可在短时间内可能是不同的。最终，以系统中最长的链条作为共识结果，即大家认可的账本内容。因此在使用中，当我们在本地看到某个交易被打包在区块链后，还需要等待若干后继块产生，等待若干块又称为等待确认（confirmation）。这样做的目的是防止由于延迟带来的账本不一致，具体细节将在第 2 章介绍。可见一致性协议的设计既要安全，以保证全网中各个节点存储的数据能够达成共识；还需要有效的激励机制，给予一定的经济奖励（即虚拟货币机制）来维持并且验证网络运行的节点，从而保证架构的稳定健康运行。

图 1-3 展示了比特币一类的数字货币的交易示例。Bob 要给 Alice 进行转账，他需要创建一条交易，声明转账的付款人、收款人以及转账金额。之后 Bob 在这条交易上添加自己

的数字签名，并将交易发布到区块链网络上。这条交易被记账节点验证后打包广播，并通过共识（一致性）协议达成全网一致。Alice 在确认看到交易被记录，且该交易后面还有若干区块陆续被记录后（通常 6 ～ 12 个块）就可以认为自己已经收到了 Bob 的转账。

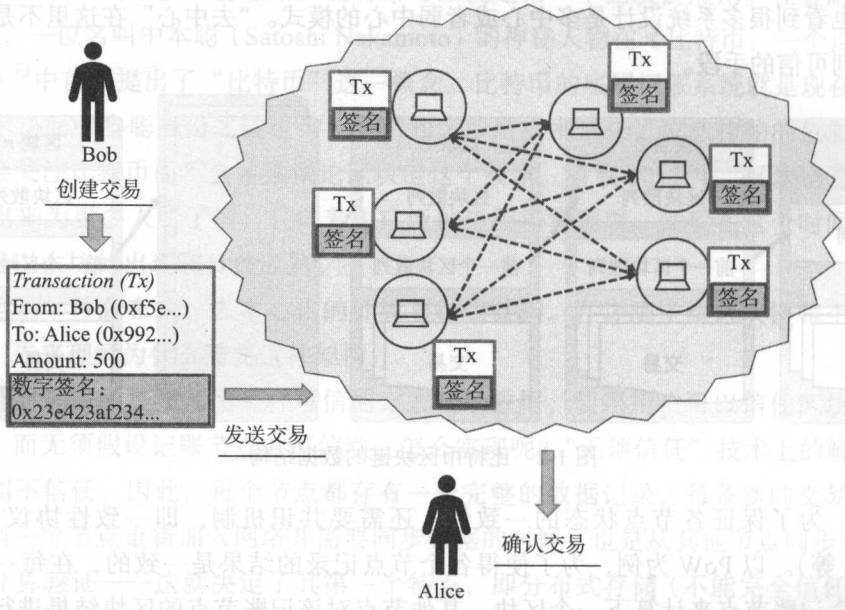


图 1-3 区块链的网络结构

对于比特币的原理和使用我们有了一定的了解，接下来的问题是区块链是不是只能支持数字货币这一种应用呢？在其他业务场景中，当逻辑复杂、资产多样化的时候，又该如何利用区块链呢？

1.2 以太坊设计思路与特色技术

随着比特币开始受到开发者等技术人员更多的关注，一些利用比特币网络实现不同于比特币逻辑的代币交易，或者除代币之外其他数字资产交易的新项目开始出现。由于比特币不太灵活，这些项目大多基于比特币系统做了一些改变，添加了一些新的特征和功能，然后独立地运行在不同的节点上。或者说，每一个新项目都要重复、独立地建立一个类似比特币的系统。能不能设计一个更通用的系统呢？通过应用层的编写，让不同的数字资产运行在统一的平台之上？以太坊的发明者 Vitalik Buterin 就在思考这个问题。

1. 以太坊的诞生

在 2013 年下半年，Vitalik Buterin（当时他才 19 岁）提出了“以太坊”的概念——一种能够被重编程用以实现任意复杂计算功能的单一区块链，这种新的区块链包含了之前众

多区块链项目的大多数特征。2014年，以太坊基金会成立，Vitalik Buterin、Gavin Wood和Jeffrey Wilcke创建了以太坊项目，作为下一代区块链系统。今天，以太坊[⊖]作为全球最为知名的公有区块链项目之一，同时拥有全球最大的区块链开源社区。简单地说，以太坊是一个有智能合约（Smart Contract）功能的公共区块链平台。用智能手机打个比方，如果说以太坊是智能手机的操作系统，那么智能合约就是上面搭载的应用（App）。有了以太坊，用户可以直接开发自己的区块链应用，而无须担心底层的区块链系统。到目前为止，以太坊上有880多个应用[⊖]。

2. 以太坊与比特币的异同

与比特币类似，以太坊是一个去中心化的区块链平台。在这个区块链平台上有众多节点参与，它们组成了一个P2P网络，这些节点彼此平等，没有任何一个节点有特殊的权限，也不存在由一个或多个节点进行协调或调度。以太坊网络中的各个节点都可以发出“交易”，也可以进行“记账”，即记录并执行网络上发出的“交易”。这些交易会被节点打包成一个个“区块”，其中每个区块包含上一个区块的索引，因此这些区块依次相连接，形成一条区块链。如上文所述，这些节点之间通过共识机制以达成数据一致性，从而形成一个整体。早期版本的以太坊像比特币一样使用“工作量证明”（Proof of Work，PoW）这种共识机制来保证一致性。

以太坊与比特币不同的地方有很多，从性能表现以及特性上来看，主要有以下几点区别。

- 以太坊有更快的“出块”速度以及更先进的奖励机制。目前，比特币的出块时间平均为10min，而以太坊的出块间隔为12s，这意味着以太坊具有更大的系统吞吐量和更小的交易确认间隔。
- 以太坊支持智能合约，用户可以自己定义数字资产和流通的逻辑，通过以太坊虚拟机几乎可以执行任何计算，而比特币只能支持比特币的转账。这一点意味着以太坊可以作为更通用的区块链平台，支持各种去中心化应用（DApp）。

另外，以太坊的社区更加活跃。显然，不像比特币一样满足于虚拟货币，以太坊积极地探索新技术，不断地对系统升级更新。而且其相关技术生态更加完善，在Ethereum官方的GitHub上有147个项目，其中不仅有各种不同语言版本的客户端，还有智能合约编译器、集成开发环境，以及未来将要采用的“股权证明”（Proof of Stake，PoS）协议和各种技术文档。

3. 以太坊的特色技术

如上文所述，以太坊是一个**可编程的区块链**。形象一点地理解，在以太坊区块链上发

⊖ Ethereum: https://www.ethereum.org/。

⊖ https://www.stateofthedapps.com/。

送的交易不仅仅可以是转账金额，还可以是调用一段代码，而该代码可以由用户自定义。因此可以想象，在以太坊区块链上处理的交易逻辑不再是单一的转账，而可能是任意的函数调用；记录在区块链账本里的不仅仅是账户余额，还有函数调用后变量的新状态。因为代码可以任意定义，所以应用就都可以在区块链上运行了。

支持用户在以太坊网络中创建并调用一些复杂的逻辑，这是以太坊区别于比特币区块链技术最大的挑战。以太坊作为一个可编程区块链的核心是以太坊虚拟机（EVM）。每个以太坊节点都运行着 EVM。EVM 是一个图灵完备的虚拟机，这意味着通过它可以实现各种复杂的逻辑。用户在以太坊网络中发布或者调用的“智能合约”就是运行在 EVM 上的。智能合约和 EVM 将在第 4 章介绍。

所谓智能合约其实就是一段 EVM 可执行的代码，熟悉面向对象编程的读者可以将一个智能合约实例理解成一个对象。简单来说，用户编写一个智能合约类似于编写一个类，其可以在这个类里定义各种变量以及函数。当用户将这个智能合约发布到以太坊网络中时，相当于给这个类生成一个对象，合约发布之后用户会得到一个合约地址，相当于合约对象的指针。当网络中的用户调用这个智能合约时，可以直接给这个合约地址发送“交易”，并声明本次调用的函数名称和参数，使得智能合约执行对应的逻辑。无论发布还是调用智能合约，智能合约的信息都被附在“交易”中，以交易的形式发布到网络中。因此以太坊网络中的节点接收到这些交易后，其中的 EVM 会执行对应的合约代码，最后各个节点通过 PoW 或 PoS 等达成共识，合约的内容和状态也就实现了全网一致。

这里给出一个简单的例子。下面这段代码就是一个智能合约 SimpleStorage，里面只有一个变量 storedData，以及 set 和 get 方法，有编程基础的读者可以很轻松地理解。

```
contract SimpleStorage {
    string storedData;
    function set(string s) {
        storedData = s;
    }
    function get() constant returns (string) {
        return storedData;
    }
}
```

图 1-4 展示了在以太坊网络中创建智能合约的过程。当 Bob 将一个包含智能合约信息（如上例代码）的交易发送到以太坊网络中后，节点的 EVM 执行这个交易并生成对应的合约实例，图中的“0x6f8ae93..”代表了这个合约的地址。节点间通过共识机制达成一致后，这个合约就正式生效了，之后用户就可以调用 SimpleStorage 合约了。

图 1-5 展示了在以太坊上调用智能合约的过程。Bob 同样以交易的形式在“To”字段填上 SimpleStorage 合约的地址，在“Data”字段填上调用的方法（set）和参数（“Hello”），