区块链中的自适应广播路由分配算法

秦 毅
(重庆电子工程职业学院人工智能与大数据学院 重庆 401331)

摘 要 针对区块链中单播消息导致的数据包重复传输问题,提出一种区块链同步服务的自适应广播路由分配算法,该算法用于区块链的认证、授权和计费(authentication, authorization, and accounting, AAA)服务。将网络拓扑的问题特征和区块链的数据库验证概念进行数学模型描述。根据覆盖网络拓扑中的分散处理,提出一种应用层广播方法,实现自适应路径设计和链路分配,用于分析密码信息在消息中传播到共享网络池中的所有主机的过程。构造广播树作为覆盖网络拓扑,以最小延迟改善信息验证能力,包括通过适当的传输路径选择来限制传输和计算延迟。实验结果表明,自适应动态 AAA 架构和路径选择使区块链运营商能够有效地做出决策并实现更安全的服务。

关键词 区块链 同步服务 自适应广播 路由分配 认证、授权和计费

中图分类号 TP309 **文献标识码** A DOI:10.3969/j.issn.1000-386x.2019.08.046

ADAPTIVE BROADCAST ROUTING ALLOCATION ALGORITHM IN BLOCKCHAIN

Qin Yi
(*School of Big Data and Artificial Intelligence, Chongqing College of Electronic Engineering, Chongqing 401331, China*)

Abstract Aiming at the problem of duplicate transmission of packets caused by unicast messages in the blockchain, this paper proposed an adaptive broadcast routing assignment algorithm for the blockchain synchronization service. The algorithm was used for authentication, authorization, and accounting(AAA) services of blockchain. We described the problem characteristics of network topology and the concept of database validation of blockchain by mathematical model. According to the decentralized processing in overlay network topology, an application layer broadcasting method was proposed to implement adaptive path design and link assignment. And it was used to analyze the process of cryptographic information propagating to all hosts in the shared network pool. A broadcast tree was constructed as an overlay network topology to improve information validation capability with minimum latency, including limiting transmission and computing latency through appropriate transmission path selection. Experimental results show that the adaptive dynamic AAA architecture and path selection enable blockchain operators to make effective decisions and achieve more secure services.

Keywords Blockchain Synchronization service Adaptive broadcast Route assignment Authentication, authorization, and accounting

0 引 言

区块链是一种技术,允许双方之间记录交易的每个主机可验证和永久分散的分类账[1-2],较为流行的应用是加密货币(如比特币)。当 Internet 上的其他主机完成验证时,最新块将成为附加区块链的一部分,每个块包括使用哈希函数和加密密钥加密的多个事务[3],块头中的元数据集被记录到关系中并与其他块链接。Internet 上的任何主机都可以联合验证事务作为验证贡献者,并检查块是否正确[4]。这种主机的协作称为比特币应用程序中的挖掘。

当使用区块链时,该区块具有一系列过程(如交易和支付)的数字签名,其可以追溯到个人以进行识

收稿日期:2018-11-23。重庆市高等教育教学改革研究重点项目(162071);重庆市教育委员会科学技术研究项目(KJ1729408)。
秦毅,副教授,主研领域:计算机应用。

别、验证和确认。节点保持的块被分散以便共享给每个主机,这种分散的系统可以保护区块链免受篡改、删除和修改[5]。为了保持分类账副本的一致性,主机需要就事务达成一致。广播机制提供由其中一个节点创建的用于暂时提交事务的新块,并且以规则的间隔同步。该块被广播并分发给所有主机以进行验证和确认[6],完成确认的最快节点收到奖励或加密货币金额,传输和计算延迟是矿工的关键因素[7]。

广播路由机制的服务质量(QoS)路由的目标是找到具有足够可用资源的可行路径,以满足网络中节点的 QoS 要求并实现有效的资源使用[8]。延迟、带宽、延迟抖动、吞吐量和丢包率是将块广播到所有矿工主机的路由策略的 QoS 测量[9]。研究确定优化问题的可行路径,并找到了成本最低的可行解决方案,文献[10]对各种 QoS 路由算法进行了分析,分为源路由算法、分布式路由算法和分层路由算法。文献[11]提出了最小生成树(MST),涉及无向生成树的分配,当在网络中使用 MST 时,考虑 QoS 问题是必要的,受路由树的 QoS 约束的最短路径问题,MST 问题是 NP 难问题。

目前区块链已经被用于复制因特网上所有矿工设备的交易数据。网络资源管理已通过软件定义网络(software-defined networking, SDN)和网络功能虚拟化(network function virtualization, NFV)适应资源容器化[12]。SDN 是一种可编程机制[13],可以动态灵活地控制路由路径和链路管理,实现端到端通信。NFV 是网络功能可以转换为基于软件的应用程序的概念,可以用于区块链广播方法。启用 SDN 后,可以通过网络状态向应用程序通知详细路由和流量负载信息,这有助于有效地为应用级广播选择覆盖网络转发节点。

通过区块链技术引入了三个主要的数据处理任务:1) 中间矿工的交易处理协调;2) 交易处理监控的会话;3) 交易数据到区块链的分布式写入[15]。在所有节点收到完整消息后,矿工可以收到奖励,这称为工作证明,表明每个参与者将验证结果发送到区块链。生成的块在构建分布式分类账的过程中连接到现有的区块链,首先生成此块的主机有责任将块广播到必须将该块存储在网络中的其他主机。但是,在大多数情况下,单播用于发送消息,因为 Internet 路由器上没有启用广播,或者默认情况下使用正在运行的网络协议 TCP / IP 进行切换。单播消息导致许多相同的数据包被重复传输,这可能会导致网络拥塞。

针对以上问题,本文提出了一种区块链自适应广播路由新方法,该方法从数据库验证的概念出发,将区块链作为分布式 AAA 模块在 Internet 上的矿工主机进行模拟。针对每个矿工主机的加密信息,提出一种应用层广播方法,用于分析加密信息在消息中如何传播到共享网络池中的所有主机,构造广播树作为覆盖网络拓扑,以最小的延迟提高信息验证能力,包括通过适当的传输路径选择来限制传输和计算延迟。本文将区块链的比特币应用的概念扩展到一种新的分布式虚拟机 AAA 系统结构。

1 区块链广播数学模型

根据 AAA 体系结构,将块或事务引入虚拟机(VM)中初始化的 VNFS,网络拓扑由 VM 初始化,块的信息,例如用户标识、身份验证、授权和委托,使用公钥密码进行数字签名。使用广播机制在整个网络拓扑中传播密码学,每个收件人可以使用私钥验证事务,公开密钥用于认证和识别。广播消息可能引起传播延迟,总延迟被假定为沿着路径的传输和处理时间的组合。

广播是一种自动通信技术,用于所有矿工主机始终如一地验证与区块链应用程序相关的重要数据,如事务和分类账。然而,广播环境的保密性和有效性应予以考虑,维护分散的功能验证和性能是 NP 问题。该问题可以用广播树模型抽象地构造和建模,该模型具有资源分配和路由分配问题。

本文提出了一种广播方案,用于分析区块链所采用的加密信息如何将消息传播到 Internet 上的每个节点。网络拓扑可以由广播树的成本来构造,从覆盖网络的角度以最小延迟(包括传输和处理延迟)来改进信息验证能力。

x_p和 y_{bl}表示决策变量,对于块 b 的目的地 d,如果路径 $p \in P_{bd}$被选中,则 x_p为 1,否则为 0,其中 P_{bd}表示块 b 的目的地 d 的候选路径集合 $d \in D_b$,D_b表示块 b 的目的地集合。如果块 b 采用链路 l,则 y_{bl}为 1,否则为 0。目标函数为:

$$\min \sum_{b \in B} \sum_{l \in L} (\alpha_l + \beta_l + \gamma_{bl}) y_{bl} \tag{1}$$

式中:α_l表示链路上的传输成本,β_l表示链路上的处理成本,γ_{bl}表示对于块 b 链路 $l \in L$ 上的惩罚成本,L 表示覆盖网络中的链路集$\{1,2,\cdots,l\}$,y_{bl}受制于:每个块 b 被选择为采用链路 l(等于 1)或不采用链路 l(等于 0),对于$\forall b \in B, l \in L$,则有:

$$\begin{cases} y_{bl} = 1 & b \text{ 选择键路 } l \\ y_{bl} = 0 & b \text{ 不选择键路 } l \end{cases} \tag{2}$$

B 表示所有需要向矿工主机广播的块$\{1,2,\cdots,b\}$,对于每个块 b,采用的链路数应大于到最远节点的跳跃时间和目标节点的数量,对于$\forall b \in B$,则有:

$$\sum_{l \in L} y_{bl} \geqslant \max\{h_b, | D_b |\} \tag{3}$$

式中:h_b表示用于发送块 b 到最远目标节点的最小跳数,D_b表示块 b 的目的地集合。对于每个块 b,每个目标节点的接入链路数应等于或小于 1,对于 $\forall b \in B$,则有:

$$\sum_{l \in I_{d_b}} y_{bl} \leqslant 1 \tag{4}$$

式中:I_{d_b}表示目标节点的传入连接,对于每个块 b,根节点的接入链路数应为 0,对于 $\forall b \in B$,则有:

$$\sum_{l \in I_{rp}} y_{bl} = 0 \tag{5}$$

式中:I_{rp}表示根节点的传入连接,对于广播到目的地 d 的每个块 b,只采用一条路径,对于 $\forall b \in B, d \in D_b$,有:

$$\sum_{p \in P_{bd}} x_p = 1 \tag{6}$$

式中:P_{bd}表示块 b 的目的地 d 的候选路径集合 $d \in D_b$,D_b表示块 b 的目的地集合。对于广播到目的地 d 的每个块 b,可以采用许多路径(如果选择采用路径 p,则决策变量等于 1)或不采用路径 p,则决策变量等于 0,如下所示:

$$\begin{cases} x_p = 1 & b \text{ 选择路径 } p \\ x_p = 0 & b \text{ 不选择路径 } p \end{cases} \tag{7}$$

式(7)的约束条件为:$\forall b \in B, p \in P_{bd}, d \in D_b$,对于广播到目的地 d 的每个块 b,如果采用路径 p,则路径 p 上的所有链路应设置为 1,对于 $\forall b \in B, l \in L, d \in D_b$,则有:

$$\sum_{p \in P_{bd}} x_p \delta_{pl} \leqslant y_{bl} \tag{8}$$

式中:δ_{pl}表示指示函数,如果链路 l 在路径 p 上,则为 1,否则为 0,对于每个块 b 广播到所有目的节点,如果已采用链路 l,则采用的总次数 l 应小于目的节点数,对于 $\forall b \in B, l \in L$,则有:

$$\sum_{d \in D_b} \sum_{p \in P_{bd}} x_p \delta_{pl} \leqslant |D_b| y_{bl} \tag{9}$$

为了找到块 b 的惩罚成本,将链接的惩罚(如果在块 $b-1$ 中被采用)和块 $b-1$ 的惩罚成本线性组合,对于 $\forall b \in B, l \in L$,有:

$$\gamma_{bl} = wC_l^E \cdot y_{(b-1)l} + (1-w)\gamma_{(b-1)l} \tag{10}$$

式中:w 表示以前处罚的权重,C_l^E表示链路 $l \in L$ 上的惩罚成本。

对于广播策略的中继转发选择通过一个实例进行说明,以包含 3 个节点的网络进行举例,如图 1 所示。对于网络中广播中继转发一般是以减少能耗为目的的,但同时会参考节点位置分布和所处环境进行选择。本次实例忽略参数变化,重点研究最小传输能耗与位置分布的关系,图 1 中网络中的 3 个节点都具有全向天线,D_{12}、D_{13}、D_{23}分别代表节点 1 和 2 之间的距离、节点 1 和 3 之间的距离、节点 2 和 3 之间的距离。

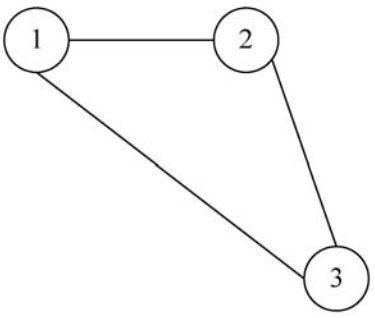


图 1　包含 3 个节点的网络

对于图 1 中网络,节点 1 有两个广播再转发策略:(1) 直接向节点 2 和 3 进行消息广播,此时能耗为 E;(2) 通过中继节点 2 向节点 3 广播数据包,此时能耗为 $E' = E_{12} + E_{23}$。根据文献[16]无线广播算法中概率模型与能耗模型的计算表达式,当 $E > E'$时,第 2 种广播策略能耗更小,当 $E < E'$时,第 1 种广播策略能耗更小,当 $E = E'$时,第 1 种广播策略和第 2 种广播策略能耗相同。

通过文献[16]中对不同位置的节点的能耗分析,可以得出:在网络拓扑结构中,广播策略中使得能耗最小的路径选择,与节点位置相关,因为位置不同,使得有时通过中继转发消耗的能量小,有时直接传输信息消耗的能量小。本文对于广播策略中路径设计和优化过程是将广播模型作为一个最小生成树问题求解本文目标函数的最小成本,具体过程在第 2 节中给出。

2　路由路径设计和优化链路分配

为了找到本文目标函数的最小成本,将本文模型看作为一个最小生成树(Minimum Spanning Tree, MST)问题。如果图形具有 n 个顶点,则每个树具有 $n-1$个边;最小化总边缘权重。图 2 显示了连接的无向加权图,应用 MST 算法后,可以找到该图的 MST。

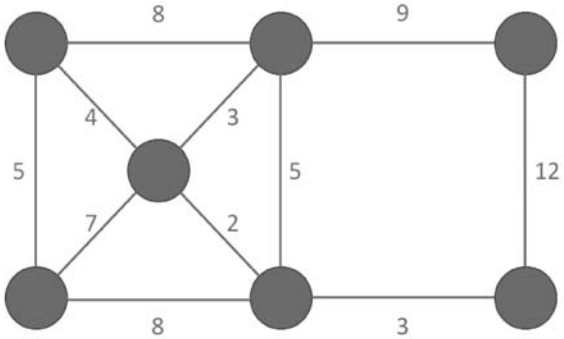


图 2　无向加权图网络拓扑

图 3 显示了连接图中所有节点的最小总边缘权重生成树。

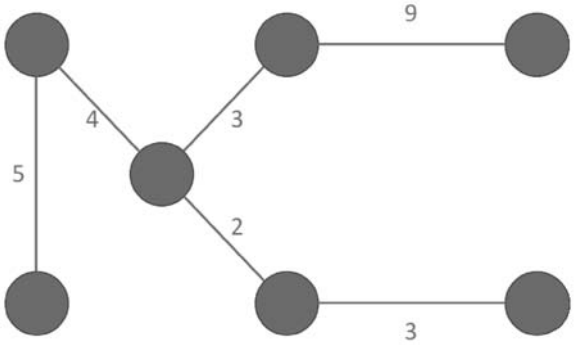
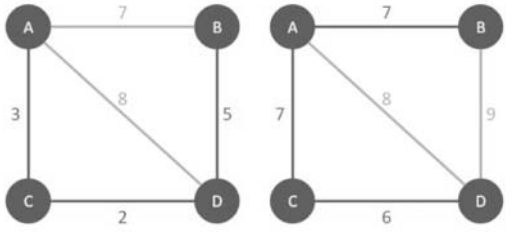


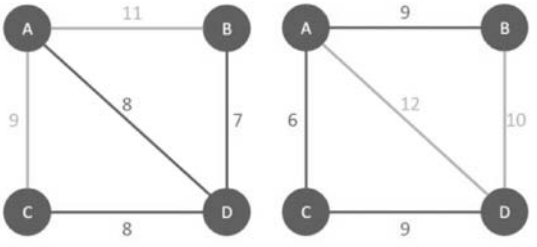
图 3 最小生成树

在本文数学模型中,链路上的最大聚合延迟是沿树的源节点到目的节点之间的传输延迟和进程延迟的组合,但是,应考虑广播环境的机密性和有效性。通过在使用链路后添加惩罚,解决方案将避免广播选择相同的链路,为每个广播块提供随机拓扑。

图 4 说明了每个区块链中的块进行随机拓扑的过程,图 4(a)给出了第一广播块的 MST,链路上的总成本将是传输成本和处理成本的组合,没有重复惩罚。但是在图 4(b)中,对图 4(a)中选择的链路添加了重复惩罚,因此应用了不同的拓扑。通过使用式(10)来计算图 4 中的惩罚,其中,权重 $w=0.5$,$C_l^E=8$。如图 4(c)所示,连接 A 与 B、A 与 C 的链路不再被选中,因为增加惩罚成本后的总成本使其成为不是更好的选择。重复惩罚导致图 4(a) -(d)中的不同拓扑,从而产生随机路由路径和高安全广播环境。



(a) 第一个广播块的 MST (b) 添加重复惩罚



(c) 随机路径 (d) 随机路径

图 4 每个广播块的 MST

3 实 验

根据第 2 节中所提公式,将每个块广播视为单独的 MST 问题,每条链路上的权重是传输成本、处理成本和重用链路的惩罚成本的组合。本文实验使用 Prim 算法来解决随后的块广播中的每个单独的 MST 问题,并获得作为实验结果的总目标值。本文实验硬件环境是笔记本电脑,具有 Windows 7 系统,4 GB 运行内存,500 GB 磁盘内存,实验软件环境是 MATLAB 2013a。实验中广播树的构造如图 5 所示。

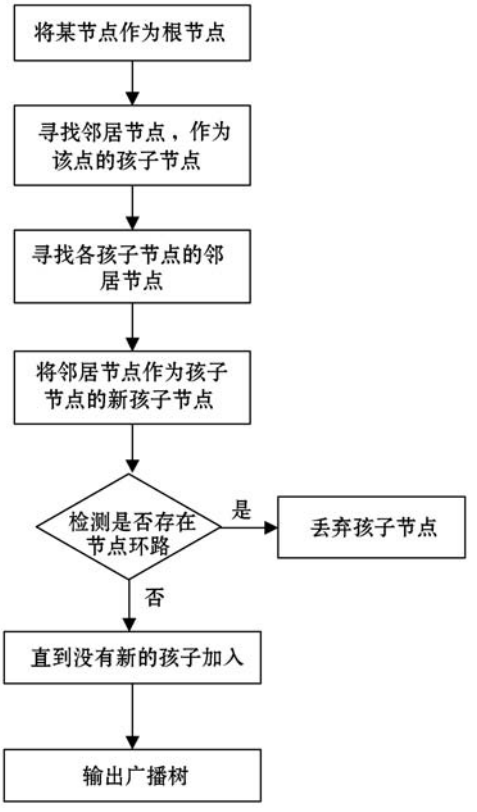


图 5 构造广播树流程

当网络链路处于空闲状态时,广播树路由使用最短路径进行路由,当网络链路处于繁忙状态时,使用多径路由对网络压力进行分担,从前 k 条最少跳数的备用路径中随机分配路径进行广播。

在本文中,进行了几个实验来验证提出的模型,并比较不同情况下的结果。为了比较实验之间的差异,将固定值分配给模型中的某些给定参数,表 1 显示了实验中使用的给定参数的属性。通过将随机数分别乘以传输权重和计算权重,随机分配每个链路的传输成本和每个节点的计算成本。如果在以下实验中未用作独立变量,表 1 中指定的值是每个参数的默认值。

表 1 实验参数

给定参数	值
所有需要广播的块	10
网络中的节点集	10
核心节点和边缘节点的比率	0.3 ~ 0.7
重复惩罚的权重	0.5
传输权重	300
计算权重	100
重复权重	100
每个边缘的传输成本	随机
计算每个节点的成本	随机

首先对节点数量和模型的目标值之间的关系进行实验。在现实使用中,节点的数量可以被认为是系统的规模。拥有更大规模系统的运营商将拥有比在小型系统中运营的运营商更多的节点。为了进行实验,在每个测试中保持每个参数的值相同。这种情况首先将节点数调整为 10, 20,…, 90,结果如图 6 所示。

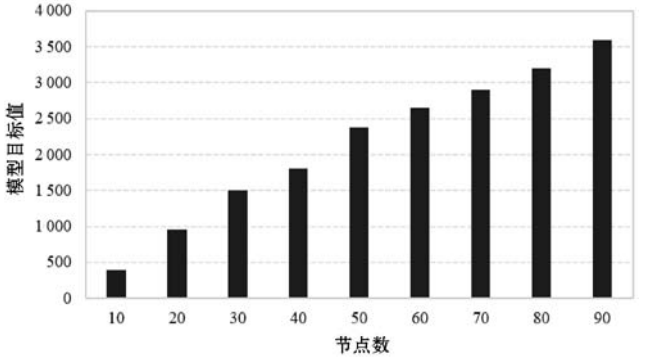


图 6　不同节点数的目标值

图 6 显示了当节点数增加时,目标值随之增加的趋势,这是因为更大规模的系统意味着操作员必须花费更多来操作整个系统。

然后对不同核心比率时成本变化趋势进行实验验证。在实验中,所有节点分为两种类型:核心节点和边缘节点,在现实使用中,系统也分为核心和边缘计算节点。核心计算节点具有更强大的计算能力但节点之间的传输成本更高,边缘计算节点功能较弱但传输成本较低。为这两种节点类型分配不同的核心比率,从 0.2 到 0.6(边缘比率分别为 0.8 到 0.4),结果如图 7 所示,提供了有关不同分配比率的模型中提到的不同类型成本变化的信息。

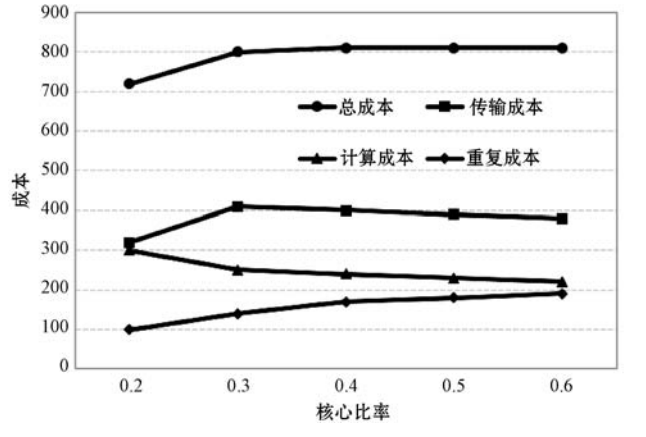


图 7　不同核心比率的成本分布评估

从图中可以看出,随着核心比率的增加,总成本也会增加,但逐渐达到最大值,传输成本以类似的方式呈现,但最后略有减少。这是因为当核心比率首先增加时,一些节点加入核心节点,从而增加传输成本,但核心比率不断增加,两种节点的传输成本平衡并达到最大值。

当核心比率增加时,计算成本降低并达到最小值,设置为核心节点的边缘节点越多,系统的计算能力就越高,这导致整个系统的计算成本下降。当核心比率增加时,重复成本增加并逐渐达到最大值。这表明当两种类型的节点的比率变得更加平衡时,其重复成本增加,较低的核心比率将导致较低的重复成本。

最后对各种重复惩罚权重时成本的变化进行实验。在现实使用中,重复惩罚权重可以被看作是操作者反复使用链接的紧迫性。图 8 给出了从 0 到 1 分配权重的实验结果。

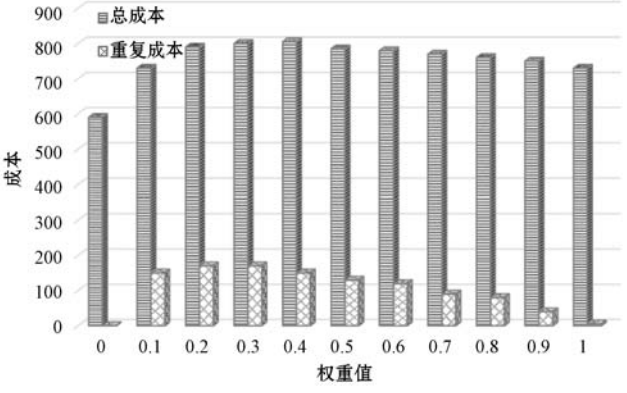


图 8　不同权重值的成本评估

当权重较小时,重复成本增加,但当权重大于 0.4 时,重复成本逐渐降低。由于惩罚权重的增加,重复成本首先增加,然而在重复成本增加之后,系统将尝试寻找另一路径以实现较低的总成本。当选择不同的路径时,重复成本降低。

4　结　语

为了将 SDN 和 NFV 技术适应于分布在边缘和核心云环境中的云基础设施,提出了区块链的资源管理。本文提出一种区块链同步服务的自适应广播算法,广播转发节点在共享网络架构中采用具有各种拓扑的广播和路由策略,以通过较少重用传输链路来最小化处理和传输延迟,分配给 AAA 服务、块或事务。通过计算实验说明本文方法的可行性和有效性。

参　考　文　献

[1] Zyskind G, Nathan O. Decentralizing privacy: Using blockchain to protect personal data[C]//Security and Privacy Workshops (SPW), 2015 IEEE. San Jose, CA, USA: IEEE, 2015: 180 - 184.

[2] 何蒲, 于戈, 张岩峰, 等. 区块链技术与应用前瞻综述[J]. 计算机科学, 2017, 44(4):1 - 7.

[3] Vukolić M. The quest for scalable blockchain fabric: Proof-of-work vs. BFT replication[C]//International Workshop on

Open Problems in Network Security. Springer, Cham, 2015: 112 – 125.
[4] 周致成, 李立新, 李作辉. 基于区块链技术的高效跨域认证方案[J]. 计算机应用, 2018, 38(2):316 – 320.
[5] Iansiti M, Lakhani K R. The truth about blockchain[J]. Harvard Business Review, 2017, 95(1): 118 – 127.
[6] Aitzhan N Z, Svetinovic D. Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams[J]. IEEE Transactions on Dependable and Secure Computing, 2018, 15(5): 840 – 852.
[7] Kiayias A, Russell A, David B, et al. Ouroboros: A provably secure proof-of-stake blockchain protocol[C]//Annual International Cryptology Conference. Springer, Cham, 2017: 357 – 388.
[8] 孔祥彬, 沈苏彬, 李莉. 一种基于 SDN 网络的 QoS 路由选择方案[J]. 计算机技术与发展, 2018(2):102 – 106.
[9] Guck J W, Van Bemten A, Reisslein M, et al. Unicast QoS routing algorithms for SDN: A comprehensive survey and performance evaluation[J]. IEEE Communications Surveys & Tutorials, 2018, 20(1): 388 – 415.
[10] Tomovic S, Radusinovic I, Prasad N. Performance comparison of QoS routing algorithms applicable to large-scale SDN networks[C]// EUROCON 2015-International Conference on Computer as a Tool (EUROCON), IEEE. Salamanca, Spain: IEEE, 2015: 1 – 6.
[11] Al-rubaye M, Salameh H B, Jararweh Y. Minimum spanning tree-based multicast routing protocol for dynamic spectrum access networks: A multi-layer probabilistic approach [C]// 2016 7th International Conference on Computer Science and Information Technology (CSIT). Amman, Jordan: IEEE, 2016: 1 – 6.
[12] Sharma S, Miller R, Francini A. A cloud-native approach to 5G network slicing[J]. IEEE Communications Magazine, 2017, 55(8): 120 – 127.
[13] Nunes B A A, Mendonca M, Nguyen X N, et al. A survey of software-defined networking: Past, present, and future of programmable networks[J]. IEEE Communications Surveys & Tutorials, 2014, 16(3): 1617 – 1634.
[14] Han B, Gopalakrishnan V, Ji L, et al. Network function virtualization: Challenges and opportunities for innovations [J]. IEEE Communications Magazine, 2015, 53(2): 90 – 97.
[15] Watanabe H, Fujimura S, Nakadaira A, et al. Blockchain contract: A complete consensus using blockchain [C]// Consumer Electronics (GCCE), 2015 IEEE 4th Global Conference on. Osaka, Japan: IEEE, 2015: 577 – 578.
[16] 程红举, 黄行波. 不可靠通信环境下无线传感器网络最小能耗广播算法[J]. 软件学报, 2014(5):1101 – 1112.

(上接第 274 页)

[26] Chen L, Yu T, Chirkova R. Wave cluster with differential privacy[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York: ACM, 2015.
[27] Xu W H, Chang Y, Qin Z. A Framework for Classifying Uncertain and Evolving Data Streams[J]. Information Technology Journal, 2011, 10(10):1926 – 1933.
[28] Mehmood R, Bie R, Dawood H, et al. Fuzzy clustering by fast search and find of density peaks[C]//International Conference on Identification, Information, and Knowledge in the Internet of Things. IEEE, 2016:258 – 261.
[29] 马春来, 单洪, 马涛. 一种基于簇中心点自动选择策略的密度峰值聚类算法[J]. 计算机科学, 2016(7):255 – 258.
[30] 王建忠. 高维数据几何结构及降维[M]. 北京: 高等教育出版社, 2012.
[31] Tenenbaum J B. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. Science, 2000, 290(5500): 2319 – 2323.
[32] Roweis S T. Nonlinear Dimensionality Reduction by Locally Linear Embedding[J]. Science, 2000, 290(5500):2323 – 2326.
[33] Saul L K, Roweis S T. Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifold[J]. Journal of Machine Learning Research, 2003, 4(2):119 – 155.
[34] Belkin M, Niyogi P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation[J]. Neural Computation, 2003, 15(6):1373 – 1396.
[35] Vinh N X, Epps J, Bailey J. Information theoretic measures for clusterings comparison: is a correction for chance necessary? [C]//Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009:1073 – 1080.
[36] Zahn C T. Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters[J]. IEEE Transactions on Computers, 1971, C-20(1):68 – 86.
[37] Gionis A, Mannila H, Tsaparas P. Clustering aggregation [J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1):4.
[38] Kim P, Kim S. Detecting overlapping and hierarchical communities in complex network using interaction-based edge clustering[J]. Physica A Statistical Mechanics & Its Applications, 2015, 417(C):46 – 56.
[39] Lancichinetti A, Fortunato S, Kertesz J. Detecting the overlapping and hierarchical community structure in complex networks[J]. New Journal of Physics, 2009, 11(3):033015.
[40] Karypis G, Han E H, Kumar V. Chameleon: hierarchical clustering using dynamic modeling[J]. Computer, 1999, 32(8):68 – 75.
[41] Zhang T, Ramakrishnan R, Livny M. BIRCH: A new data clustering algorithm and its applications[J]. Data Mining and Knowledge Discovery, 1997, 1(2):141 – 182.