



专题：先进数据网络

去中心化互联网基础设施

刘冰洋¹, 杨飞¹, 任首首¹, 魏鑫鹏¹, 杨雪¹, 王闯¹, 延志伟²

(1. 华为技术有限公司, 北京 100094; 2. 中国互联网络信息中心, 北京 100190)

摘要: 域间路由系统、域名系统和公钥基础设施等是互联网重要的基础设施, 它们是互联网的网络连通性、服务可用性和通信可信性的基础。然而, 这些基础设施或其背后的可信模型是中心化的, 存在着中心节点权限过大、单点失效等脆弱性, 降低了互联网的安全性、可靠性和平等性。为了构建一个更加安全、可靠、平等和开放的互联网, 尝试提出了去中心化的互联网基础设施 (decentralized internet infrastructure, DII), 并讨论其体系结构和各层的设计, 分析系统的可行性。DII 架构包含 3 个层次: 底层利用分布式账本技术构建基础的去中心化能力; 中间层构建 IP 地址和域名等互联网名字空间的去中心化可信管理机制, 并支持安全可信的域间路由和域名映射系统; 顶层为开放的应用层, 支持和促进创新、可信的去中心化互联网应用。

关键词: 互联网基础设施; 安全可信; 去中心化

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019200

Decentralized internet infrastructure

LIU Bingyang¹, YANG Fei¹, REN Shoushou¹, WEI Xinpeng¹, YANG Xue¹, WANG Chuang¹, YAN Zhiwei²

1. Huawei Technologies Co., Ltd., Beijing 100094, China

2. China Internet Network Information Center, Beijing 100190, China

Abstract: Inter-domain routing system, domain name system (DNS) and public key infrastructure (PKI) are the fundamental infrastructures of the internet. These infrastructures help to provide network connectivity, service availability and communication trusts between different internet entities. However, these infrastructures are all designed based on a centralized system architecture or a trusted model, which suffers from many vulnerabilities. An architecture called decentralized internet infrastructure (DII) based on decentralization techniques was proposed. The DII architecture consists of three layers. The underlying layer was the distributed ledger layer, providing decentralized trusted foundation for DII. The intermediate layer, called name space management layer, fulfills the management of internet core resources (such as IP addresses, AS numbers and domain names) and provides trustworthy mapping information between different resources. The top layer was an open application layer that could support trustworthy decentralized internet applications.

Key words: internet infrastructure, security and trusted, decentralization

收稿日期: 2019-07-10; 修回日期: 2019-08-01

基金项目: 国家重点研发计划基金资助项目 (No.2018YFB1800100)

Foundation Item: The National Key Research and Development Program of China (No.2018YFB1800100)

1 引言

今天,互联网已经发展成为了一张覆盖全球、连接数十亿设备的巨大网络,承载着对生产生活十分重要的通信任务。为了使得距离遥远、互不相识的两个节点能够可信、可靠地通信,互联网提供了相应的基础设施。从网络层到应用层,共有3个最重要的基础设施,如图1所示。

- 边界网关协议 (border gateway protocol, BGP) 通过将 IP 地址前缀关联到自治域和域间拓扑,计算域间路由,实现了全球互联网的基础连通性。
- 域名系统 (domain name system, DNS) 通过将域名映射为 IP 地址,将应用层服务名与网络层地址关联起来,使得服务能够在网络中被访问。
- 公钥基础设施 (public key infrastructure, PKI) 通过将企业身份信息和其他信息关联到公钥,将网络通信实体与真实世界的真实身份关联起来,使得通信变得可信。

当前,这些基础设施或其背后所依赖的安全可信系统都采取了如图2(a)所示的中心化的设计。该设计的基本原理是,单一可信根节点为整个系统的信任锚点,它为中间层的可信节点“背书”,中间层节点进而再为叶子节点“背书”,可信关系由根层层传递到叶子。以BGP为例,BGP本身虽然是分布式的协议,但其可信基础是资源公钥基础设施(RPKI),用以验证BGP地址前缀宣告的合法性、保障BGP的安全。RPKI采用了

中心化、层次化的机构,RPKI的系统设计的具体介绍如图2(b)所示。RPKI以IANA为根节点,全球五大地区的RIR为第二层节点,向下可能还包含国家NIR节点,各个运营商为叶子节点。RPKI自顶向下的信任关系靠逐级颁发的资源证书(resource certificate, RC)来承载,RC用于声明某个节点对某个IP地址前缀的所有权。最终,拥有资源证书的叶子节点可以通过发布路由源认证(route origin authorization, ROA)授权某个自治系统(AS)号,允许该AS在发出的BGP消息中进行该节点持有的IP地址前缀的宣告。相应地,BGP路由器在收到BGP消息后可以利用RPKI形成的证书链来验证BGP前缀宣告的合法性。类似地,DNS系统也是以根服务器为中心的层次化系统,各级顶级域名服务器为中间节点,权威域名服务器为叶子节点。DNS系统的安全扩展DNSSEC仍然依赖于DNS的系统结构,自上而下逐级进行公钥背书和签名,其基本安全原理仍然是中心化的。而PKI系统也同样采取了具有中心根节点的层次化结构。

以上这类中心化的系统存在一些根本的问题。该类系统中,中心化的权威节点是系统的信任锚点,其权利过大,可以单方面地通过撤销证书或者数字签名,移除对后继节点的信任背书或者给虚假节点授信,从而损害真实节点的利益。由于这些基础设施是全球化的,所以一个中心节点的恶意操作带来的负面影响是全球性的。比如一个中心节点可以为其他国家的节点进行信任背书,撤销授信或者为虚假节点授信,都会对其他

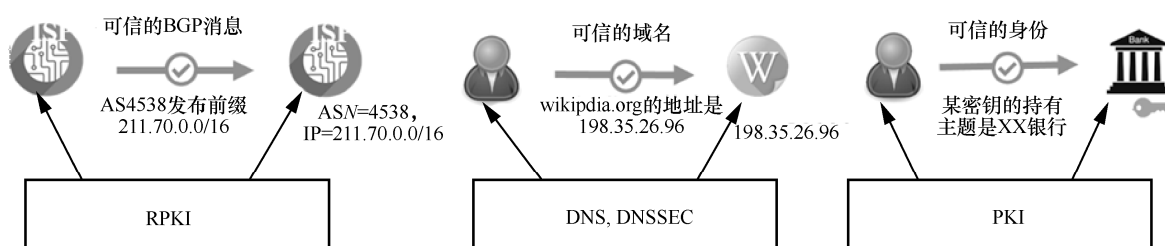


图1 互联网的重要基础设施

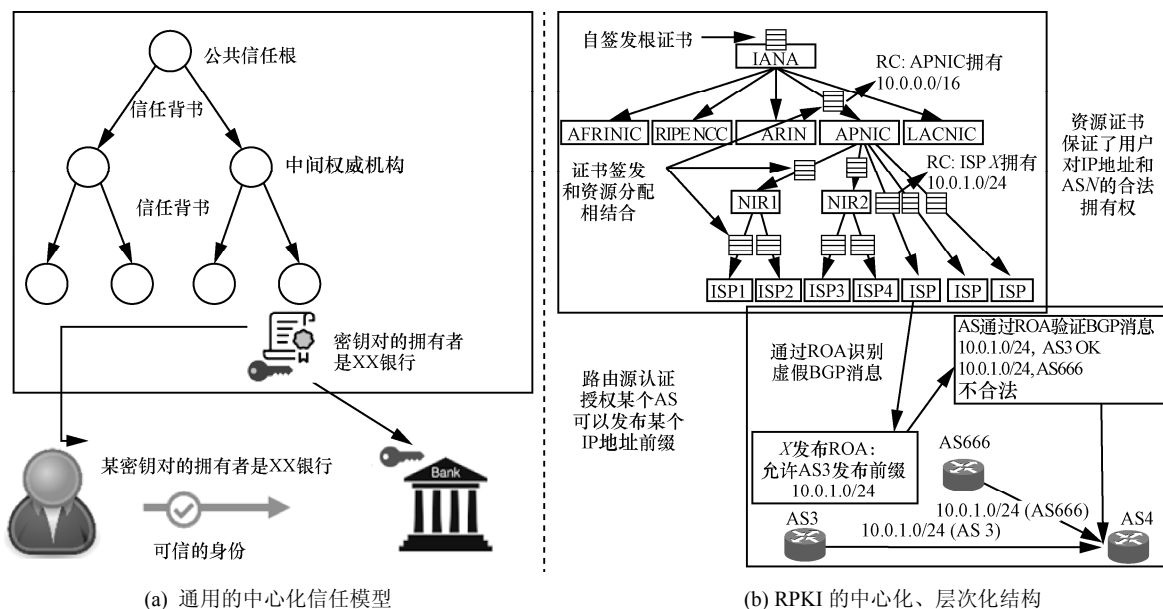


图2 当前互联网可信基础设施

国家的可信服务带来破坏性的影响。这些中心权威节点可能有各种各样的原因去做这样的操作，比如因为被黑客控制或者管理员错误配置。一个中心节点也可能无法做到完全中立，由于利益冲突、政治或法律原因而存在偏见。以下列举了一些现实中由于前述中心化架构而导致的安全事件。

- 2013 年，一条归属于某个俄罗斯网络的路由源认证（ROA）被从 RPKI 系统中移除，这导致相应的地址前缀在 RPKI 中被归属到了其他的运营商，使得原网络的 BGP 宣告变成 invalid（不合法），有可能使得该网络的相应前缀失去可达性^[1]。
- 2010 年，某域名被其权威服务器（位于美国的 DNS 服务提供商 EveryDNS）删除，导致该域名在全球范围内无法访问^[2]。
- 2011 年，荷兰著名的证书颁发机构（CA）DigiNotar 的服务器被黑客攻击，颁布了 531 个虚假证书，内容包括 Yahoo!、Mozilla、Google 等国际化企业的域名，导致很多用户被劫持到虚假网站，造成这些

企业和用户的损失^[3]。

综上所述，中心化的基础设施和其安全信任模型已经难以适应互联网的全球化属性，甚至有可能阻碍互联网长期可信和健康发展。因此，本文提议对基础设施进行去中心化改造，从而加强互联网的可信与平等基础，促进互联网的可持续发展。

在后文中，提出了一种去中心化的互联网基础设施（decentralized internet infrastructure, DII）的架构设计，并通过该架构来讨论基础设施去中心化的技术可行性以及其所带来的好处。本文将主要围绕基础设施去中心化带来的收益和挑战来阐述概要设计，并给出初步的可行方案，而 DII 系统和相关协议的细节实现和实验的数据并非本文关注重点，将不详细展开讨论。值得一提的是，互联网基础设施的去中心化改造是宏大而长期的议题，本文只发起初步的探讨，后续推进仍需与业界共同付出长期而艰辛的努力。

2 去中心化互联网基础设施的架构设计

本文提出的去中心化互联网基础设施（DII）的架构包含 3 个层次：底层利用分布式账本技术

构建基础的去中心化能力；中间层构建 IP 地址和域名等互联网名字空间的去中心化可信管理机制，并支持安全可信的域间路由和域名映射系统；顶层为开放的应用层，支持和促进创新、可信的去中心化互联网应用，如图 3 所示。分布式账本层为名字空间管理层提供去中心化的基础能力，主要关注构建去中心化的底层基础平台。名字空间管理层在分布式账本层提供的基础平台上，实现互联网名字空间的去中心化。同时，又为依赖这些名字空间的应用层 App 提供可信基础，实现更贴近用户业务的应用开发。

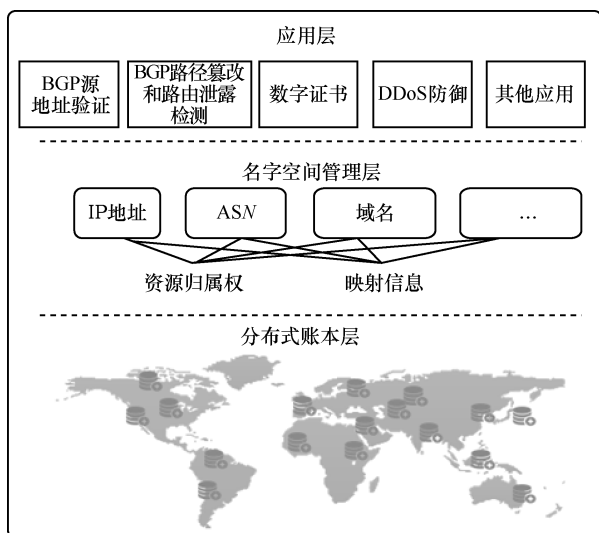


图3 DII的整体架构设计

3 分布式账本层

近些年来，以区块链为代表的分布式账本技术得到了快速的发展。然而，分布式账本更重要的价值，是作为去中心化的平台，支持去中心化的网络应用。DII 利用分布式账本的如下能力，构建去中心化的互联网基础设施。

(1) 去中心化的系统结构

系统中没有恒定的特权节点（尽管在某段时间内有些节点拥有更大的权力，但如果该节点用特权作恶，则会被其他节点替代），长期来看所有节点的地位是平等的。这符合去中心化互联网

基础设施的信任模型。

(2) 分布式共识机制

这是分布式账本技术的核心，所有节点采用去中心化的机制达成共识。DII 需要共识机制来实现互联网名字空间等网络资源所有权的唯一性以及相关应用的一致性。

(3) 智能合约

运行于分布式账本之上的计算环境，先进的智能合约能够支持图灵完备的计算模型，理论上支持任何应用程序。DII 对于互联网名字空间的管理需要复杂的逻辑，而开放的应用层则需要支撑任意的应用程序，因此需要智能合约技术。

(4) 可信交易的能力

各个账户之间可以通过交易相互转账，这赋予了分布式账本内在的价值传递能力。利用该能力，DII 不仅可以为第三方的开发者提供技术平台，还能够提供技术变现能力。

3.1 DII 对分布式账本的技术需求

当前的分布式账本主要分为两类：公有链和联盟链，然而这两类分布式账本均存在一定问题，不能直接用于 DII 系统。

公有链以比特币^[4]和以太坊^[5]为代表，运行在公有互联网之上，开放给任意节点加入，具有很好的动态性。但是由于其不对加入节点做任何限制，导致节点数量过于庞大且系统过于暴露，进而导致系统共识效率低下，且恶意攻击层出不穷。DII 作为互联网核心资源（IP 地址、ASN、域名等）的管理平台，从安全的角度考虑，不希望任何人都能不加约束地随意加入；从需求角度讲，也并非任何人都有申请核心资源的需求（例如只有 ISP 或者较大组织才有申请大块地址的需求）。公有链并不满足以上要求。

联盟链以超级账本^[6]为代表，一般运行在局部网络中，只有被认证准入的节点才能够加入，节点数量较少。相比公有链，联盟链具有更高的性能和更好的安全性，即使出了问题也可以通过



认证渠道进行追责和解决。但是，联盟链技术难以直接应用于 DII。这是因为 DII 作为全球互联网的基础设施，参与的节点数量是非常庞大的。而当前联盟链依赖静态配置的准入列表，难以适应庞大系统的动态性。比如 Hyperledger 就需要手动静态配置节点列表，列表的更新需要系统重启，甚至会引入中心化管理配置。

3.2 准入控制机制

现有公有链没有准入机制，但有很好的动态性，节点可以动态加入和下线。而联盟链拥有准入机制，但是系统扩展性一般较差，容纳节点数有限。为了仅允许合格的组织（如运营商、企业、高校）等动态加入系统申请资源，简单给出一种同时实现背书准入机制和动态节点管理的机制，该机制允许 DII 系统中的已有背书节点为新节点背书或者停止背书，从而实现节点加入和离开。

该机制的基本原理如图 4 所示。DII 系统初始化会有一个背书者（endorser）列表，该列表中记录了 DII 系统内有背书能力的节点信息。所有要加入的节点均需要通过一个或多个背书者的背书才允许加入该系统。图 4 中以各大地区的互联网注册中心 RIR（AFRINIC、APNIC、ARIN、LACNIC、RIPE NCC）以及各个国家的注册中心 NIR（如 CNNIC、JPNIC 等）作为背书机构，只

有被至少一个背书机构准入的组织才有资格加入 DII，其账户信息才被写入由智能合约维护的准入节点列表。实际使用中，该列表的初始化确定需要通过线下的沟通和协商。背书机构的名单也可以增加或减少，这由背书机构彼此投票来确定。具体的加入流程如下。

（1）假如某个 ISP X 要加入 DII 系统， X 需要首先在本地下载 DII 系统的区块链客户端，并生成自己的账户（account）和节点标识符（node ID）。

（2） X 通过线下的方式向某个背书者申请背书，并将自己的信息（加入 DII 设备的 IP 地址信息、 X 的账户和节点标识符）发送给某个背书者 A 。同时 X 还要通过线下方式获取背书者 A 的相关信息（如能证明其身份的某个公钥）。

（3）背书者 A 在 DII 系统中发起背书交易，将 X 的节点标识符信息写入区块链的背书清单（endorsement list）中。背书清单保存有当前系统里所有的节点背书信息。

（4）背书者 A 向 X 提供的 IP 地址发起节点询问（ping node）消息，确定该背书申请确实是 X 发起的；节点询问消息中还要携带 A 的签名，用于 X 验证 A 的身份。节点询问消息为以太坊系统中节点间的握手消息为准，此处以此消息为例，不同系统可能有不同消息，但功能大致类似。

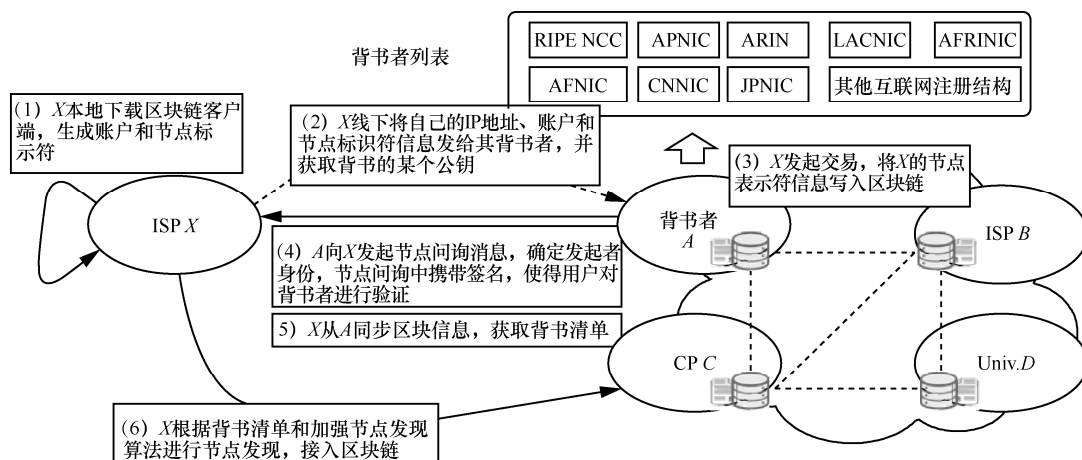


图 4 DII 准入控制机制基本原理

(5) X 验证 A 的身份后, 从 A 同步区块信息, 并获取背书清单。

(6) X 根据背书清单和相关节点发现算法进行节点发现, 接入 DII 系统。

虽然 DII 系统中仍然存在 RIR 和 NIR 等作为背书者的角色存在, 但他们的职能与现有 RPKI 等中心化系统不一样。首先, 在 DII 中, 这些背书者只做申请组织的身份背书和准入认证, 但并不主导资源分配。一个组织一旦被准入, 即可独立地申请和拥有相关资源, 且资源归属权不依赖于背书者。其次, 一个组织可以从多个背书机构得到背书, 因此即便有背书机构单方面撤销背书, 该组织仍然符合准入条件。因此, DII 大大消除了这些组织的中心化权利。

以上只是给出了准入控制的一个基本原理, 在具体实现过程中, 可以依照上述原理对已有的分布式账本技术 (如以太坊等) 进行改造和增强, 也可以按照实际需求实现定制化的分布式账本。此外, 可以根据不同的应用需求来设计多个背书者列表, 并根据不同的背书者列表实现相应的准入控制实例; 这些背书者列表等信息以及后续所有 DII 系统维护的信息均可以通过智能合约或者类似的方式来实现。总之, 以上问题都是需要在未来的研究和具体实现中进一步详细设计和深入解决的, 本文不再详细展开。

4 名字空间管理层

互联网名字空间是 TCP/IP 协议的核心, 也是互联网基础设施的核心。互联网最重要的两大基础设施, 即 BGP 与 DNS, 都是围绕名字空间展开的。其中 BGP 将 IP 地址前缀映射到 ASN (AS 号) 和 AS path (AS 路径), 从而计算出 IP 地址空间的域间路由。可信的 IP 地址前缀归属权和映射关系十分重要, 否则将存在 BGP 前缀劫持和路径劫持等网络攻击。DNS 将域名空间映射到 IP 地址空

间, 从而让服务在网络层能够被访问。可信的域名归属权和映射关系也十分关键, 否则将存在域名缓存污染等漏洞。

可信可靠的名字归属和名字映射是基础设施可信可靠的基础。正因如此, 将名字空间管理层作为 DII 架构的中间层, 通过去中心化的可信的名字归属和映射, 保障互联网基础设施的安全可信, 进而支持可信的上层应用。

4.1 IP 地址和 AS 号管理

当前 IP 地址和 ASN 的可信基础设施为 RPKI。RPKI 沿用了全球 IP 地址分配系统的中心化树状结构, 一方面通过中心权威机构保障 IP 地址和 ASN 的保证归属权的唯一性, 另一方面要评估申请者的合理使用需求, 在分配过程中防止名字空间的耗尽并防止路由碎片化。

下面分析采用去中心化的方式实现全球 IP 地址分配的可行性。如果 IP 地址初始归属权确定, 则 IP 地址转移可以采用交易的方式在分布式账本上完成, 过程相对简单, 不展开阐述。此处主要介绍地址的初始分配过程。由于 IPv4 地址空间已经分配完毕, 此处聚焦 IPv6 地址。DII 系统中, 考虑 IPv6 地址分配的基本单位为 /32, 与 32 位 ASN 的可分配数量是相近的。

IPv6 地址分配的系统如图 5 所示。已被准入的组织可以发起 IPv6 地址申请。以 ISP B 为例, 它首先发起一个分布式账本交易, 内容是申请一个 /32 IPv6 地址的一年使用权, 并支付地址使用年费。其他节点收到该交易后, 通过智能合约检查申请者的合法性和年费, 并采用稀疏委托 (sparse delegation) 算法为 ISP B 计算出一个适合它的地址前缀。该算法的基本功能是为每个申请者计算连续的地址空间, 从而防止地址碎片化及路由膨胀。由于算法是确定的, 全网对地址的分配也就是一致的、唯一的。使用权到期之前, 申请者要发起一个续约的交易并续交年费, 否则该前缀将被智能合约重新放回地址池。

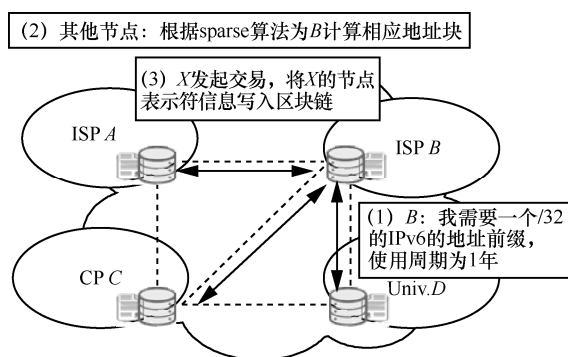


图5 去中心化的IP地址管理

采用上述的地址分配方法，IPv6的地址耗尽攻击是很难实现的。当前全球针对IPv6/32前缀的年费价格在2000美元左右。以该价格计算，耗尽全部IPv6地址空间需要约 8×10^{12} 美元，约占全球GDP的10.5%。即便有单一利益集团支付了 this 金额，也只是耗尽了一年的使用权，过期之后其他组织仍可正常申请地址。

ASN也可以通过类似的方法得以实现，受限于篇幅，本文不再赘述。

4.2 域名管理

域名的管理与IP地址有所不同。首先，域名是层次化的，IP地址空间是扁平的；其次，域名空间现实中是不可耗尽的，而IP地址空间是有限的。对于DII来说，最应关心的是属于全人类共同资产的域名空间，而非归属于某个国家或机构的域名空间。本文重点讨论通用二级域名（如example.com、example.net等）。这是因为：

- .com、.net等通用顶层域名属于人类共同资产，可以自由申请的就是相应的二级域名；
- 在二级域名归属权确定后，相应的三级域名一般受同一管理域内管理，属于私有资产，不在去中心化管理范围内。此外，国家码顶级域名（.cn、.jp等）需要线下协商，受篇幅所限不在本文中讨论。

域名管理逻辑可以在单独的智能合约中实现。不同于IP地址申请者（以大型组织结构为主），域名申请者包括大量的个人和小微组织，其部署

分布式账本、承担维护开销的意愿较低。因此，可以利用当前域名管理体系中的域名中介代替申请者参与去中心化域名申请过程，同时要避免权力的集中化。

去中心化的域名申请与转移如图6所示，二级域名（SLD）的申请者首先要生成一对公私钥（公钥 pk_X 和私钥 sk_X ）。它将自己的公钥 pk_X 和要申请的域名example.com提交给某个域名代理A。A先检查该域名是否仍然闲置，如果闲置，则在分布式账本中发起一个交易，内容是 pk_X 申请域名example.com。其他节点收到交易后，同样要检查域名的可用性，如果可用，则写入“example.com的所有者是 pk_X ”。域名的申请与IP地址不同，前者采取“先到先得”的办法，后者需要用算法防止地址碎片化。域名使用年费以及续期和逾期的处理办法与IP地址类似，这里不再赘述。

虽然DII采用了域名中介在分布式账本上代替申请者管理域名，但域名的所有权仍然受申请者自身控制，不存在中心化问题。这是因为，分布式账本中写入的是域名归属于 pk_X ，而只有持有相应私钥 sk_X 的申请者才能证明对域名的拥有权。即使域名代理A不为该申请者提供服务，申请者也可以通过其他的中介进行域名管理，而不会被中介绑定或挟持。以图6中的域名转移为例，域名持有者需要通过私钥签名，才能操作相关域名的转移。其申请时可以通过中介A，转移时可以通过中介B，从而独立掌握了域名的控制权。

由于域名解析数据量巨大且动态性极高，把全部解析数据（如将域名www.example.com映射到某些IP地址）全部存储在分布式账本上是不现实的。因此，DII仅将对安全域名解析最关键的信息存储在分布式账本中。域名example.com的持有者仅将自己的公钥和域名的权威名字服务器的地址存储在分布式账本中，如图7所示。这些信

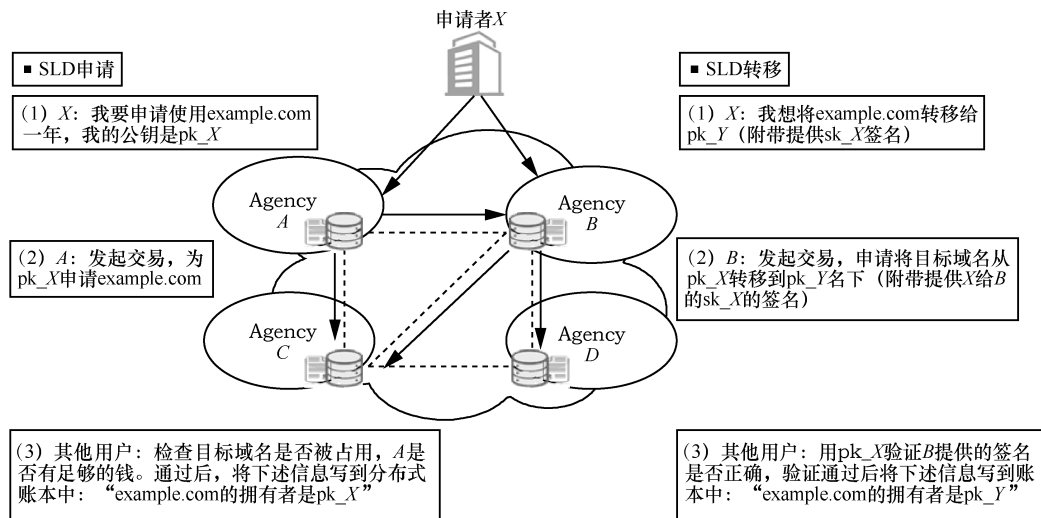


图6 去中心化的域名申请与转移

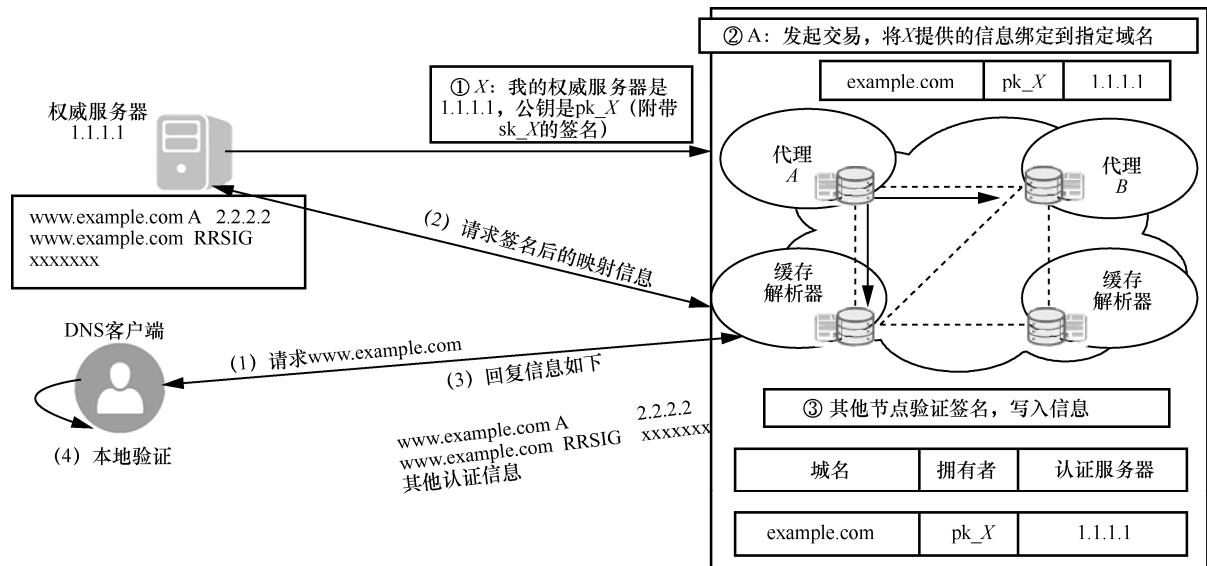


图7 安全高效的域名解析

信息的特点是占用的存储量小、动态性小, 不会对分布式账本的性能和可扩展性带来过大的挑战。而存储量大、动态性高的信息存储在权威服务器中。

为了保证域名解析数据不被篡改和污染、提供安全的域名解析, 域名持有者需要对解析数据签名。DNS解析服务器根据分布式账本中存储的权威服务器地址发起解析请求, 权威服务器将解析数据和签名均应答给解析服务器。解析服务器或客户端可以使用存储在分布式账本中的公钥对解析数据和签名进行验证, 保证域名解析的安全。

5 应用层

底层分布式账本的基础能力和中间层可信的名字空间管理能够支撑安全可信的去中心化网络应用, 比如基于可信ASN和在线交易能力的跨域端到端服务质量保障, 基于可信IP地址和远程可信交易的近源DDoS防御服务等。本节主要描述去中心化的BGP源地址验证、BGP路径篡改检测、路由泄露检测以及去中心化的公钥基础设施等若干个应用。



5.1 BGP 源地址验证

明确了地址和 ASN 归属权后, 地址的所有者就可以发布 ROA 在路由系统中保护自己的前缀。如图 8 所示, ISP B 可以发起交易, 内容是 IP 地址前缀到 ASN 的映射, 授权某些 ASN 发布相应的前缀。其他节点收到该交易后, 先验证地址的所有权, 验证通过后再写入该 ROA。可见, 地址的所有者对于 ROA 有完全的控制权限, 并不依赖于任何第三方权威或者相关资源证书。各个网络可以读取 ROA, 并生成路由的验证列表, 写入 BGP 路由器, 进行 BGP 源地址验证, 提升路由安全。

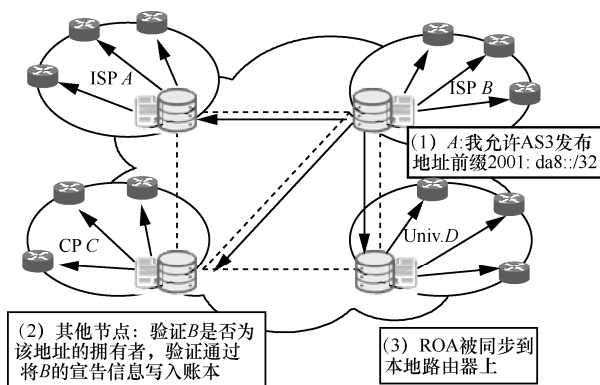


图 8 BGP 源地址验证

5.2 BGP 路径篡改检测

当恶意 AS 通过发布虚假 AS path 劫持路由时, 通过现有 BGP 本身很难发现 AS path 被篡改, 导致路由被劫持到恶意 AS, 造成巨大危害。图 9 给出了一种 BGP 路径篡改的例子, AS400 伪造前缀 20.20.0.0/16, 并发布虚假 AS path (400, 200)。AS600 收到 2 条到达 20.20.0.0/16 的 AS path (400, 200) 和 (500, 100, 200), AS600 根据最短路径原则, 优选 AS400 到达 20.20.0.0/16。结果是, AS600 中目的地址为 20.20.0.0/16 的流量被劫持到 AS400 中。

基于 DII 系统, 可以采用如下的 BGP 路径验证方法来检测 BGP 路径篡改。

AS 的拥有者可以通过交易的方式在 DII 上发

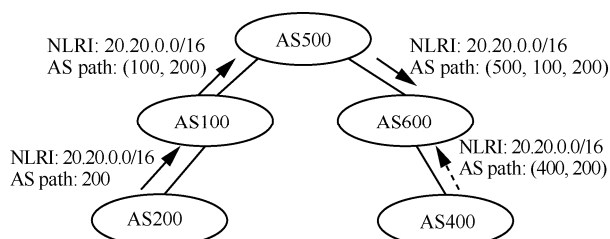


图 9 BGP 路径篡改

布和其他 AS 之间的邻居关系, 其他节点会验证发布者是否为该 AS 的真正拥有者, 验证通过后, 该 AS 的邻居关系会被其他节点写入 DII 系统。由于发布虚假信息会导致某个 AS 的流量被劫持或进入黑洞, 受害者只有该 AS, 因此, AS 的真正拥有者不会有动机来发布虚假信息, 即 DII 系统维护的 AS 邻居关系表是安全并且可信的。路由器收到 BGP 前缀更新消息时, 可以利用 DII 系统维护的 AS 域间邻居关系, 对 BGP AS path 信息进行逐跳验证。

仍然以图 9 为例。假设 AS200 发布的邻居关系见表 1, 当 AS600 收到 AS400 发布的 AS path (400, 200) 时, 可以通过表 1 验证 AS200 同 AS400 域间邻居关系的可信性。显然 AS200 没有发布同 AS400 域间邻居关系, 表明 AS200 没有同 AS400 建立域间邻居关系。因此 AS600 可以识别出 AS400 发出的 BGP path 信息为虚假信息。

表 1 AS200 发布的域间邻居关系

ASN	邻居 ASN	关系
200	100	提供商 (provider)
200	300	对等者 (peer)

5.3 BGP 路由泄露检测

互联网工程任务组 (IETF) 的标准 RFC7908^[7] 针对 BGP 路由泄露进行了分类, 本节提供检测该 RFC 定义的前 4 种路由泄露类型的方法。具体路由泄露类型如下。

- AS 将接收自提供商的路由通告给了其他的提供商;

- AS 将接收自对等者的路由通告给了其他的对等者；
- AS 将接收自提供商的路由通告给了它的对等者；
- AS 将接收自对等者的路由通告给了它的提供商。

AS200 收到来自提供商 AS100 的前缀 30.30.0.0/16, 通告给了它的对等者 AS300, 该行为符合上述 BGP 路由泄露类型 3, 如图 10 所示。AS300 根据现有的 BGP 无法识别出 AS200 泄露了前缀 30.30.0.0/16, 如果 AS300 到 30.30.0.0/16 的流量优选 AS200 转发, 导致流量被错误地吸引到 AS200, 如果 AS200 负载能力不足, 可能导致流量转发延迟或者被丢弃。

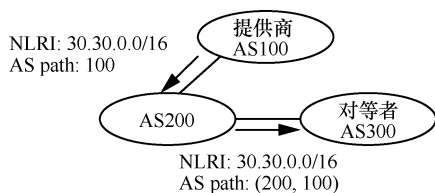


图 10 BGP 路由泄露

DII 系统提供的域间邻居关系同样也可以支持 BGP 路由泄露检测。以图 10 为例, 假设 AS200 发布邻居关系见表 1。当 AS300 收到来自对等者 AS200 的前缀更新消息时, 可以通过查询 AS 域间邻居关系结合 AS path 对路由通告的合法性进行检测。通过 AS path (200, 100) 得知 AS200 发布的路由来自 AS100, 通过域间邻居关系表得知 AS100 是 AS200 的提供商。AS200 将来自提供商 AS100 的前缀发送给对等者 AS300, 符合 BGP 路由泄露类型 3, 检测出 BGP 路由泄露。

5.4 去中心化的公钥基础设施

传统网站的可信性主要通过 PKI 证书来保障两类信息的实现: 域名的所有权和运营者的真实身份。只提供域名所有权证明的证书被称为 DV 证书, 而同时提供这两类信息保障的证书

被称为 EV 证书, 一般认为后者的可信性更高。基于第 4 节中介绍的域名管理, DII 可以利用智能合约提供的信息来提供网站运营者对该域名的所有权, 即 DV 证书不再依赖任何第三方数字证书认证机构 (CA), 网站运营者可以自行向其用户提供智能合约相关信息来证明其对该域名的所有权。

在此基础上, 可以在 DII 中引入 CA 角色作为参与节点。CA 节点可以通过智能合约发布对其他节点账户的身份背书信息。逻辑上, 该身份背书信息完全独立于域名管理系统信息。如果某个节点在已有身份背书的同时也拥有某些域名, 则其可以同时提供其网站用户相关的域名所有权信息和身份背书信息。用户可以基于 DII 来对这两类信息进行验证, 从而实现 EV 证书级别的认证。

需要说明的是, 网站在通信中使用的公钥仅与域名所有权绑定, 而不与身份背书绑定。因为域名所有权的证明已经完全地去中心化, 安全性更高, 而后者仍然依赖于 CA。

在 DII 中, 一个网站可同时从多个 CA 获得身份背书, 以避免当前 PKI 证书体系下多个 CA 单方撤销证书而造成网站信任锚点的情况。在最坏的情况下, 即使所有 CA 同时取消对某个节点账户的身份背书, 也无法剥夺其域名所有权, 最多将该账户所运营网站的可信性从 EV 级别降到 DV 级别, 从而削弱了中心化 CA 单方面操作造成的影响。

6 名字空间管理层演进方案

由于现有的互联网核心资源 (包括 IP 地址、ASN 和域名) 的分配和管理已成为事实并广泛使用, 不可能一次全部转移到 DII 系统上, 本节提供了一种可以从当前集中式资源管理方式进行演进方案, 在不影响资源使用的基础上, 逐步实现核心资源的去中心化管理。



6.1 IP 地址管理和 ROA 能力

DII 系统可以在不改变现有 IP 地址分配系统的基础上,通过收集现网 BGP 前缀信息、获取 IP 地址归属权和 ASN 映射关系,进而维护在 DII 系统中,为后续应用提供真实可信的信息。具体方案如图 11 所示。

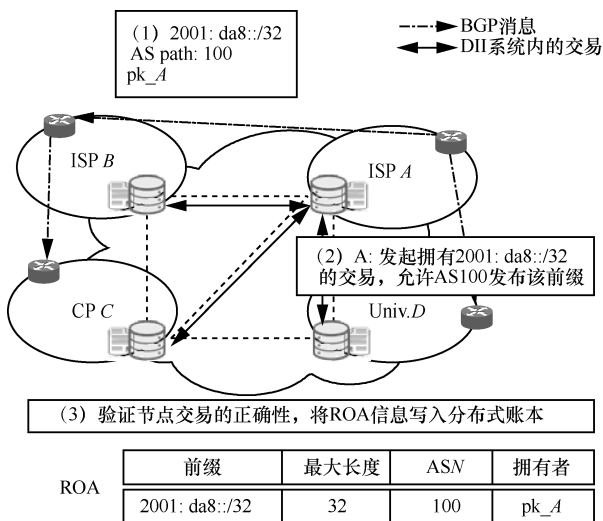


图 11 IP 地址管理演进方案和 ROA 管理

步骤 1 ISP A 通过现有 IPv6 地址分配机制获得地址 2001:da8::/32。ISP A 向其他 AS 发布 BGP 消息,该消息携带地址前缀 2001:da8::/32、AS path 100 以及所有者的公钥信息 pk_A。该消息实际上实现了两个功能,一是将 IP 地址 2001:da8::/32 和公钥 pk_A 进行了绑定,在步骤 3 中被其他节点用来验证该 IP 地址在 DII 系统中的归属权;二是将 IP 地址 2001:da8::/32 和 AS100 进行了绑定,在步骤 3 中被其他节点用来验证 ROA 详细的合法性。公钥信息 pk_A 可以通过扩展现有的 BGP 更新(update)消息来携带,具体的消息格式不在本文阐述。

步骤 2 ISP A 在 DII 系统中发起 IP 地址归属权和 ASN 映射关系的交易。交易内容是“公钥 pk_A 所对应的私钥持有者(即 ISP A)是 IPv6 地址 2001:da8::/32 的所有者,并通过 AS100 发布相应的前缀”。为了防止作恶,路由被短暂地劫持,

该路由需要在网络中稳定存在一段时间。例如 10 天内,稳定存在 5 天以上,其他节点才会认为 ISP A 是该地址的真正所有者。ISP A 待路由满足上述规则后,再发起交易,确保网络中的节点都认同该前缀相关信息的可信性,避免交易失败。

步骤 3 验证节点收到交易后,获取 IP 地址前缀、所有者的公钥信息和授权进行 BGP 宣告的 ASN。然后验证节点再从路由器获取该地址前缀的相关路由信息,包括在本地存储路由的存活时间、ASN、地址所有者的公钥等信息,对交易的信息进行验证。如果路由器上的消息和交易消息一致,则验证通过。验证节点达成共识后,将 IP 地址的归属信息和 ROA 信息写入分布式账本。

通过上述方法,路由的 ROA 存储在分布式账本中。各个网络可以读取 ROA,并生成路由的验证列表,写入 BGP 路由器,进行 BGP 源地址验证,提升路由安全^[8]。地址的所有者对于 ROA 有完全的控制权限,并不依赖于任何第三方权威,避免中心节点单方面撤销授信或者为虚假节点授信,对可信服务带来破坏性的影响。

6.2 ASN 管理和域间邻居关系

DII 系统可以在不改变现有 ASN 分配系统的基础上,通过收集现网 BGP 信息,将 ASN 的归属信息存储在分布式账本中,进而支持 AS 所有者在分布式账本中存储可信的 AS 域间邻居关系,实现去中心化的 AS 域间邻居关系管理。ASN 的管理类似于 IP 地址管理方式,具体方案如图 12 所示。

(1) ISP B 通过现有 ASN 分配机制获得 ASN: 200。ISP B 向其他 AS 发布前缀信息时,携带 ASN: 200 和该 ASN 所有者的公钥信息 pk_B。该方案需要扩展 BGP 更新消息,携带 ASN 所有者的公钥信息,从而绑定 ASN 和所有者的关系,具体的消息格式不在本文阐述。

(2) ISP B 发起拥有 ASN: 200 的交易,交易内容是“公钥 pk_B 所对应的私钥的持有者(即 ISP B)是 AS200 的所有者”。为了防止作恶,

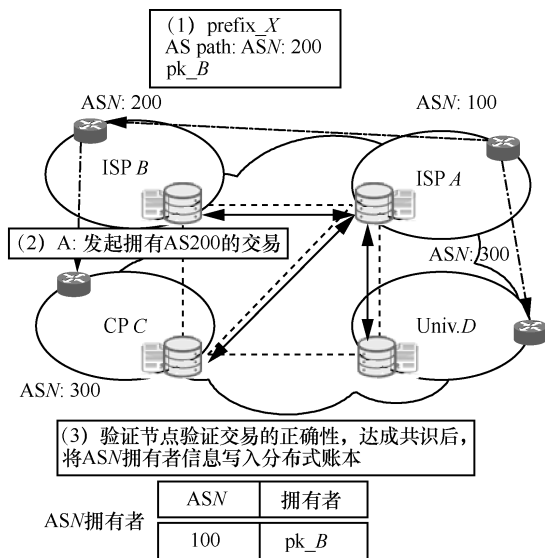


图 12 ASN 管理演进方案和邻居关系管理

ASN 信息需要在网络中稳定存在一段时间, 再发起交易, 确保网络中的节点都认同 ASN 所有者信息可信性, 避免交易失败。

(3) 验证节点收到交易后, 根据 ASN 获取 ASN 所有者的公钥信息, 验证通过后接收该交易。验证节点达成共识后, 将 ASN 所有者的信息存储在分布式账本中。

明确了 ASN 的所有权后, ASN 的所有者可以发起该 AS 域间邻居关系的交易。验证节点收到该交易后, 先验证 ASN 的所有权, 只有 ASN 的

所有者才能发起 ASN 的相关交易。验证节点达成共识后, 将 AS 域间邻居关系存储在分布式账本中, 见表 1。可信的域间邻居关系可以应用于 BGP 路径篡改和路由泄露检测, 方法详见第 5 节。

6.3 域名管理

DII 可以在不改变基于现有域名分配系统的基础上, 基于既成事实的域名所有权状态来实现 DII 系统里的域名所有权的初始化。基本原理如图 13 所示。

(1) 假设某个用户 A 已经通过现有域名管理系统获得域名 example.com。用户 A 需要启用该域名, 并将该域名对应网站的 IP 地址维护到现有的 DNS 系统中。对应的网站 (以 www.example.com 为例) 需要有提供验证信息的能力。

(2) 用户 A 在 DII 系统上发起交易, 宣告对 example.com 的拥有权, 并将该域名和某个公钥 PK 绑定, 即声明公钥 PK 对应的私钥的持有者才是该域名的拥有者。

(3) 验证节点 (如 C) 收到交易后, 从当前 DNS 系统中获取目标网站对应的 IP 地址。

(4) 验证者 C 和该网站建立连接并执行验证; 验证通过后, C 认为该交易合法。

(5) 多数节点验证通过后, 该域名的所有权

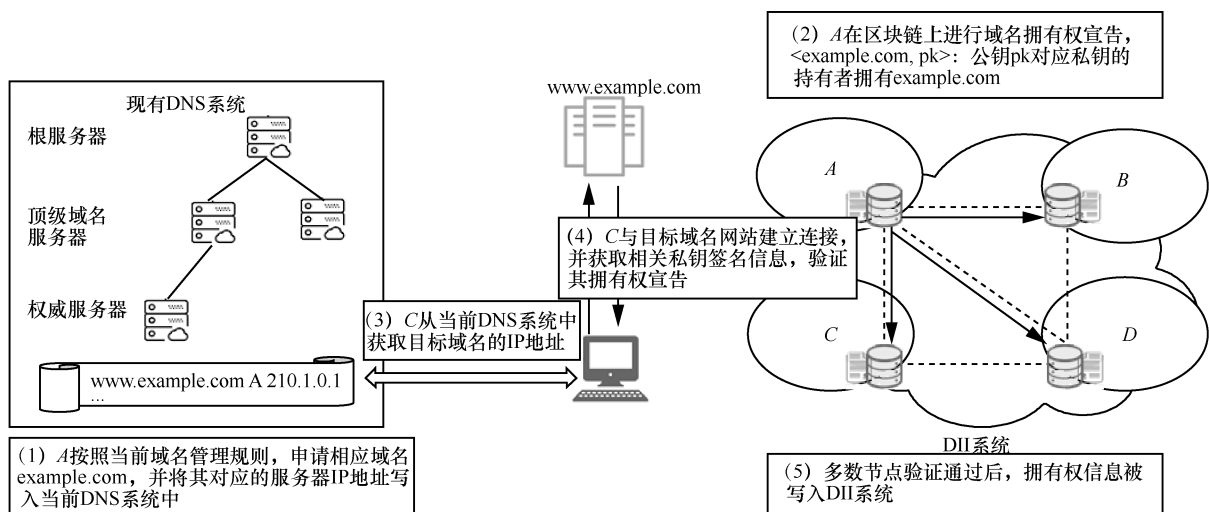


图 13 域名管理过渡方案



信息被写入 DII 系统。

验证的方式取决于网站自身, DII 不做任何限制, 本文提供两种思路供读者参考。第一种方式, 网站上可以发布某个随机数和相关私钥 sk 对应的签名, 验证者凭 DII 系统上获取的公钥 pk 可以对签名进行验证, 验证通过则说明该域名的拥有者持有的私钥 sk 和 DII 上发布的公钥 pk 是一对密钥对, 即私钥 sk 的持有者就是域名的拥有者。第二种方式, 网站可以使能某种交互能力, 验证者生成一个随机数, 用公钥 pk 加密并发送该网站, 要求网站回复私钥 sk 解密后的随机数, 解密的随机数正确, 说明该网站拥有 DII 系统上公钥 PK 对应的私钥 sk 。

明确了域名的所有权后, 域名的所有者可以撤销现有 DNS 系统内的发布信息, 并将该信息维护在 DII 系统上, 进而可以用图 7 所示的方式来提供 DNS 服务。

以上 3 种过渡方案仅给出了如何将既成的资源所有权事实同步到 DII 系统中。一旦资源所有权信息同步到 DII 系统中以后, 就可以在 DII 系统上增加新的资源回收能力, 资源过期后由智能合约回收到 DII 资源池。资源池中的资源则可以使能第 4 节所述完全去中心化的管理方式。资源的回收和再分配将受到国际机构间的策略和博弈的影响, 超出了技术范畴, 本文不再详细讨论。

7 相关工作

互联网基础设施的去中心化议题已经得到了越来越多的关注。国际互联网工程任务组 (IETF) 成立了去中心化互联网基础设施研究组 (DINRG)。在 DINRG, 斯坦福大学的团队提出了适合互联网基础设施的去中心化一致性协议 SCP^[9]; 伯克利大学提出去中心化的映射系统^[10]; 去中心化身份基金会推动去中心化身份系统^[11], 受到了学术界和工业界的广泛关注。

针对 RPKI 的中心化问题, Cloud Flare 推出

了 RPKI Transparency 技术和产品^[12], 但只是采用事后审计的方式来检查中心节点的恶意行为, 无法做事前预防。单点问题的本质是中心节点的权威和数据, DII 通过分布式共识取代了中心权威节点, 通过分布式账本取代了中心数据库, 从根本上解决了中心化问题。西班牙加泰罗尼亚理工大学 (UPC) 分析采用 Proof of Stake 算法管理 IP 地址空间^[13]。西班牙马德里卡洛斯三世大学 (UC3M) 则基于以太坊开展了用区块链管理 IP 地址空间的实验^[14]。

针对于域名系统的中心化问题, 业界已经有多个去中心化域名项目, 如 Namecoin^[15]、BNS (Blockstack Naming System)^[16]、ENS (Ethereum Naming Service)^[17] 等。这些方案都聚焦于解决中心化问题, 但也带来了域名解析过程中的其他问题。Namecoin 和 BNS 中, DNS 客户端仍然需要无条件信任某些本地的域名解析器, 而无法对解析结果进行验证。这些本地域名解析器可通过单方面的信息篡改对客户端和域名持有者作恶。ENS 中, 客户端虽然可以对解析结果进行验证, 但是每次请求需要付出巨大的验证开销。

针对 PKI 的中心化问题, 谷歌提出了证书透明化 (certificate transparency, CT) 方案^[18]。该方案利用一个额外的 Log 系统来记录证书发布的所有历史, 同时各网站合法运营者会通过监控器 (monitor) 实时监测相关证书的发布行为。一旦出现 CA 作恶的情况, 监控器会向合法的网站运营者进行告警。CT 在一定程度上可以缓解 PKI 体系的中心化问题, 但由于它是一种被动式的防御方案, 只能在事后进行弥补, 无法从根本上阻止中心化问题带来的恶意行为。

8 结束语

互联网作为人类共同的通信网络, 其基础设施只有做到平等, 才会得到所有人的信赖, 才会支撑互联网的长期健康发展。本文仅尝试对互联

网基础设施的去中心化进行初步的探讨,受篇幅所限难以深入技术细节。互联网技术发展至今,基础设施相关的技术和运行都积累了大量的知识和经验,去中心化的过程势必十分艰难。此外,分布式账本技术用于全球互联网基础设施仍待进一步地研究和试验验证,名字空间管理方案也需要与业界共同探讨与完善。期待这个议题能够得到业界更加广泛的关注,通过共同的讨论和试验推进去中心化互联网基础设施的发展和成熟。

参考文献:

- [1] HEILMAN E, COOPER D, REYZIN L, et al. From the consent of the routed: improving the transparency of the RPKI[C]//ACM SIGCOMM Computer Communication Review, August 17 - 22, 2014, Chicago, Illinois, USA. New York: ACM Press, 2014, 44(4): 51-62.
- [2] VANCE A. WikiLeaks struggles to stay online after attacks[EB]. 2010.
- [3] ZETTER K. DigiNotar files for bankruptcy in wake of devastating hack[EB]. 2011.
- [4] NAKAMOTO S. Bitcoin: a peer-to-peer electronic cash system[EB]. 2008.
- [5] WOOD G. Ethereum: a secure decentralised generalised transaction ledger eip-150 revision [EB]. 2017.
- [6] CACHIN C. Architecture of the hyperledger blockchain fabric[C]//Workshop on Distributed Cryptocurrencies and Consensus Ledgers, July 25, 2016. [S.l.:s.n.], 2016.
- [7] IETF. Problem definition and classification of bgp route leaks:7908[S]. 2016.
- [8] LIETF. An infrastructure to support secure internet routing: 6480[S]. 2012.
- [9] IETF. The stellar consensus protocol (SCP):raft-mazieres-dinrg-scp-05 [S]. 2018.
- [10] IETF. Delegated distributed mappings:draft-wastom-dinrg-delmap-01[S]. 2018.
- [11] AW N. Decentralized Identity – what lies ahead of Us: the open (Interesting) research issues[EB]. 2018.
- [12] Cloudflare. Cloudflare cirrus[EB]. 2018
- [13] IETF. An analysis of the applicability of block chain to secure ip address allocation, delegation and bindings: draft-paillisse-sidrops-blockchain-01[S]. 2017.
- [14] ANGIERI S, GARCÍA-MARTÍNEZ A, LIU B, et al. An experiment in distributed Internet address management using blockchains[J]. arXiv:1807.10528, 2018.
- [15] KALODNER H A, CARLSTENA M, ELLENBOGEN P, et al. An empirical study of name coin and lessons for decentralized namespace Design[C]//WEIS, June 2015, Delft Netherlands. [S.l.: s.n.], 2015.
- [16] ALI M, NELSON J, SHEA R, et al. Blockstack: a global naming and storage system secured by blockchains[C]//2016 {USENIX} Annual Technical Conference, June 22-24, 2016, Denver, USA. [S.l.: s.n.], 2016: 181-194.
- [17] Ethereum. Ethereum Name Service[EB]. 2018.
- [18] IETF. Certificate transparency:6962[S]. 2013.

[作者简介]



刘冰洋 (1985-), 男, 博士, 华为技术有限公司主任工程师, 主要从事网络技术方面研究工作, 包括互联网体系结构、网络协议、安全可信、服务质量等。



杨飞 (1975-), 男, 博士, 华为技术有限公司主任工程师, 主要从事网络技术方面研究工作。



任首首 (1988-), 男, 博士, 华为技术有限公司高级工程师, 主要从事网络技术方面研究工作。

魏鑫鹏 (1986-), 男, 华为技术有限公司高级工程师, 主要从事网络技术方面研究工作。

杨雪 (1980-), 女, 华为技术有限公司高级工程师, 主要从事网络技术方面研究工作。

王闯 (1975-), 男, 华为技术有限公司高级技术专家, 主要从事网络技术方面研究工作。

延志伟 (1985-), 男, 博士, 中国互联网络信息中心基础技术实验室副主任、研究员, 主要从事互联网名址路由关键技术及下一代网络架构方面的研究工作。