可信互联网中服务提供者身份的可信问题①

卢文哲[1,2,4], 马 迪[1,2], 毛 伟[1,2,3,4]

[1](北龙中网(北京)科技有限责任公司, 北京 100190)
[2](互联网域名系统北京市工程研究中心, 北京 100190)
[3](中科院计算机网络信息中心, 北京 100190)
[4](中国科学院大学, 北京 100049)

摘 要: 可信互联网是近年的研究热点. 介绍了当前互联网在网域路由、域名解析和可信应用 3 个层面在身份可信方面的主要进展, 包括 RPKI、DNSSEC 以及网站可信标识体系, 介绍了 3 项技术的主要原理和作用, 同时介绍了三项技术目前在互联网社区部署的进展情况. 最后特别强调了网站可信标识体系的重要性和意义.

关键词: 可信网络; 可信互联网; RPKI; DNSSEC; 网站可信标识

Trusted Identification of Trusted Internet

LU Wen-Zhe[1,2,4], MA Di[1,2], MAO Wei[1,2,3,4]

[1](KNET Co., Ltd., Beijing 100190, China;)
[2](Internet Domain Name System Beijing Engineering Research Center, Beijing 100190, China)
[3](Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)
[4](University of Chinese Academy of Sciences, Beijing 100049, China;)

Abstract: Trusted internet has been a research hotpot in recent years. This paper introduces trusted identification of inter-domain routing, DNS resolving and trusted-application, including PRKI, DNSSEC and trusted website identification. Finally, we emphasize specially on the importance and significance of trusted website identification.

Key words: trusted network; trusted internet; RPKI; DNSSEC; trusted website identification

1 概述

随着互联网的快速发展, 互联网已经极大的改变了人类社会生活的方式, 但是与此同时, 也伴随产生了很多新的问题, 比如恶意攻击、垃圾邮件、计算机病毒等等, 这些问题阻碍了互联网的进一步发展. 学术界和产业界不断对这些问题进行深入研究, 并将其中比较重大的研究成果规划到下一代互联网中. 清华大学的林闯教授指出: 新一代互联网的特征可以用“可控、可信、可扩展”来概括, 同时, 可信的互联网络的特性包括: 安全性、真实性、可审计性、私密性、抗毁性和可控性[1]. 我们认为, 可信互联网的特点还可以概括为身份可信、传输可信、行为可信. 其中“身份可信”是其它特性(如可控性、可审计性、抗毁性等)的重要基础, 身份可信包括了服务提供者的身份可信和服务使用者的身份可信两个完全不同的领域. 我们在研究互联网的身份可信问题时, 主要研究的是服务提供者的身份可信, 因此如果没有特别说明, 本文提到的身份可信指的是服务提供者身份可信的问题.

实践证明, 互联网上的身份可信问题是一个比较复杂的问题, 直到现在, 钓鱼网站、域名劫持、假冒 IP 地址广播等等各种互联网的身份不可信问题仍然不时出现, 其中一些事件还会对社会正常运行造成一些严重后果, 比如在 2010 年 1 月 12 日, 中国著名搜索引擎百度的域名(baidu.com)就被劫持, 页面被换成黑色以及伊朗国旗, 给百度造成的直接经济损失高达 700 万人民币[2]. 在我们的研究过程中, 我们发现解决互联网的身份可信问题, 应该从 3 个层面来开展: 域间路由领域的身份可信、域名解析服务的身份可信、以及

①基金项目:中国下一代互联网示范工程基金(CNGI-12-02-007);国家发改委 2012 年信息安全专项基金(发改办高技[2013]1309 号);北京科委创新环境与服务体系建设基金(Z131101000113046);北京市怀柔区 2012 年度优秀人才培养基金(京怀组字[2012]46 号)
收稿时间:2014-02-21;收到修改稿时间:2014-03-27

互联网应用服务的身份可信, 也就是从 IP 路由层面、域名解析层面和互联网应用层面分别来解决互联网的身份可信问题, 三个层次互相支持, 可以构成一个比较完整的互联网身份可信体系.

目前在互联网社区中, 这三个层面都已经开始了身份可信的研究和实践, 并取得了很大的进展. 在域间路由领域, 已经发展出了 RPKI 并发表了 14 篇相关的 RFC; 在域名解析服务的身份可信领域的 DNSSEC 发展最为成熟, 在 2005 年就已经发布 RFC4033、RFC4034、RFC4035, 并在 2008 年发布 RFC5155, 目前已经进入大规模部署阶段, 截止 2014 年 1 月 31 日, 已经有 238 个 TLD(Top Level Domain)已经部署了 DNSSEC[3]; 在互联网应用层面解决服务提供者身份可信的问题是一个比较新的方向, 国内的产业界比较活跃, 应用最广泛的是基于域名技术的可信网站标识技术.

2 域间路由领域的身份可信问题

互联网上有时候会有人因为各种原因, 比如被黑客攻击或者错误配置了参数, 广播了不属于自己的地址, 这就是路由劫持. 路由劫持对互联网的正常运行危害非常大, 会导致大面积的网络瘫痪和服务不可用. 一些比较著名的路由劫持事件有：1997 年 4 月的“AS 7007 Incident” 事件, 2004 年的 12 月土耳其 TTNet 路由劫持事件, 2006 年 1 月 Con-Edison 路由劫持事件以及 2008 年巴基斯坦电信劫持 YouTube 流量事件等. 路由劫持事件的发生, 主要原因是因为 BGP 协议缺乏对路由内容正确性进行校验, 从而导致对路由信息的肆意修改. 目前互联网社区正在计划部署的 RPKI(Resource Public-Key Infrastructure)技术, 正是为了增强域间路由安全、防止路由劫持而提出的一项技术, 通过这项技术, 可以建立起一套基于 X.509 的域间信任体系, 从而杜绝各种非授权的地址广播行为. RPKI 是基于 X.509 证书体系, 保证互联网 IP 地址和 AS 号码等互联网基础号码资源正确性的基础设施, 通过对 X.509 证书的扩展以及若干签名对象, RPKI 实现了对互联网基础号码资源授权, 并以路由源声明(Route Origin Attestation, ROA)的形式进行验证.

RPKI 包括两种证书：CA 证书和终端实体(End Entity)证书. CA 证书是用来证明某个实体对互联网基础号码资源(IP 地址和 AS 号)的所有权; 终端实体证书是用来对 IP 地址前缀的路由源信息添加签名的. 传统上, IP 地址和 AS 号码的分配是有全球五大地址注册机构(Regional Internet Registry, RIR)负责的, 所以, 在 RPKI 体系中, IP 地址和 AS 号码的所有权是主要由五大地址注册机构认定的, 并通过对传统的 X.509 证书以关键扩展(Critical)的方式体现在 X.509 的 CA 证书中, RFC3779[4]就是该扩展的标准化文件, RFC3779 详细规定了 IP 地址和 AS 号作为扩展域在 X.509 证书中的编码格式. 另外, 当需要授权某个 AS 为其特定的 IP 地址前缀通告路由可达信息时, IP 地址的所有者使用其持有的 CA 证书签发一个 EE 证书来声明这种关系, 然后再用 EE 证书产生一个“路由源声明”签名项. 这样, 一个路由源声明则包含了一个 AS 号码和若干 IP 地址前缀的对应关系, 真实性是由 CA 证书、EE 证书的签名来保证的. 当 RPKI 需要验证时, RPKI 验证者(Relying Party, RP)通过验证 ROA 来验证一个路由通告中的 IP 地址资源、AS 号码资源以及两者的对应关系是否是持有者签署声明的.

同时, RPKI 相关支撑技术的标准化工作正在 IETF 域间路由安全(Secure Inter-domain Routing, SIDR)工作组持续开展, 2012 年 2 月, IETF 的 SIDR 工作组发布了 14 个和 RPKI 相关的 RFC 文档(RFC6480~RFC6493), 涵盖 RPKI 的体系结构、操作模型、密钥算法、证书策略、运行管理等支撑 RPKI 运行的各个关键环节, 标志着 RPKI 在技术层面已经基本成熟. RPKI 成为 ICANN(The Internet Corporation for Assigned Names and Numbers)下一个重点推广和部署的互联网基础设施. 伴随着 APNIC(亚太互联网信息中心)、RIPE NCC(欧洲互联网信息中心)等全球五大地址注册机构开始提供基于 RPKI 的资源认证服务, RPKI 及其衍生的资源认证服务将对互联网的运行产生重大影响.

当然, 实际工作中还有一些复杂问题需要进一步改进. 比如, RPKI 的认证关系由互联网码号资源的分配关系决定, APNIC、RIPE NCC 互联网注册机构通过签发证书作为 RPKI 的一般“信任锚点”(Trust Anchor, TA)对 RPKI 体系施加管理和控制, 但是有人认为“本地信任锚点”(Local Trust Anchor,LTA)也应该成为传统信任锚点的补充和加强, 具体实现方法仍然需要互联网社区范围中进一步讨论.

3 域名解析服务的身份可信问题

互联网域名系统 DNS(Domain Name System)是互联网服务的重要基础设施之一, 自从问世以来, 一直承担着互联网最基础的寻址职能, 同时也是互联网上最成功的协议之一. 不过, 由于当年 DNS 系统在设计之初, 互联网的应用场景远没有现在复杂, DNS 协议的设计者并没有考虑太多的安全问题, 但是现在, DNS 也遇到了越来越多的、安全方面的挑战. 特别是近年来, 和域名系统有关的互联网攻击技术不断提升, 也出现多次影响很大的攻击事件. 一般来说, DNS 服务系统的安全漏洞主要分为以下几类:

①域名劫持攻击(DNS hijacking)缓存中毒攻击;

②DNS 洪泛攻击(DNS Flooding);

③DNS 放大攻击(DNS Amplification Attack);

④DNS 动态更新缺陷攻击(DNS Dynamic Update Vulnerabilities);

其中, 域名劫持攻击就是域名解析服务的身份可信问题研究的范围. 域名劫持攻击是通过攻击域名解析服务器(DNS)、或伪造域名解析服务器、或修改用户 hosts 文件等方法, 把目标网站域名解析到错误的地址而达到使用户无法访问正确的目标网站的目的. 域名劫持攻击是一种假冒域名持有者真实身份的、普通用户难以识别的、危害很大的钓鱼手段(Phishing). 在 2008 年 7 月, 美国 IOActive 公司 Dan Kaminsky 公布了一个非常严重的 DNS 安全漏洞, DNS 缓存中毒攻击(DNS Cache Poison)[5]. 和以往的 DNS 劫持不同, DNS 缓存中毒攻击不需要获得 DNS 的管理权, 也不需要修改用户的 hosts 文件, 而是利用了递归 DNS 服务器(Recursive DNS Server)的缓存机制以及对查询应答包缺少有效检查的问题, 而展开的一种攻击行为, 由于 DNS 的缓存机制, 被攻击的 DNS 服务器可能会造成很大面积的破坏, 时间上也更加持久. DNS 缓存中毒攻击是近年来最新发展起来的域名劫持的攻击手段之一, 是域名劫持攻击中最有代表性的一种. 在 DNS 缓存中毒攻击过程中, 攻击者通过使用大量虚假的域名使 DNS 缓存中毒, 或者通过猜测 ID 序列号使 DNS 缓存中毒. 由于缓存中的错误数据是恶意构造的, 正常的 DNS 请求在收到应答后, 被指向一个错误的服务器, 可能会访问一个带有病毒的文件, 或者被钓鱼. 另外, 由于 DNS 的 forward 功能, 遭受缓存中毒攻击的 DNS 服务器还可能将错误的记录发送给其它的 DNS 服务器, 从而导致更大范围的破坏.

目前, DNSSEC(Domain Name System Security Extensions)技术是解决 DNS 领域身份可信问题的最主要手段. DNSSEC 是在传统的 DNS 协议基础上引入了公共密钥加密技术, 使 DNS 权威服务器的区管理员对其管理的区文件进行数字签名, 从而给递归服务器提供了一个检验手段, 以确保递归服务器查询得到的数据的正确性和完整性. 关于 DNSSEC 的细节描述, 在 RFC4033[6]、RFC4034[7]、RFC4035[8]中进行了详细规定,2008 年, IETF 又发布了一个 NSEC3(RFC5155[9])标准, 以提高 DNSSEC 隐私保护能力. DNSSEC 包括签名和验证两个过程,域名权威服务器用私钥对资源记录(Resource Record, RR)进行签名, 域名递归服务器用权威服务器的公钥对收到的应答信息进行验证. 如果验证没有通过, 则表明该应答报文在解析过程中可能被篡改了. 在 RFC4033 中规定, 一个验证解析器(validating resolver)应该可以验证分辨出以下四种结果:

(1) 安全的(secure): 验证解析器有一个信任锚点(trust anchor), 有一个信任链(a chain of trust), 并且在应答中能够验证所有的签名;

(2) 不安全的(insecure): 验证解析器有一个信任锚点, 也有一个信任链, 但是由于信任链不完整, 因而无法完成一个完整的验证, .

(3) 伪造的(Bogus): 验证解析器有一个信任锚点, 但是但是信任链上的签名验证是错的. 导致错误的原因可能是因为受到攻击, 也可能签名已经过期或者是签名使用了不支持的算法等等.

(4) 不确定(Indeterminate): 没有信任锚点, 验证过程无法开始.

在完全部署 DNSSEC 后, 可以确保最终用户在访问互联网时, 链接到特定域名所对应网站或服务上. 当然, 这并不会解决互联网安全的所有问题, 但这一环节是相当重要的一个环节. 因此, 推动 DNSSEC 的完全部署是一项非常重要但是难度很大的工作. ICANN 正在全球范围内, 特别是顶级域名(TLD, Top Level Domain)领域内, 推动这项工作. 目前, 根据 ICANN 提供的信息, DNSSEC 的部署范围已经超过了 50%[10]. 可以预见, 在不远的将来, DNSSEC 将成为一项普遍部署的互联网安全的基础服务.

4　互联网应用服务的身份可信问题

在互联网上，有很多钓鱼网站(Phishing Website)并不是通过 AS 号码劫持技术，也不是通过域名劫持技术来进行欺诈行为，也就是说网域广播和 DNS 解析都是正确的，而这些不良分子是通过和被仿冒者相似的域名或者相似的页面也假冒网站的. 互联网上最终用户在确认一个域名的持有者的身份时，主要的方式是通过各个顶级域名(Top Level Domain, TLD)提供的 Whois 服务进行查询，但是由于一方面 whois 服务过于专业化，大部分普通用户并不熟悉，另一方面，域名注册时很多域名持有者的身份并没有被严格核验过，最终用户在核验域名持有者的身份时十分不便，因此，给不良分子制作假冒网站留下了可乘之机. 由此可见，高可信性是下一代互联网一个非常重要的特征，互联网上的所有计算机系统需要建立高可信的网络服务，如何构建一个安全、可生存和可控的可信互联网正在成为人们关注的焦点.

北龙中网(北京)科技有限责任公司(以下简称中网公司)在国家发改委信息安全专项的支持下，设计的“网站可信标识体系”，是在可信应用方面，做出的重要尝试，网站可信标识体系结构包括 3 个角色，分别是网站、可信应用以及标识权威机构[11]. 其中，网站指的是网络服务提供者，指待认证的实体；可信应用指的是支持网站可信标识的应用，包括浏览器、搜索引擎、即时通讯软件等. 可信应用可以对具有可信标识的网站进行验证，并向最终用户展示标识的信息；标识权威机构指的是具备认证网站真实信息能力，能够发布网站可信标识的机构. 标识权威机构对网站进行验证，为网站发放可信标识，同时对可信标识进行管理，并提供其验证的可信标识详细信息查询和网站详细验证信息查询.

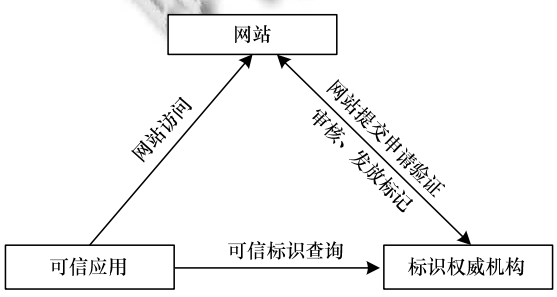

图 1　网站可信标识体系结构

网站可信标识体系执行的具体过程如下：

①网站所有者向标识权威机构提交资料信息注册，申请标识.

②标识权威机构对申请信息审核后，将被验证网站的标识数据按指定格式生成网站可信标识数据，发布到查验服务平台上.

③可信应用需要获取网站的可信标识信息时，应通过身份标识查验协议访问标识权威机构的查验服务，根据返回结果获得网站的可信标识信息.

④可信应用对标识验证通过后，在应用上展示该网站的可信标识，并将标识中的内容向用户展示，提示用户正在访问的网站信息.

⑤用户可通过可信应用上展示的可信标识跳转至验证该网站的标识权威机构网站查看完整的验证信息.

由以上流程可以看出，可信应用向标识权威机构发送的可信标识查询请求这个过程是整个体系的核心，这个查询过程使用的是一个基于 DNS 查询协议的服务接口，可查询某一网站是否经过标识权威机构验证，和相应的验证信息. 针对每个网站的验证信息的集合，称为网站可信标识对象. 网站可信标识对象由标识权威机构生成，并发布到该标识权威机构的身份标识查验服务上. 标识权威机构通过验证信息的集合来生成网站可信标识对象. 信息集合包括可辨别的网站域名、IP 地址，用户名称以及一个可选的包含用户附加信息的唯一性标识符. 唯一性标识符内容的确切格式未做规定，而留给标识权威机构(IA)去定义. 唯一性标识符可以是诸如对象标识符、日期或是说明有关可辨别用户名的有效性的证书的其他形式. 协议中的标识对象是不可伪造的，可使用数字证书签名技术或 DNSSEC 技术，保证标识对象的真实性.

中网公司已经将本体系和相关协议实现为“可信网站验证开放平台”，这个平台得到了产业界的广泛认可，各种主流互联网应用正逐步接入本开放平台，下面两张图展示的是淘宝浏览器和阿里云搜索引擎对接开发平台后的使用效果图.

图 2　可信网站验证服务在淘宝浏览器的应用

图 3 可信网站验证服务在阿里云搜索引擎中的应用

同时, 中网公司正在全国信息安全标准化技术委员会和中国通讯标准化协会的组织领导下, 将该体系和相关标准制定为国家标准和行业标准.

5 结语

针对互联网中的服务提供者身份可信问题, 本文从网域路由、域名解析和互联网应用服务 3 个层面讨论了互联网中服务提供者身份可信问题的理论研究的成果和实践, 包括解决网域路由的身份可信问题的 RPKI 体系的原理和最新进展、反域名劫持的 DNSSEC 技术的主要原理和最新进展、以及网站可信标识体系的主要原理和最新进展.

可以说, 高可信性是互联网下一步发展的重要方向, 而目前互联网在高可信性方面的发展, 也正是沿着以上三个方向展开, 深刻认识并积极参与可信互联网的建设, 对互联网的良性发展具有不可估量的正面作用. RPKI 和 DNSSEC 技术都是由国外互联网社区主导推进的, 而网站可信标识体系是由我国互联网社区发明并持续推动的, 具有重大意义.

参考文献

1 林闯,田立勤,王元卓.可信网络中用户行为可信的研究.计算机研究与发展,2008,45(12):2033–2043.

2 2010 年百度域名被劫持事件.维基百科.http://zh.wikipedia.org/wiki/2010 年百度域名被劫持事件

3 Internet Corporation for Assigned Names and Numbers. TLD DNSSEC Report. http://stats.research.icann.org/dns/tld_report/archive/20140131.000101.html

4 Laurie B, Sisson G, Arends R. X.509 Extensions for IP addresses and AS identifiers. RFC3779, 2008.

5 Wikipedia. Dan Kaminsky. http://en.wikipedia.org/wiki/Dan_Kaminsky

6 Arends R, Austein R, Larson M, Massey D, Rose S. DNS security introduction and requirements. RFC4033, 2005.

7 Arends R, Austein R, Larson M, Massey D, Rose S. Resource records for the DNS security extensions. RFC4034, 2005.

8 Arends R, Austein R, Larson M, Massey D, Rose S. Protocol modifications for the DNS security extensions. RFC4035, 2005.

9 Laurie B, Sisson G, Arends R, Blacka D. DNS security (DNSSEC) hashed authenticated denial of existence. RFC5155, 2008.

10 Internet corporation for assigned names and numbers. DNSSEC Surpasses 50%. http://blog.icann.org/2014/01/dnssec-surpasses-50/

11 卢文哲,杨风雷,高宁,毛伟.网站可信标识架构与查验协议研究.计算机工程,2014,40(1):20–24.