云计算在信息安全中的应用

邹 维 丁 羽 韩心慧 杨文瀚
北京大学

关键词：信息安全 病毒查杀 网页木马 漏洞挖掘

引言

互联网时代的信息安全面临着高强度、高时效的技术对抗，零日（zero-day）攻击、僵尸网络以及用户隐私信息大面积泄露等都是新的挑战。然而，云计算的迅速发展，除了给信息安全带来挑战外，也为信息安全提供了有效的技术支撑。

云计算既是一种技术，也是一种商业模式，其特点是将大规模计算及存储资源在基础设施即服务（infrastructure as a service，IaaS）、平台即服务（platform as a service，PaaS）、软件即服务（software as a service，SaaS）三个层面为用户提供不同类型的、弹性化的服务，并在云端对其进行管理。云计算已经在信息搜索、邮件服务、客户关系管理、信息存储等领域取得了巨大的成功。

目前许多信息安全活动都以服务形式呈现。基于云计算模式的信息安全服务，称作安全即服务（security as a service，SecaaS）。一般认为它是云计算中软件即服务的一个子类。据高德纳（Gartner）预测，2011年至2013年全球SecaaS业务将增长三倍。云安全联盟（Cloud Security Alliance, CSA）的SecaaS工作组于2011年发布的报告[1]，列举了10大类SecaaS服务：

1 身份与访问管理（identity and access management，IAM）

2 数据防丢失（data loss prevention，DLP）

3 Web安全（Web security）

4 电子邮件安全（Email security）

5 安全评估（security assessments）

6 入侵管理（intrusion management）

7 安全信息与事件管理（security information and event management，SIEM）

8 加密（encryption）

9 业务连续性与灾难恢复（business continuity and disaster recovery）

10 网络安全（network security）

服务基本涵盖了直接面向用户的各类安全检测和防护的内容，只是未提及软件安全漏洞挖掘。本文将介绍云计算在信息安全领域的三类具体应用：

基于云计算的病毒查杀 属于第10类服务，其核心是云和客户端的协同问题；

基于云计算的网页木马检测 属于第3类服务，其核心是海量数据安全性的分析问题；

软件漏洞并行挖掘 不属于以上10类服务，是第5类服务的重要支撑，复杂计算的并行化执行是其面临的主要挑战。

基于云计算的病毒查杀

密歇根大学乔恩·奥博赫德（Jon Oberheide）等人在2008年USENIX Security Symposium会议上发表的题为“CloudAV: N-Version Antivirus in the Network Cloud”论文中，首次提出基于云计算的病毒查杀思想[2]。该技术借用云服务平台提供的强大灵活的计算能力，将病毒扫描分析这一计算与资源密集型任务转移到云平台中。通过对云计算平台的计

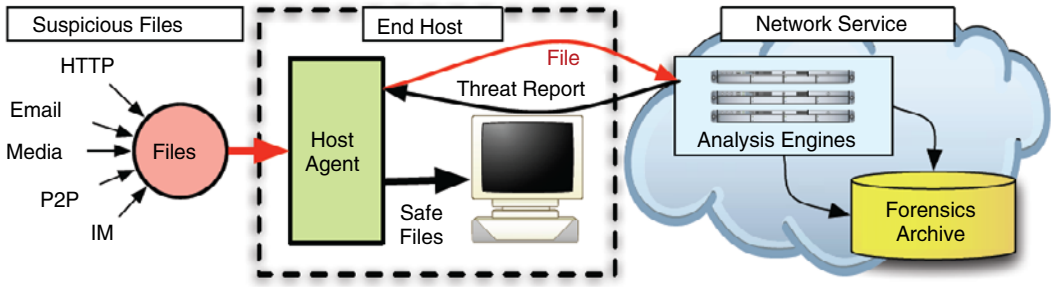

图1 CloudAV架构图[2]

算与存储资源的充分利用，CloudAV不但降低了客户端防病毒软件的负担，同时提高了对未知、多态病毒的平均识别率。

在CloudAV杀毒模式中（图1），用户终端的杀毒程序自动将敏感文件上传到云端进行分析，云端借助大型病毒特征库，通过分析引擎对汇集的成千上万份敏感文件样本进行并行化检测，将处理结果返回给客户端杀毒程序，从而减轻了用户终端负载，增强了查杀效果，也改善了用户体验。

相对于传统的杀毒软件，CloudAV具有以下优势：

1 在云端易于部署基于虚拟机沙箱的恶意行为动态分析工具，以及时发现未知病毒；

2 借助云计算中MapReduce等技术对病毒样本做并行化静态分析，大大提升分析效率；

3 通过在云端部署多种来自不同厂商的病毒检测引擎，对病毒样本进行交叉检测，可提升检测准确率；

4 在云端可维护大型甚至巨型病毒特征库，一旦分析引擎发现了来自某个客户端的一种新型病毒，则会迅速入库，让成千上万的客户端共享最新、最全的病毒信息。

基于云计算的病毒查杀技术（简称“云杀毒”）的核心思想是“云查—端杀”，即利用云端强大的计算能力，进行病毒分析和检测，客户端跟据结果进行病毒杀灭。目前主流的防病毒软件厂商都不同程度实现了云杀毒，其中包括赛门铁克、奇虎、卡巴斯基、瑞星等公司。虽然各厂商所采用的技术框架及原理基本相同，但在具体实现上各有特色。

赛门铁克公司将客户端的文件标为三个不同的信任等级，并根据用户所选择的信任等级确定扫描范围；奇虎公司在云端建立了大型黑白名单数据库，以提升云端运行效率；瑞星公司在云端部署了木马/恶意软件自动分析系统来进行分析；卡巴斯基公司的客户端杀毒程序在发现疑似样本时，会自动将在该公司检测网中首次出现的时间、使用人数、已信任该样本的用户比例等信息提供给用户进行判断，系统根据判断结果进行相应处理。云端杀毒由于需要从客户端外传用户文件等信息，一些用户往往会担心个人信息泄露，卡巴斯基公司的做法减小了用户的担忧，但需要用户具备基本的防病毒知识和更多的人机交互。

云杀毒技术经过4年的发展，已全面进入实用阶段。这是云计算在信息安全领域一项最成功的应用。随着4G宽带无线通信技术的普及应用，移动互联网也将迎来云杀毒时代。

基于云计算的网页木马检测

网页木马是近年来出现的一种新形态恶意代码，它通常被人为植入Web服务器端的HTML页面中，目的在于向客户端传播恶意程序。当用户通过客户端浏览器访问该页面时，它会利用浏览器及其插件的漏洞将恶意程序自动植入客户端电脑中。网页木马的表现形式是一个或一组有链接关系、含有恶意代码（通常用JavaScript等脚本语言编写）的HTML页面，恶意代码在该（组）页面被客户端浏

览器加载、渲染的过程中被执行，并利用浏览器及插件中的漏洞隐蔽地下载、安装、执行病毒或间谍软件等恶意可执行代码。与主动传播的网络蠕虫不同，网页木马采取“守株待兔”的被动传播方式，当用户浏览页面时，网页木马以网页内容的正常下载为掩护，隐蔽地完成感染和入侵，因此被称为“Drive-by-Download”。

谷歌的一项研究表明其搜索结果中平均约1.3%的网页被感染木马[4]，网页木马已经成为网络攻击与病毒传播的重要手段。

网页木马的检测一般分为高交互式和低交互式两类：（1）高交互式检测通常基于虚拟技术模拟一种浏览器环境，通过透视度量、控制流完整性检测等技术对网页的动态行为进行检测；（2）低交互式检测主要包括基于已知漏洞特征匹配的检测方法、基于机器学习的方法和基于统计的方法。

网页木马的检测非常适合在云计算环境中进行，因为分析一个单独网页安全性的时间代价并不大，而且网页数量众多，借助爬虫技术可以爬取到足够多的网页进行分析。因此使用云计算来进行大规模网页并行分析是十分有效的。

谷歌的安全浏览（safe browsing）服务可以为使用搜索引擎的用户提供URL的安全性警示，包括钓鱼网站（phishing）和网页木马（malware page）两项安全性检测。提到谷歌的网页木马检测，人们往往会想到Stopbadware[5]。它是一个与恶意代码（或“坏件”，badware）作斗争的非盈利组织，源于哈佛大学的Stopbadware.org项目。该组织定期公布一个全球恶意网站黑名单。有人误认为谷歌是通过这一黑名单比对来确定搜索结果中是否包含恶意网页。其实不然，谷歌采用的是尼尔斯·普洛斯（Niels Provos）等人研制的自动检测系统[3]，该系统具有以下3个特点：

1 采用基于MapReduce模型的启发式算法进行网页裁剪，大幅度缩减检测页面的数量；

2 通过虚拟机上的IE浏览器沙箱（Sandbox）技术，对网页进行动态行为分析；

3 结合多种来自不同厂商的反病毒引擎扫描结果，进行动态验证。

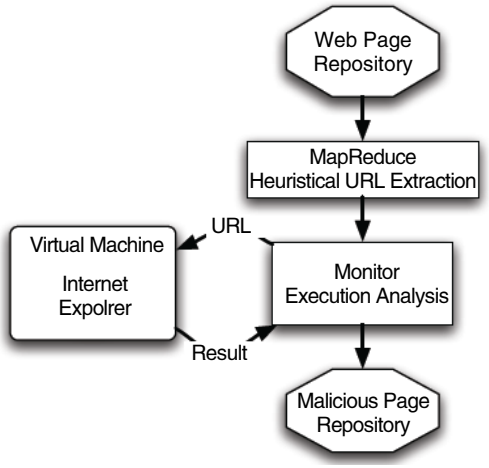

图2 谷歌网页木马检测[3]

图2显示了基于以上关键技术形成的检测方法。

网页木马本质上是针对客户端万维网（Web）浏览器安全漏洞的渗透攻击代码，通过破坏客户端软件运行时刻的完整性，获得远程代码执行机会并进一步植入恶意代码。因此，对网页木马的检测可归结为动态完整性遭到破坏与否的检测问题。

奇虎公司的360云安全系统的网页安全性检测采用了机器学习方法，每天处理5TB以上的数据。北京大学互联网安全技术北京市重点实验室研发的Web网页木马监测系统[6]，采用海量网页过滤、轻量级并行化沙箱、挂马链分析、网页木马场景收集与重放等技术，形成一个基于云计算架构的Web网页在线安全服务平台，为客户网站提供持续的Web浏览安全监测服务。近三年来协助中国教育和科研计算机网（CERNET）紧急响应组（CCERT）为网内3.5万个网站提供持续的网页木马检测服务。

软件漏洞并行挖掘

软件安全漏洞（security vulnerability）是指可以被攻击者用于破坏目标系统安全策略的软件错误。安全漏洞是网络攻防的重要信息资源，信息技术大国特别是美国已形成一条围绕安全漏洞发现、评估、交易、利用的产业链[7]。

漏洞的发现包括两个方面：（1）漏洞挖掘：通过对软件程序（包括二进制执行程序）的静态或动态分析，发现安全漏洞的过程；（2）漏洞逆向分析：通过对软件补丁或恶意代码的逆向分析，发现补丁试图弥补、防护的软件漏洞，或者发现恶意代码试图攻击的软件漏洞。

经过几十年的发展，漏洞挖掘逐渐形成静态挖掘和动态挖掘两个主要技术方向。各种软件漏洞挖掘技术都会不可避免的遇到路径爆炸问题。为了缓解路径爆炸，提高软件漏洞分析效率，研究者开始将目光转向并行漏洞挖掘方向。云计算服务能够提供大量弹性的计算资源，为并行化软件漏洞挖掘创造了条件。比较杰出的几项并行漏洞挖掘工作：

斯坦福大学Saturn项目[8] 谢依晨（Yichen Xie）等人针对可被有限状态机（DFA）刻画的漏洞模式，构造了并行漏洞挖掘系统Saturn。该系统扩展了布尔抽象理论，将程序整型变量、路径谓词抽象为布尔（Bool）矢量和谓词，强化对指针和结构体的分析，在遍历程序路径时，将漏洞触发条件编码为Bool谓词，通过判定Bool谓词的可满足性从而判断漏洞是否存在。Saturn系统充分利用函数依赖关系，结合函数摘要技术，并行分析彼此不依赖的函数。这项并行技术取得了出色的实验结果，使对Linux内核锁异常的分析时间从原来的23小时缩减至50分钟。

加州大学伯克利分校的Vulnivore项目[9] 瓦格纳（D. Wagner）等人研发的Vulnivore系统关注于格式化字符串（format string）漏洞，将此类漏洞的挖掘并行化。在一个月的时间内，完成对Debian 3.1 Linux中3692个代码包分析，成功发现数千个安全隐患。

Sun公司的Parfait系统[10] 该系统使用并行、迭代化模块和启发式策略进行漏洞挖掘。Parfait 系统以具体需求为驱动（demand driven），采用“瀑布”模型，整合不同的漏洞挖掘模块和启发式策略，进行并行化、迭代式漏洞挖掘。

瑞士Cloud9系统[11] 赛尔蒂（L. Ciortea）等人在Amazon EC2云计算平台上实现了Cloud9系统。它能够在符号化遍历程序执行空间过程中，并行遍历不同的执行路径，在不同的计算节点间调度遍历任务，并保持负载均衡。基于云计算的并行符号执行与负载均衡技术是Cloud9的两项创新。其结构见图3。

北京大学互联网安全技术北京市重点实验室基于云计算的并行漏洞挖掘 实验室在面向安全的反编译、基于污点跟踪与符合执行的漏洞挖掘等方法研究方面有一定积累，已发现20多个零日安全漏洞（指首次被发现、尚无补丁的漏洞）[12,13]。实验室软件安全课题组基于G. Edward Suh安全脆弱性模型，尝试从源代码中找出可能的软件漏洞脆弱点，结合动态污点分析和二元决策图（binary decision diagram）理论，使用自底向上的方法计算各个潜在脆弱点的可达性。该方法具有良好的可伸缩、并行化能力，因此适用于云计算环境。该课题解决两个关键性技术难题：（1）脆弱性Sink（指程序中出现缓冲区溢出等异常的位置）的筛选问题，如何从大型复杂软件中挑选出潜在的脆弱性Sink。（2）在已知潜在脆弱性Sink的情况下，采用哪种并行、高效方法，以辨别出Sink是否是真实的漏洞。

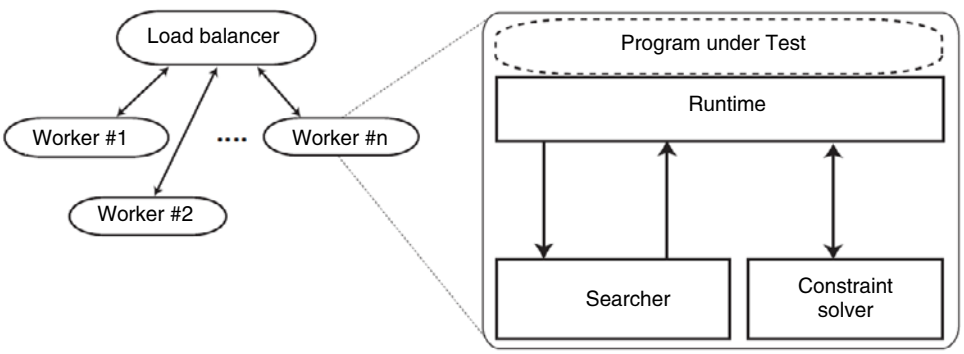

图3 Cloud9结构图

现有的研究结果显示，并行技术能够显著提高漏洞挖掘技术的效率，

是未来漏洞挖掘工作的一个重要研究方向。需要指出的是，由于程序结构的复杂性，其运行状态及可执行路径数量随程序代码行数迅速增长，同时许多程序分析方法理论性强、实现复杂度高，因此漏洞挖掘的并行化实现难度远高于前面介绍的“云查杀”和基于云计算的网页木马检测。正因如此，国内外在基于云计算的软件漏洞挖掘研究方面仍处于探索阶段，许多理论研究工作有待实际验证。

结语

本文介绍了云计算在病毒查杀、网页木马检测、软件漏洞挖掘方面的典型应用。前两类应用技术相对成熟，已进入实用阶段，第三类尚处于研究阶段，有待进一步探索。可以预见，云计算作为继个人电脑、互联网之后的第三次信息革命产物，在信息安全领域的应用前景将十分广阔。■

致谢：本文得到国家发改委国家信息安全专项2010课题、国家242信息安全计划项目（2011A40）的资助。

邹　维

CCF高级会员。北京大学研究员。主要研究方向为互联网安全检测、软件安全、新形态网络安全等。

zou_wei@pku.edu.cn

丁　羽

CCF学生会员。北京大学博士生。主要研究方向为软件安全。

dingelish@pku.edu.cn

韩心慧

CCF会员。北京大学高级工程师。主要研究方向为恶意代码检测与防范、互联网安全威胁监测等。

hanxinhui@pku.edu.cn

杨文瀚

北京大学博士生。主要研究方向为新形态网络安全。

yangwenhan@pku.edu.cn

参考文献

[1] CSA SecaaS, Defined Categories of Service, 2011, https://cloudsecurityalliance.org/research/working-groups/security-as-a-service

[2] J.Oberheide, E.Cooke, F.Jahanian. CloudAV: N-Version Antivirus in the Network Cloud, USENIX Security Symposium, 2008

[3] N.Provos, D.McNamee, P.Mavrommatis, K.Wang, N.Modadugu. The ghost in the browser analysis of web-based malware, In Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets (HotBots'07), 2007

[4] N.Provos, P.Mavrommatis, M.A.Rajab, F.Monrose. All your iFRAMEs point to Us, Proceedings of the 17th Usenix Conference on Security symposium (SS'08), 2008

[5] StopBadware: http://www.stopbadware.org/

[6] 北京大学互联网安全技术北京市重点实验室：http://seclab.pku.edu.cn

[7] 韦韬, 王贵驷, 邹维. 软件漏洞产业：现状与发展, 中国信息安全, 2010年第1期, 2010

CCF YOCSEF CLUB 6月在光明网举办

2012年6月14日晚，CCF YOCSEF在光明网举行CLUB活动。CCF YOCSEF主席袁晓如、副主席王涛，CCF副秘书长刘雨以及AC委员宋乐永、荣誉委员侯紫峰、委员刘剑、彭柯、秦征、肖永红、梁荣华、周黎、陈涛和华为的温向东等参加了CLUB。会议由宋乐永主持。

会上，安卓越公司COO方生与大家分享了当前移动互联网产业发展的新趋势、新技术以及对移动互联网人才的需求等。大家围绕移动互联网的软件基础环境、移动开发人才需求等进行了深入讨论。与会人员还参观了光明日报社。

[8] Y.Xie，A.Aiken. Saturn: A Scalable Framework for Error Detection using Boolean Satisfiability. ACM Transactions on Programming Languages and Systems (TOPLAS), Vol.29(3), May 2007

[9] K.Chen, D.Wagner. Large-Scale Analysis of Format String Vulnerabilities in Debian Linux, Programming Languages and Analysis for Security(PLAS’07), 2007

[10]C.Cifuentes, B.Scholz. Parfait –Designing a Scalable Bug Checker, Proceedings of the 2008 workshop on Static analysis (SAW’08), 2008

[11]L. Ciortea, C. Zamfir, S. Bucur, V. Chipounov, G. Candea. Cloud9: A Software Testing Service, ACM SIGOPS Operating Systems Review, Vol. 43 Issue 4, Jan 2010, P.5-10.

[12]Tielei Wang, Tao Wei, Guofei Gu, Wei Zou: TaintScope: A Checksum-Aware Directed Fuzzing Tool for Automatic Software Vulnerability Detection. IEEE Symposium on Security and Privacy 2010: 497-512.

[13]Tielei Wang, Tao Wei, Guofei Gu, Wei Zou: Checksum-Aware Fuzzing Combined with Dynamic Taint Analysis and Symbolic Execution. ACM Trans. Inf. Syst. Secur. 14(2): 15 (2011)