

基于小云审大云的云平台可信评测体系结构与技术研究

刘川意¹, 潘鹤中^{1,2}, 梁露露³, 王国峰^{1,2}, 方滨兴^{1,2,4}

(1. 哈尔滨工业大学(深圳)计算机科学与技术学院, 广东 深圳 518055; 2. 北京邮电大学计算机学院, 北京 100876;
3. 中国信息安全测评中心, 北京 100085; 4. 东莞电子科技大学电子信息工程研究院, 广东 东莞 523000)

摘要: 提出一种“小云审大云”的云平台可信评测架构, 通过引入独立第三方对云平台的可信性进行动态、实时的远程可信数据收集、验证、审计和评价。采用数据流可视化、监控与脱敏技术保障租户数据权益, 为租户提供选择可信云平台的依据, 为云平台提供证明自身可信的凭据。并突破以下关键技术: 1) 基于独立第三方的数据采集、行为接入、API代理; 2) 持续、实时的远程数据收集和数据分析; 3) 非强侵入式证据收集。

关键词: 云计算; 可信评测; 权限管控; 数据脱敏

中图分类号: TP311

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2016.00102

Cloud Trustworthiness Evaluation as a cloud service: architecture, key technologies and implementations

LIU Chuan-yi¹, PAN He-zhong^{1,2}, LIANG Lu-lu³, WANG Guo-feng^{1,2}, FANG Bin-xing^{1,2,4}

(1. School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China;

2. School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China;

3. China Information Technology Security Evaluation Center, Beijing 100085, China;

4. Dongguan University of Electronic Science and Technology Electronic and Information Engineering Institute, Dongguan 523000, China)

Abstract: A "big clouds audited by a small cloud" scheme was proposed, by introducing an independent trusted third party (TTP) dealing with run-time data collection, verification, audit and evaluation remotely, in a continuous and data-driven model, compared with traditionally certification based audit. The TTP mainly adopts data flow visualization, data monitoring and encryption to protect the rights of users. It provides the basis for users to choose a trusted cloud platform and for cloud platform to prove own trusted credentials. In-depth study, the following key technologies were broken through: 1) the introduction of an independent trusted third party as an intermediate layer between cloud platform and users as well as administrators; 2) continuous, real-time remote data collection and data analysis; 3) strong non-intrusive evidence gathering.

Key words: cloud computing, trusted reviews, permissions control, data encryption

1 引言

在云计算模式下, 租户将自己的数据和业务

系统迁移到云服务商的云计算平台上, 失去了对这些数据和业务系统的直接控制能力。另一方面, 从体系结构层次来看, 租户处于上层, 并不能有

收稿日期: 2016-05-19; 修回日期: 2016-07-08。通信作者: 刘川意, cy-liu04@mails.tsinghua.edu.cn

基金项目: 国家高技术研究发展计划(“863”计划)基金资助项目(No.2015AA016001); 广东省产学研合作基金资助项目(No.2016B090921001); 山东省自主创新及成果转化专项基金资助项目(No.2014ZZCX03411)

Foundation Items: The National High Technology Research Program of China (863 Program) (No.2015AA016001), Production-Study-Research Cooperation Project in Guangdong Province (No.2016B090921001), Innovation Project in Shandong Province (No.2014ZZCX03411)

效获知和监控云提供商的底层基础设施和云操作系统。事实上，云管理员客观上具备窃取用户隐私数据的特权^[1, 2]。

本文实现了一个真实攻击场景（适用于各主流虚拟化平台，如 OpenStack^[3]、QEMU+KVM^[4, 5]、Xen^[6]、vSphere^[7]、Hyper-V^[8]等）；租户在云上的 Linux 虚拟机，其 Root 密码是用户自己设定的，以防止别人入侵。但云管理员可直接对 Linux 虚拟机磁盘文件进行修改，从而绕过虚拟机登录验证，掌握虚拟机的 Root 权限，如图 1 所示，其主要步骤如下。

1) 云管理员登录云平台后，可查看虚拟机的详细信息，其中包括作为虚拟机唯一标识的云主机 ID。

2) 云管理员登录运行该虚拟机的物理机，进入虚拟机实例的存储位置，在记录下虚拟机 ID 号后，云管理员便可找到其对应的云主机实例，并将其复制挂载到另一个服务器上。

3) 挂载成功后，云管理员可以找到挂载文件的用户密码（shadow）配置文件，进一步找到用户设置的虚拟机口令。编辑此文件，删除对应的登录用户名和密码，绕过用户云主机的登录验证环节。

4) 删除成功后，当云管理员使用该挂载文件重新启动这台虚拟机时，便不再需要口令就可以直接进入该虚拟机。进入后，云管理员便可以查看、修改甚至恶意删除虚拟机的任何数据。

云平台或云管理员是否存在对租户数据或权益的侵害行为，称为云平台的可信问题。

通过对主流云平台体系结构、权限管理、日志

和审计等进行系统梳理和分析，总结出影响云平台可信性的安全风险主要有 3 个方面（如图 2 所示）。

1) 云平台体系结构栈中存在安全风险

通常情况下，租户访问自己的虚拟机 VM（virtual machine）的流程如下。首先通过网络进入云平台网关节点，通过云平台管理系统定位到虚拟机 VM 所在的物理服务器，然后经由虚拟机管理器（VMM，virtual machine monitor）和虚拟交换机（virtual switch）将请求和数据发送给 VM 或截获来自 VM 的事件和数据，最终 VM 中的操作系统和业务应用程序将对租户的数据进行处理。在上述整个流程中，若任意一个组件存在安全风险，用户的隐私则存在被侵犯的可能，即存在不可信的因素。云平台体系结构栈中的安全风险，除了主机安全、网络安全、应用安全这 3 种传统的安全点以外，存在云平台特有的安全风险。例如，由于虚拟化平台自身存在安全漏洞导致攻击者对客户主机的非法侵入（图 2 标号 1），由于虚拟化平台中多租户共享隔离机制缺陷导致某租户以自身虚拟机为跳板对其他虚拟主机的非法侵入（图 2 标号 2）等。

2) 云平台管理权限划分存在安全风险

云系统中租户对虚拟机和数据的控制权缺失，并且对云平台提供商的信任有限。云管理员有可能偷窥、泄露，甚至篡改用户信息。数据中心基础设施、云平台管理系统、VMM 等都会给云管理员以相应的管理接口对云平台的物理服务器和虚拟化组件进行管理，通过这些管理接口，云管理员往往可以偷窥、泄露，甚至篡改用户信



图 1 一个真实的攻击场景

operations center) 系统模型, 主要用来关联并分析来自不同信息源的安全事件。

在学术研究上, Bhatt^[9]等指出, SOC 是一个安全运营中心, 用来监控企业 IT 资产 (包括 IT 网络、外围防护系统, 如防火墙和入侵检测系统、应用程序服务器、数据库和用户账户等) 安全相关的事件。SOC 主要包括 SIEM 工具和 SA (security analyst) 分析员。SIEM 主要通过传感设备和事件日志从不同来源收集事件, 把不同事件记录模式处理成标准化表示, 存储结果并利用规则库分析安全相关的事件。通过分析事件信息发现网络和企业周边可能出现的恶意行为时触发警报, 由 SA 分析判断警报属于恶意事件还是正常行为。Bhatt 等^[9]总结出 SOC/SIEM 主要存在如下主要缺点及参考思路。1) SOC/SIEM 成本较高, 主要用来处理大量不同来源的事件, 不适用于小规模网络。2) 由于数据量大, 包含大量无用数据或噪音数据, 需要解决存储、查找、分析、可视化等问题。3) 由于事件种类繁多, 规则不好确定, 若规则太具体, 容易产生漏报; 若规则太抽象, 则容易产生大量误报。另外规则库需要及时学习和更新。4) 与企业网络运营是隔离的, 一些正常的操作, 如配置、测试也有可能产生报警, 带来不必要的开销。且如果事件源被入侵, SIEM 则有可能收集到被恶意篡改的事件。Yen 等^[10]依据 SIEM 技术提出一种基于日志分析的网络异常行为框架 Beehive, 主要通过大范围收集和分析 HTTP 代理服务器日志, 从日志中提取特征, 然后使用聚类发现和确定主机异常行为, 但需要大量的人工分析和评定工作。针对 SIEM 系统处理并关联大量低质量的事件产生的误报问题, 有人提出利用警报排序算法识别 SA 需要注意的高质量的警报, 如利用概率理论或证据理论^[11,12]实现警报排序方案。

目前, 业界已有相关的 SIEM 产品。根据 2016 年 Gartner 发布的 SIEM 市场分析报告^[13], 攻击检测与响应及合规管理是 SIEM 市场发展的驱动力。IBM QRadar^[14]、Splunk^[15]和 LogRhythm^[16]是 SIEM 供应的重要厂商, 其中, Splunk 能提供 SaaS 服务模式, 可以部署在公有或私有云中, 但其使用成本较高, 对使用者的要求较高。一些开

源项目, 如 ELK^[17]、OpenSOC^[18]有进一步发展空间, 但对现有市场的影响较小。

2.2 云平台权限划分和验证

云平台体系结构栈中存在安全风险, 云平台中的恶意管理员或入侵攻击者都可能对用户隐私造成威胁。对此, 相关研究针对云平台资源或数据作权限划分, 或在特权层制定数据访问验证策略, 以提高云平台的可信性, 使特权管理员无法轻易窃取用户隐私。

自服务云计算体系架构 (SSC)^[19]修改了虚拟机管理的权限模型, 将原有管理域划分为无法管理用户虚拟机的系统管理域和管理用户虚拟机的用户管理域, 以防止恶意的云管理员窃取用户虚拟机数据, 并且为用户使用虚拟机提供服务接口。这种方法依赖云平台物理芯片, 且需要修改 VMM 和用户 VM 镜像, 通用性受到限制。

CloudVisor^[20]在云平台最高特权层, 将虚拟机管理器置于非特权层, 使用硬件辅助技术实施 VMM 与虚拟机之间的内存隔离。CloudVisor 仲裁 VMM 和虚拟机之间的数据交互, 使 VMM 访问不到虚拟机的隐私数据。传统的虚拟化功能, 如资源管理、VM 创建和销毁、调度等仍然由 VMM 完成。所有从客户虚拟机的控制转移首先由 CloudVisor 处理, 由 CloudVisor 对相关数据做保护, 然后转移到 VMM。同样 CloudVisor 也需要借助物理芯片, 其改变了云平台权限逻辑, 带来一定的性能损失。

Overshadow^[21]和 Inktag^[22]研究在虚拟机不可信的情况下保护用户程序的安全性, 通过 VMM 和用户程序的协商保护用户程序的隐私。Overshadow 利用虚拟化阴影页表技术, 提供两套阴影页表, 包括系统页表 and 用户程序页表。当系统要访问用户数据时使用系统页表, 访问到加密的用户内存数据; 当用户程序访问数据时使用用户阴影页表, 访问到明文内存视图, 防止用户数据被恶意虚拟机用户窃取。Overshadow 使用单密钥加密页面, 由 VMM 管理, 密钥一旦暴露则所有数据都会泄露; 需要借助物理芯片, 控制转移需要页表切换和加解密操作, 使执行性能降低。

Inktag 提出半验证机制, 利用不可信的操作系统帮助验证自己的行为。在修改一个进程的页

表之前,操作系统必须首先执行超级调用使页表更新与高级应用程序的请求相关联,并且不可信的操作系统通知 VMM 页表更新,显著减少管理程序的复杂性。Inktag 基于半虚拟化技术,需要借助物理芯片,并且需要修改用户程序使其与 VMM 程序配合,通用性大大降低。

2.3 数据监控

用户对于云提供商是否能够对其数据提供隐私保障,云应用程序是否按照预定的方式安全执行是持怀疑态度的,通过对用户数据加密的方式可提高云服务的可信性,可预防云管理员窃取数据隐私。对敏感数据在用户端先加密再上传到云服务商,用户掌管解密密钥,从而只有用户自己拥有数据的访问权,云服务商只能得到密文。用户可控的加密已经成为近期隐私保护研究的热点。

第1种研究方案在应用程序客户端和用户之间对数据加密,如 ShadowCrypt^[23]作为浏览器扩展程序运行,在应用程序访问数据之前对数据加密。它使用安全隔离的 JS 和 DOM 操纵页面的输入,当用户输入明文时使用密文替换明文交给应用程序,当检测到应用程序包含的密文数据时解密得到明文返回给用户,从而应用程序不能访问明文,也不能破坏入侵隔离的 JS 和 dom。ShadowCrypt 不能加密非文本数据,不支持移动程序。Mimesis Aegis^[24]在移动平台提出隐私保护方案,通过处于应用程序和用户之间的 7.5 层加解密隐私数据,不仅提供很好的数据隔离性,而且保留原有的用户体验。Mimesis Aegis 适用于移动平台,只能保护文本型数据。

第2种研究方案在应用程序客户端和应用程序服务端之间对数据加密,如 Mylar^[25]是对 Meteor JavaScript 框架的扩展,通过在应用程序客户端和应用程序服务端之间对隐私数据加密来保证隐私性和安全性。Mylar 主要针对开发者,需要开发者调用 API 完成数据加密,应用程序需要在 Meteor JavaScript 框架上开发,影响了向后兼容性。Virtru^[26]对电子邮件加密预防网络邮件供应商,如 Gmail 偷窥用户的数据。但 Virtru 数据保护方案只适用于几家 Webmail 厂商,不能扩展到其他网络应用程序和移动应用程序。

第3种研究方案在应用程序服务端和数据库之间对数据加密,如 CryptDB^[27]在服务端程序和数据库之间加密隐私数据,数据以密文形式存储在数据库中,这样数据库管理员无法窃取用户的隐私。CryptDB 使用数据库代理保护用户隐私,其位于应用程序和数据库之间,将密文查询转换为对密文的操作,同时对返回的密文结果解密返回给用户。当应用程序客户端或服务端存在恶意程序时无法保护用户隐私。

3 体系结构

评价云平台的可信性需综合考虑 5 大方面的可信要素,包括人员可信、治理可信、云操作系统可信、基础设施可信、法规遵从可信。人员可信是指云服务商工作人员的录用、筛选、离岗和管理需要遵从规定的流程和规范;治理可信是指云服务商需对云平台进行符合安全标准的配置、监控、维护和管理;云操作系统可信是指云平台上的虚拟机镜像和模板应受到防护,网络资源需进行逻辑隔离,系统权限需划分;基础设施可信是指云服务上的机房地地理位置、硬件设施供应链和机房监控应符合标准;法规遵从可信包括重要产品和系统的合规检测、合同或协议中网络安全管理要求的体现、可移植性标准的遵守和企业信誉的保证。

本文设计并实现的云平台可信评测体系,综合分析以上 5 大可信要素,为租户提供选择可信云平台的依据,为云平台提供证明自身可信的凭据,其体系架构如图 3 所示。

可信第三方(TTP)服务网关通过逻辑串联、物理并联的方式部署在被评测云平台的前端,租户及云管理员的特权操作需通过 TTP 网关接入云平台。TTP 网关按照功能划分可分为“云平台特权行为分析与审计”部分以及“数据流可视化、监控与脱敏”部分。

“云平台特权行为分析与审计”用于持续、实时地采集云管理员对云平台的管理操作记录,并基于云管理员操作的分级机制对危险操作和越权操作进行预警。该模块实现了非强侵入式的证据收集,即在不修改云平台原有功能、不影响云服务性能、无需云服务商配合的前提下,收集云管

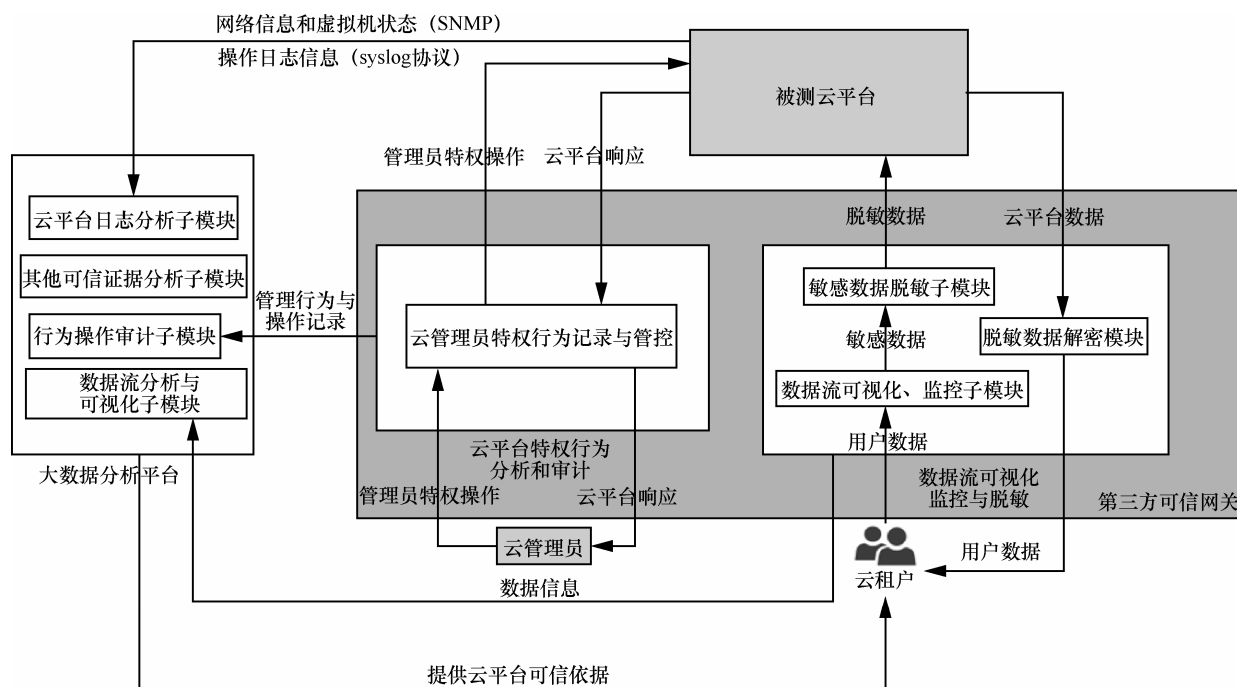


图3 面向第三方的云平台可信评测体系结构

理员操作记录。“云平台特权行为分析与审计”部分面临的主要挑战在于如何收集云平台管理行为和操作，并保证证据的可信性。

“数据流可视化、监控与脱敏”部分分为3个模块。“数据流可视化与监控子模块”截获并分析用户使用云服务时的上传数据流，找到敏感数据位置，同时为大数据分析平台提供可信性分析证据。“敏感数据脱敏子模块”加密敏感数据，并将脱敏数据传递给云平台。当用户下载云上数据时，“下载数据处理审核子模块”会解密用户敏感数据，并审核数据的完整性。“数据流可视化、监控与脱敏”部分的关键问题在于如何监控数据在租户和云平台之间的流动。

大数据分析平台综合了TTP服务网关采集的可信证据和云平台自身提供的可信证据，对云平台的可信性进行全面系统的分析。大数据分析平台分为4个子模块。云平台日志分析子模块：收集并分析云服务商主动提供的虚拟机日志和网络日志，评估云平台的运行状况和可信等级。行为操作审计子模块：对云管理员特权行为记录进行分析，及时发现云管理员的恶意行为，实现云平台管理行为的监控。数据分析与可视化子模块：分析用户上传云平台的数据，识别其中的敏感数

据并对数据流进行可视化的分析。其他可信证据分析子模块：综合云服务商提供的人员可信、基础设施可信、法规遵从可信相应证据清单，分析云平台的可信性。云平台日志分析子模块、行为操作审计子模块、数据分析与可视化子模块的综合评测结果对应可信要素的云操作系统可信和治理可信。

4 云平台特权行为分析和审计

云平台特权行为分析与审计模块主要针对两类数据进行评测：第一是可信第三方主动采集的云管理员特权行为数据，应用到体系结构中的云管理员特权行为记录与管控模块和行为操作分析与审计子模块，4.1节主要介绍这一部分数据的记录与审计；第二是云平台自身的日志数据，通常需要调用云平台API或由云服务商提供，应用到体系结构中的云平台日志分析子模块，4.2节主要介绍这些数据的采集与分析。

4.1 云管理员特权行为记录与审计

云管理员特权行为记录与审计功能部署于可信第三方代理内，如图4所示。该模块记录云管理员对于云上物理机、虚拟机、网络设备或云本身的操作并进行分析和预警处理。对于越权操作

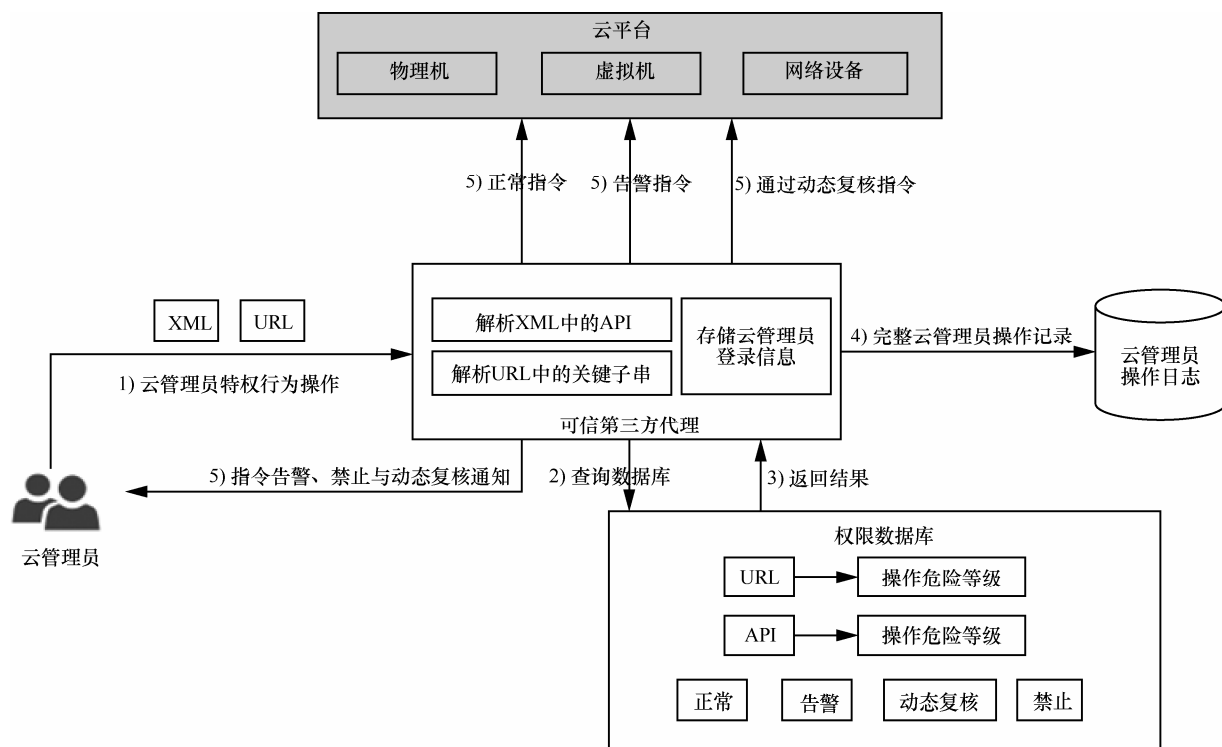


图4 云管理员操作采集、记录与审计

和危险敏感操作，该模块会给出相应的处理（告警、申请动态符合或禁止命令）。

4.1.1 云管理员特权行为记录

云管理员特权行为记录具体细节和流程如下。

根据云管理员其具体职能的不同，可信第三方会预先为其划分权限（即云管理员可执行的操作、敏感操作和危险操作，具体分为正常、告警、禁止、动态复核4个等级），云管理员的权限划分和权限对应的请求关键字（如vSphere云平台的API、OpenStack云平台的URL关键字串）保存在权限数据库中。

云管理员通过客户端（vSphere云平台）或Web端（OpenStack云平台）登录时，可信第三方代理服务器将记录其会话信息，包括用户ID、登录时间、操作的云平台信息、IP地址等，并将其与token值对应并存储。

云管理员发出服务请求时，代理服务器解析请求XML中包含的API信息或URL中的关键字序列，并识别请求中的token值从而得知云管理员信息。代理服务器根据API信息或URL关键字序列查询权限数据库，找到对应的操作

危险等级并进行相应处理。代理服务器将完整的云管理员请求过程记录在“云管理员操作日志数据库”中，从而完成了云管理员特权行为记录的采集。

4.1.2 云管理员特权行为等级划分、分析和审计

云管理员特权行为等级划分：根据记录中云管理员操作划分的4个等级，行为操作审计子模块进行如下处理。

1) “正常”等级，表示该操作是云管理员权限内操作，这类操作直接送交云平台执行。

2) “告警”等级，表示这类操作虽为操作云管理员权限内操作，但包含敏感指令和信息。第三方将此类指令送交云平台执行的同时，将通知云管理员进行敏感操作。

3) “禁止”等级，表示该类指令属于危险指令，第三方会截断该指令并警告执行操作的云管理员。

4) “动态复核”等级，表示该类指令不属于该管理员权限范围，但仍属正常操作范畴。可信第三方代理将会通知系统管理员对该云管理员操作进行动态审核，审核通过后可送交云平台执行，

反之拒绝操作并警告云管理员。

云管理员特权行为分析与审计：根据云管理员特权行为记录，分析模块可以检查云管理员是否进行了危险操作，包括：1) 在云平台上保存或安装与业务无关的程序、工具、脚本；2) 对不属于本管理员权限范围内的资源发起请求；3) 连续登录失败或输入错误口令；4) 执行了删除虚拟机、创建系统账号、配置访问授权、执行系统管理功能、审计系统事件或访问事件日志等敏感特权命令；5) 对虚拟机模板文件或镜像进行了篡改；6) 某些操作造成了大规模的系统开销或资源损耗等。分析模块综合考虑危险操作出现的次数和频率，同时解析云管理员上传文件的内容（是否含有恶意代码和冗余信息），进而评估云平台的可信程度，为用户提供可信依据。

4.2 云平台可信日志采集与分析

4.2.1 可信日志采集

为深度验证云平台的可信性，本文从以下几个方面收集并整理云平台日志信息。

1) 云平台操作日志：是指用户或云管理员操作云平台时产生的系统日志，主要基于虚拟机或物理机本身的 syslog，获取交换机和防火墙日志信息、apache 服务器信息、Web logic、Windows 和 Linux 系统日志等。

2) 云平台配置信息：主要采集 OpenStack 和 vSphere 上的主机、云平台、网络设备、虚拟机实例、虚拟机镜像、存储设备等配置信息。配置收集器的具体采集内容包括以下几方面。①管理员权限配置：扫描云平台自身的云管理员权限配置情况，生成权限划分报告；②云平台接口检测：通过扫描云平台控制节点的管理接口，给出被测云平台开放了哪些接口（如访问 vSphere 云平台可通过 SSH、Direct Control UI、ESXI Shell、vCenter Server 等接口）；③防火墙和网络配置信息。

3) 云平台虚拟机运行环境信息：获取租户虚拟机的内存数据，并对这些数据结合虚拟机所安装的操作系统和用户程序，解析其结构和语义，从而达到重构出运行环境状态和系统数据，如虚拟机的 CPU 利用率、内存利用率、进程数、硬盘读写、I/O 频率等方面的可信证据。

4) 云平台网络信息：主要包括两部分网络信

息：①基于 SNMP，感知云主机的运行状态及云主机端口流量信息；②基于 sFlow^[28]和 OVS(open vSwitch) 技术^[29]，获取网络的端口流量信息，包括端口发送量、端口接收量、分组丢失数、错分组数等。

由于云平台日志信息的采集部分依赖于云服务商自己提供数据或接口，因此验证评测证据的可信性和完整性尤为重要。本文将云平台日志数据与可信第三方的“小云”数据进行对比，甄别云平台数据中人为篡改的痕迹，从而实现云平台日志的可信性检查。

4.2.2 可信日志分析

1) 云平台操作审计。针对云平台操作日志，审查内容包括：①进程审计，主要检测用户态下是否运行了可疑或恶意程序；②模块扫描，查看虚拟机运行环境加载的内核态驱动程序和动态链接库；③文件审计，查看用户环境在运行时打开了哪些文件。综合以上三者，分析用户使用云平台时，云平台是否按照用户意愿运行，是否植入了不良文件和操作程序。

2) 云平台配置审核。主要分析云平台配置信息和网络信息，根据系统设定的安全基线，进行安全检查，分析云平台的安全状况。①云平台身份验证与用户管理：包括系统与用户密码策略审查，对云平台进行严谨的权限划分，如资源管理权限和用户创建权限等。②云平台安全配置审查：评估防火墙配置确保网络安全，评估网络隔离技术（如 vlan、vxlan）确保虚拟机网络隔离安全，检测云平台非必需端口提高云平台安全可信。③加密与安全证书：核心组件通信是否启动证书检查。④云平台上虚拟机安全配置：限制虚拟机操作系统复制到剪贴板的敏感数据，移除虚拟机上不必要的硬件设备。⑤检测云平台配置文件完整性。

3) 虚拟机监控管理分析。针对虚拟机运行环境信息对虚拟机的状态进行综合性的分析与审计。首先，对存储在数据库中的状态数据进行去重归处理，然后进行分析审计：①对虚拟机物理资源如 CPU、内存、硬盘空间、I/O 频率的使用情况以时间为轴进行可视化，对资源使用超过用户指定安全阈值的进行标记；②分析虚拟机的内核崩溃信息，如内核崩溃的时间、导致崩溃的

进程和原因等。综合分析,将以上的两点结合云平台操作日志,形成状态异常七元组(进程号、进程启动时间、进程退出时间、指令(服务)、读写目录、物理资源使用情况、内核崩溃信息),并根据异常状态产生原因对其进行分类,可分为操作系统级别异常、服务级别异常和应用级别异常。

5 数据流可视化、监控与脱敏

数据流可视化、监控与脱敏通过分析、获取和加密用户的敏感数据,使其在组织内部和任何公有云、私有云之间进行自由穿梭,满足在使用云应用服务的过程中保障敏感数据的安全性和私密性,从而增强了云平台的可信性。以下主要从云应用服务数据流可视化、数据监控与数据脱敏3个方面进行阐述。

5.1 数据流可视化

数据流可视化的内容包括用户、数据、访问设备、来源等。数据流可视化首先要能截获用户和云服务之间的应用层数据流,如用户通过浏览器访问云应用服务时,要能截获并分析http、https等主流云应用协议,对不同的应用(主流的邮箱、网盘和在线SaaS等)进行适配,获取满足不同需求的应用层数据。

本文通过使用安全访问代理的方式保护用户的云应用层数据。用户访问云应用服务时需要经过安全访问代理,访问代理截获并过滤应用层数据流,识别应用层协议并对敏感数据进行加密,之后再用户的访问请求转发给云应用服务;云应用服务回复内容也首先交给访问代理,由访问代理识别密文并解密,经过检验与审计将结果返回给用户,从而提供数据流可视化与监控,保护用户敏感数据,有效防止数据泄露。

5.2 数据监控

数据监控自动识别和解析不同云应用服务的协议,通过语义分析和特征库提取得到敏感数据内容。协议识别和语义分析依据协议请求的URL等特征调用具体的协议处理逻辑做敏感数据的提取,具体的协议处理逻辑对内容格式(如key-value、JSON、XML、multi-part等)进行分析,找到用户的关键数据内容。

通过解析各种格式的数据流得到敏感数据的属性和内容,如邮件的主题和正文。每种协议有固定的数据格式,通过分析各种协议格式找到共同的处理函数抽象成工具库,则具体的协议处理会调用相关工具库,分析和提取敏感数据。如果适配新的应用程序,只需要识别协议格式所属的类别,并添加部分代码调用相关工具库即可识别并监控新应用程序的敏感数据。为了适配某一应用程序协议的变化,对每一类应用程序要维持一个特征库,分析协议时根据特征库获得协议的敏感数据。这样当协议发生改变时,协议检测逻辑可检测到协议特征变化,更新特征库,从而不需要改变具体协议处理逻辑自动适应协议的变化。

数据监控利用访问代理可识别和分析不同协议的多种数据格式,如文本、文件、图片、音视频、应用自定义的格式数据等,从而实时监控敏感数据流动情况,防止数据泄露。

5.3 数据脱敏

本文实现数据脱敏的手段是对敏感数据加密保护用户数据隐私。当数据监控获取到云应用协议的用户敏感数据后,需要调用加密模块对敏感数据加密。数据脱敏使用标准加密算法,且需要在密文前添加元数据记下密文和密钥之间的对应关系及密文的特征标志。根据需要加密数据的长度从原协议数据流中取出敏感数据并加密,然后将密文置于原协议数据流相应位置,从而不改变原有协议,对云应用服务和用户透明。加密采用分块加密,属于一个请求体的各数据块拥有共同的元数据,包括密钥标志、特征数据等。同样解密模块首先匹配特征找到密文对应位置,分析密文元数据,依据密钥标志找到对应的密钥,然后解密各块恢复明文。以一个典型的应用邮箱的附件加密为例,附件加密采用分块加密并传输,每次对固定长度的数据块进行加密,直到加密完所有文件数据,最后在密文末尾添加结束标志。解密也是逐块解密,直到遇到结束标志。加密或解密模块为通用模块,分析任何协议得到要加密的敏感数据后都可调用加密功能加密,同样匹配到密文后需调用解密功能解密。数据脱敏在保护用户隐私数据的同时,在不改变云服务接口和用户使用习惯的情况下还要保持现有云应用服务的功

能,如密文搜索功能,即用户能够对加密后的敏感数据进行全功能搜索。实现思路为当敏感数据加密后上传到云应用服务后云服务会返回密文标识,访问代理得到密文标识后为敏感数据建立索引,这样用户搜索时,访问代理提取到搜索的关键词并转发给索引进行搜索,得到相应的密文标识经访问代理返回给用户,用户使用密文标识通过代理访问云服务,云应用服务返回相关密文经由访问代理解密并检验,然后将明文结果返回给用户,完成搜索功能。

如图 5 所示,加密过程为:用户访问云应用服务的请求首先到达访问代理,经过过滤进行协议分析,提取特征调用具体的协议处理逻辑解析得到要加密的敏感数据,调用加密模块加密,加密模块首先产生加密密钥,将密钥和密钥标志存入数据库,将密钥标志、特征字段作为元数据内容附在密文头,然后对数据加密,加密完成的密文及元数据放到原协议相应位置。解密过程为:用户请求数据,云应用服务收到请求并回复,回复内容首先到达访问代理,经过过滤进行协议分析,提取特征调用具体的协议处理逻辑通过扫描匹配特征字段定位到要解密的密文,根据元数据密钥标志查询数据库获得相应的密钥,解密数据得到明文放置到相应位置。搜索过程为:用户访问云应用服务的搜索请求到达访问代理,经过过滤进行协议分析,提取特征调用具体的协议处理逻辑得到要搜索的关键词,用随机串替换关键词发给服务器,服务器会返回空的结果列表;同时

访问代理使用关键字向索引发起搜索请求,索引返回搜索结果为相关数据对应的密文标识,访问代理再用搜索结果替换服务器返回的结果列表,完成搜索过程。

6 结束语

租户将数据与业务系统迁移到云上,客观上平台管理员具有偷窥和泄露租户隐私的能力。为此本文提出面向第三方的云平台可信评测体系结构与关键技术,即“小云审大云”的云平台可信评测架构,主要针对云平台特有的可信问题,研究从第三方角度如何验证、审计和评测云平台的可信性。在传统安全评测手段基础上引入独立第三方 TTP,通过 TTP 对云平台的可信性进行动态、实时的远程可信数据收集、验证、审计和评价,并采用数据流可视化、监控与脱敏技术保障租户的数据权益。可信评测可根据租户的需求,由第三方发起,为租户提供选择可信云平台的依据,为云平台提供证明自身可信的凭据。

在深入研究与系统评测的过程中,本文首先分析了云计算平台面临的安全风险,进而通过限制云平台访问入口严格管控、记录与审计云管理员特权行为和用户操作,对云服务应用层数据可视化、监控与脱敏保护用户的安全隐私,并收集与分析云平台可信日志监控云平台配置与状态,由日志分析模块综合分析多种数据对可疑事件产生告警。实施过程中主要遇到并突破了以下关键技术。

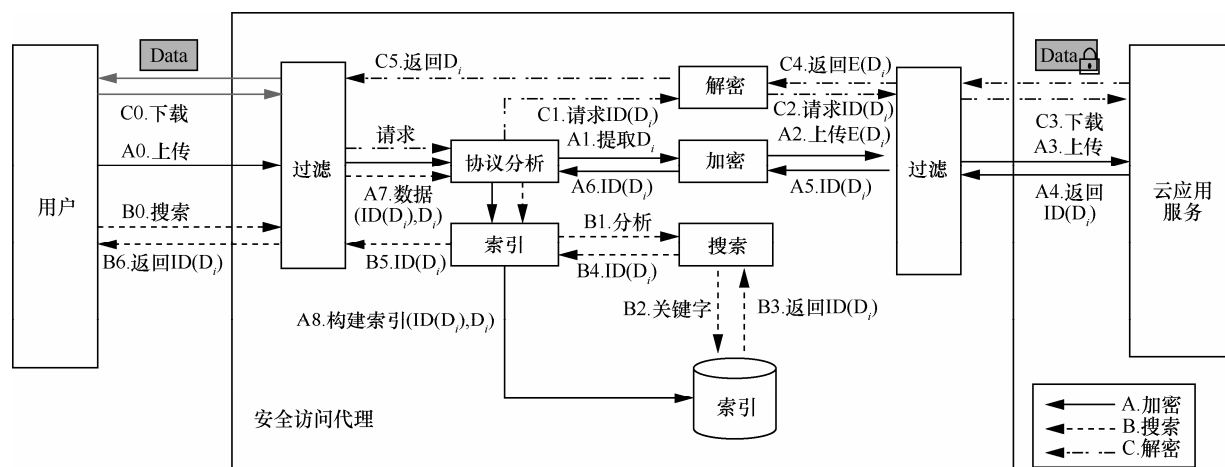


图 5 数据流可视化、监控与脱敏

1) 引入独立第三方作为云平台与用户、管理员的中间层, 独立采集和审计云平台可信证据。

2) 通过持续的、实时的远程数据收集和数据分析, 评测云平台的可信性。

3) 非强侵入式证据收集, 在不影响云平台自身性能、不影响云服务原有功能、无需修改云平台的前提下收集可信证据, 记录特权行为。

基于上述体系结构的重要研究点和未来研究方向包括以下3个方面。

1) 进一步融合多种数据, 实现多源数据的综合分析。在大数据分析平台中, 不同分析子模块对不同来源和种类的数据进行了独立的分析, 综合这些分析结果为用户提供选择云平台的可信性依据。进一步的工作可以从可信证据之间的联系入手, 在某些分析子模块中综合考虑其他子模块收集的可信证据, 以提高可信性分析的全面性。

2) 云平台日志真实性检验。针对云服务商自主采集的数据和通过云平台接口采集的数据, 可以对比模拟情况下云平台的运行环境、结合机器学习算法验证数据的真实性和完整性。

3) 基于目前的数据流可视化、监控与脱敏模块, 实现多种云服务应用的适配。为实现数据流可视化、监控与脱敏, TTP 服务网关需截获并自动分析 http、https 等主流云应用协议, 对不同的应用进行自动适配, 获取满足不同需求的应用层数据。

参考文献:

- [1] BOUCHÉ J, KAPPES M. Attacking the cloud from an insider perspective[C]//Internet Technologies and Applications (ITA). 2015: 175-180.
- [2] ROCHA F, CORREIA M. Lucy in the sky without diamonds: Stealing confidential data in the cloud[C]//2011 IEEE/IFIP 41st International Conference on Dependable Systems and Networks Workshops (DSN-W). 2011: 129-134.
- [3] [EB/OL].<http://docs.openstack.org>.
- [4] [EB/OL].<http://www.qemu.org>.
- [5] [EB/OL].http://www.linux-kvm.org/page/Main_Page.
- [6] [EB/OL].<https://www.xenproject.org>.
- [7] [EB/OL].<http://www.vmware.com/cn/products/vsphere.html>.
- [8] [EB/OL].<http://hvreport.codeplex.com>.
- [9] BHATT S, MANADHATA P K, ZOMLOT L. The operational role of security information and event management systems[J]. IEEE Security & Privacy, 2014, 12(5): 35-41.
- [10] YEN T F, OPREA A, ONARLIOGLU K, et al. Beehive: large-scale log analysis for detecting suspicious activity in enterprise networks[C]//The 29th Annual Computer Security Applications Conference. 2013: 199-208.
- [11] ZOMLOT L, SUNDARAMURTHY S C, LUO K, et al. Prioritizing intrusion analysis using dempster-shafer theory[C]//The 4th ACM Workshop on Security and Artificial Intelligence. 2011: 59-70.
- [12] ZHAI Y, NING P, IYER P, et al. Reasoning about complementary intrusion evidence[C]//The 20th Computer Security Applications. 2004: 39-48.
- [13] [EB/OL].<https://www.gartner.com/doc/3406817?ref=SiteSearch&stkw=Splunk&fml=search&srcl=1-3478922254>.
- [14] [EB/OL].<http://www-01.ibm.com/software/>.
- [15] [EB/OL].<http://www.splunk.com/>.
- [16] [EB/OL].<https://logrhythm.com/>.
- [17] [EB/OL].<https://www.elastic.co/products>.
- [18] [EB/OL].<http://opensoc.github.io/>.
- [19] BUTT S, LAGAR-CAVILLA H A, SRIVASTAVA A, et al. Self-service cloud computing[C]//The 2012 ACM Conference on Computer and Communications Security. 2012: 253-264.
- [20] ZHANG F, CHEN J, CHEN H, et al. CloudVisor: retrofitting protection of virtual machines in multi-tenant cloud with nested virtualization[C]//The 23rd ACM Symposium on Operating Systems Principles. 2011: 203-216.
- [21] CHEN X, GARFINKEL T, LEWIS E C, et al. Overshadow: a virtualization-based approach to retrofitting protection in commodity operating systems[C]//ACM Sigarch Computer Architecture News. 2008, 36(1): 2-13.
- [22] HOFMANN O S, KIM S, DUNN A M, et al. Inktag: secure applications on an untrusted operating system[C]//ACM Sigarch Computer Architecture News. 2013, 41(1): 265-278.
- [23] HE W, AKHAWA D, JAIN S, et al. Shadowcrypt: encrypted web applications for everyone[C]//The 2014 ACM Sigsac Conference on Computer and Communications Security. 2014: 1028-1039.
- [24] LAU B, CHUNG S, SONG C, et al. Mimesis aegis: a mimicry privacy shield—a system's approach to data privacy on public cloud[C]//The 23rd USENIX Security Symposium (USENIX Security 14). 2014: 33-48.
- [25] POPA R A, STARK E, VALDEZ S, et al. Building Web applications on top of encrypted data using Mylar[C]//The 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14). 2014: 157-172.
- [26] [EB/OL].<http://www.virtu.com>.

[27] POPA R A, REDFIELD C, ZELDOVICH N, et al. CryptDB: protecting confidentiality with encrypted query processing[C]//The 12rd ACM Symposium on Operating Systems Principles. 2011: 85-100.

[28] [EB/OL].<http://www.sflow.org>.

[29] [EB/OL].<http://openvswitch.org>.

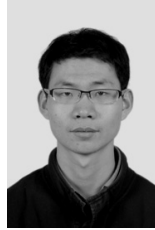


梁露露 (1985-), 男, 湖北襄阳人, 博士, 中国信息安全测评中心副研究员, 主要研究方向为云计算与大数据安全、无线网络安全。

作者简介:



刘川意 (1982-), 男, 四川乐山人, 博士, 哈尔滨工业大学(深圳)副教授, 主要研究方向为云计算与云安全、大规模存储系统、数据保护与数据安全。



王国峰 (1988-), 男, 山东济宁人, 北京邮电大学博士生, 主要研究方向为云安全、信息安全、云安全技术、数据加密技术。



潘鹤中 (1991-), 男, 辽本本溪人, 北京邮电大学博士生, 主要研究方向为云安全、信息安全。



方滨兴 (1960-), 男, 江西万年人, 中国工程院院士, 北京邮电大学教授、博士生导师, 主要研究方向为大数据、计算机网络和信息安全。