基于虚拟化技术的入侵容忍系统

云计算将计算任务分布在大量计算机构成的虚拟资源池上，使各种应用系统能够根据需要获取计算能力、存储空间和各种软件服务。云计算带来的服务整合与按需供给能够显著提高计算资源的利用率，降低每服务的能耗量，并且有效屏蔽计算资源的出错问题。然而云计算依然面临着云计算平台的容忍入侵性和新企业数据中心数据的容忍入侵性。国内外许多研究机构和组织对容忍入侵的研究大多是在真实硬件平台上采用门限密码学、秘密共享以及分布式冗余复制等技术；都要求系统具有高计算能力和高存储能力，并且这些资源一般不能分割，是针对特定应用而设计，因此资源利用率低，难以管理，通用性差；另外系统对冗余组件的数目与质量也有较高要求，并且冗余组件恢复时严重影响服务的可用性。因此，这样高昂代价的容忍入侵对于云计算服务提供商和用户来说是很难接受的，并且不适用于云计算的虚拟化。

为解决云计算平台和新企业数据中心数据的容忍入侵性问题，本文通过研究现有的虚拟化技术和秘密共享，构建了一种基于云计算的虚拟化容忍入侵系统，并基于此提出了一种基于虚拟攻击结构的云计算数据中心敏感数据容忍入侵保护的方案。该系统使用了异构容错模型、活动与被动副本、状态更新与状态转移、积极恢复和多样性策略，初步实现了在N=2F+1个副本中容忍F个副本的错误；并保证系统在无人入侵时只有F+1个活动副本在执行，其余副本置于被动模式，这样大量减少了云平台资源的消耗。最后我们验证了敏感数据的容忍入侵方案通过秘密共享实现了关键数据或敏感数据的重构特性和保密特性。

**引言**

云计算将计算任务分布在大量计算机构成的虚拟资源池上，时各种应用系统能够根据需要获取计算能力、存储空间和各种软件服务。云计算带来的服务整合与按需供给能够显著提高计算资源的利用率，降低每服务的能耗量，并且有效屏蔽计算资源的出错问题。云计算具有超大规模、虚拟化、高可靠性、通用性、高扩展性、按需服务、极其廉价等优点。但云计算依然面临着以下问题：

1. 数据的容忍入侵性：云计算最重要的两个问题是保密性和安全性。企业把企业信息、客户信息等敏感的商业数据存放到云计算服务提供商手中，通过减少对某些数据的控制来节约经济成本，因此企业必然会担心数据放在云端安全吗？会不会被泄密和窃取？
2. 云计算平台的容忍入侵性：云计算模式下，所有的业务处理都将在服务器端完成，因此云计算平台就成为入侵者攻击的核心；另外云计算平台自身也面临着各式各样无法预料的错误，比如2010年上半年，Amazon云计算服务就因为人为失误和意外停电事故出现故障，致使美国东部的少量用户失去服务并导致极少数据丢失。因此如果服务器一旦出现问题，就将导致所有的服务无法正常运行，数据无法访问。

对于云计算面临的保密性、完整性、可用性和可靠性等问题，传统的安全技术如入侵检测、防火墙、加解密等都无法完全保证，但是一种可生存技术即容忍入侵技术可以很好的弥补。本文将通过研究相关的容忍入侵技术来解决云计算平台和新企业数据中心敏感数据的容忍入侵问题。

容忍入侵主要思想就是利用分布式系统中的硬件或者软件容错技术屏蔽任何入侵或者攻击对系统功能的影响，保证系统关键功能的安全性和连续性。国内外许多研究机构和组织对容忍入侵的研究大多是在真实硬件平台上采用门限密码学、秘密共享以及分布式冗余复制等技术；这些方案都要求系统具有高计算能力和高存储能力，并且这些资源一般不能分割，是针对特定应用而设计，因此资源利用率低，难以管理，通用性差；另外系统对冗余组件的数目与质量也有较高的要求，并且冗余组件恢复时严重影响服务的可用性。当前秘密共享方案大多基于存取结构，但在设计容忍入侵系统时，要想根据系统的配置（属性）及其安全性需求确定出相应的存取结果十分困难，原因在于存取结构描述的是系统中的合格（授权）子集，而系统配置及安全性需求等描述的则是系统面临的威胁和自身的漏洞，以及在该配置下应达到的安全性目标等。另外，如果在真实硬件平台上实现秘密共享要求每个参与方都是一个真实的服务器节点，对于运行关键业务的大型平台来说，秘密共享则需要大量物理机节点的参与，这对资源的消耗和浪费是相当大的。因此这样高昂代价的容忍入侵对于云计算服务提供商和用户来说是很难接受的，并且不适用于云计算的虚拟化。

虚拟化技术自1960s诞生以来发展已经相当成熟了，但其应用大多仍是关于资源的高效管理，实际上虚拟化技术也可以用来构建安全系统，如Hans P.Reiser, Rudiger Kapitza 等人就在虚拟化技术上研究过容忍入侵系统。随着云计算的提出，研究虚拟化技术上云计算的安全性变得至关重要。

为了解决云计算平台的容忍入侵性问题，本文提出了一种基于云计算的虚拟化容忍入侵（Virtualization Intrusion Tolerance Based on Cloud Computing,CC-TIV）方案，该方案能在N=2F+1个副本中容忍F个副本的错误，并保证系统在无入侵时只有F+1个活动副本在执行，其余副本处于被动模式，这样大量减少了云平台资源的消耗。而传统的Byzantine容错算法容忍F个错误副本需要N=3F+1个副本。

对于云计算新企业数据中心敏感数据的容忍入侵性问题，本文基于CC-VIT模型提出了一种基于虚拟攻击结构的云计算数据中心敏感数据的容忍入侵保护方案，该方案将敏感数据分割成片段（划分）分发给每个参与方，保证使得只有合法的参与者才能联合他们的划分重构敏感数据，同时非法的参与者联合他们的划分不能获得关于敏感数据的任何信息。该方案中所有的参与方都是运行于虚拟机监控器上的虚拟机，因此节省了大量的物理机服务器。最后我们验证了敏感数据的容忍入侵方案实现了敏感数据的重构特性和保密特性，即该方案能有效保证敏感数据的安全性和完整性，并能有效降低物理资源的消耗。

1. **绪论**
   1. 本文背景

从2007年下旬开始，“云计算”变成了炙手可热的一个词。Google 、Amazon、IBM、微软和Salesforce等巨头们以其前所未有的速度和规模推动云计算技术及其产品的普及，一些学术活动迅速将云计算提上议事日程，支持和反对的声音不绝于耳。

从图1.1中Google对关键词“Cloud Computing”的搜索趋势反映出，“Cloud Computing”是2007年第4季度才兴起的一个词汇，并从此成为世人搜素和关注的热点。

我国对云计算的关注和研究也较早，业界和政府有关部门高度重视这一新兴的技术。为了把握云计算的实质内涵及发展趋势，探讨云计算对产业、教育和社会发展的影响，交流国内外云计算的最新研究成果，分享云计算应用的实践经验，在中国工业和信息化部、教育部、中国科协及相关部委的指导下，中国电子学会于2009年5月22日在北京中国大饭店隆重举办首届中国云计算大会，到2011年5月20日已是举办第三届中国云计算大会。每届中国云计算大会都云集了大批两院院士、IT企业精英高管、各大科研院所的专家学者等，可谓是群英荟萃，举世瞩目的盛会。

虽然云计算在中国的发展时间并不长，但最新数据显示，其发展已经具备了一定的规模，且呈现上升之势。2010年5月21日第二届中国云计算大会上，由埃森哲卓越绩效研究院与中国电子学会云计算专家委员会推出权威的云计算调研报告《中国云计算发展的务实之路》，该报告给出了当前中国云服务供应商名单节选。

在报告名单节选中，八百客位居SaaS供应商之首，是当之无愧的中国SaaS业第一品牌。800APP在企业的应用也作为唯一的CRM案例成功入选报告。其中SaaS即云计算的软件即服务（software as a Service ,SaaS）。

* + 1. 云计算的出现与发展

云计算是一种新兴的商业计算模型。它将计算任务分布在大量计算机构成的资源池上，使各种应用系统能够根据需要获取计算力、存储空间和各种软件服务。这种资源池称为“云”。“云”是一些可以自我维护和管理的虚拟计算资源等等。云计算将所有的计算资源集中起来，并由软件实现自动管理，无需人为参与。这使得应用提供者无需为繁琐的细节而烦恼，能够更加专注于自己的业务，有利于创新和降低成本。

资源池之所以称为“云”，是因为它在某些方面具有现实中云的特征：云一般都较大；云的规模可以动态伸缩，它的边界是模糊的；云在空中飘忽不定，我们无法也无需确定它的具体位置，但它确实存在于某处。资源池之所以称为“云”，还因为云计算的鼻祖之一亚马逊公司将曾经大家称作网格计算的东西，取了一个新名称“弹性计算云（EC2），并取得了商业上的成功。

我们都知道网格计算可以形象地比喻成即插即用的水电煤气厂，能够提供无穷的计算能力。类似于网格计算，我们也可以将云计算比喻成水电模式，只是云计算更倾向于将单台发电机集中起来形成超大规模的发电厂，以集中提供廉价便捷的即插即用式服务。也就是说云计算能够提供的计算能力可以像水电煤气那样作为商品出售，为用户提供更加廉价便捷的服务。最值得关注的是云计算一般是通过Web提供服务的，通过互联网可以实现任何地点任何时间使用任何服务。

有了云计算形象的比喻，我们就可以很好地理解云的本质。总的来说云计算具有以下特点：（1）超大规模：“云”的规模一般比较庞大，从几十万到上百万台服务器不等，“云”能赋予用户前所未有的计算能力。（2）虚拟化：云计算使用户能够在任何地方使用任何服务而不用关心服务存在的具体位置，服务来自于虚拟动态变化的云端。（3）高可靠性：“云”使用了数据多副本容错和计算节点同构可互换等技术来保障服务的高可靠性，云计算比本地计算机节点更可靠。（4）通用性：云计算的设计不针对特定的应用，在云计算框架下可同时支持多种不同的应用。（5）高可扩展性：“云”的规模可以动态伸缩。云计算可满足应用和用户规模增长的需要。（6）按需服务：“云”是一个庞大的资源池，用户可以按需购买并可以像自来水、电和煤气那样计费。（7）极其廉价：由于“云”的特殊容错措施可以采用极其廉价的节点来构成云，“云”的自动化集中式管理使大量企业无需负担日益高昂的数据中心管理成本，因此用户可以充分享受“云”的低成本优势。

* + 1. 云计算面临的问题

虽然云计算具有相当多诱人的优点，但是云计算面临的问题也显而易见，其中比较重要的几个问题是可靠性、可用性和安全性等问题一直是影响云计算发展和推广的一大障碍。经过长时间分析和研究，我们将云计算的这些问题归结为云计算平台和云计算数据中心的容忍入侵性问题，现简要介绍如下：

（1）云计算数据中心敏感数据的容忍入侵性：云计算最重要的两个问题是保密性和安全性。企业把企业信息、客户信息等敏感的商业数据存放在云计算服务提供商的手中，通过减少对某些数据的控制来节约经济成本，因此企业必然会担心数据放在云端安全吗？

（2）云计算平台的容忍入侵性：云计算模式下，所有的业务处理都将在服务器端完成，因此云计算平台就成为入侵者攻击的核心；另外云计算平台自身也面临着各式各样无法预料的错误，比如2010年上半年，Amazon云计算服务就因为人为失误和意外停电事故出现故障，致使美国东部的少量用户失去服务并导致极少数据丢失。因此如果服务器一旦出现问题，就将导致所有的服务无法正常运行，数据无法访问，甚至会出现数据丢失的危险。

对于云计算面临的完整性、可用性和可靠性等问题，传统的安全技术如入侵检测、防火墙、加解密等都无法完全保证，例如对于云计算新企业数据中心敏感的服务数据，当数据处于系统平台内存中进行计算时数据是非加密的，系统就面临着平台被攻陷后数据丢失的危险。虽然传统的安全技术如入侵检测、防火墙、加解密等不能解决这个问题，但是一种可生存技术即容忍入侵技术可以很好的弥补，因此本文通过研究相关的容忍入侵技术以及虚拟化来解决云计算平台的容忍入侵问题，并基于秘密共享来解决云计算新企业数据中心敏感数据的容忍入侵问题。

* 1. 容忍入侵研究现状

早在上世纪80年代中期，Dobson和Randen就提出利用不安全并且不可靠的组建来构建安全可靠的系统，这实际上就是容忍入侵思想的雏形。由于受当时相关技术局限性的限制，容忍入侵的思想没有得到广泛的应用和推广。直到近几年来，随着分布式密码学的研究，特别是秘密共享和门限密码学研究的成熟与完善，再加上分布式网络应用系统的应用，容忍入侵思想又开始进入人们的视野，并且逐渐成为信息安全领域研究的一个焦点。

容忍入侵技术就是所谓的第三代网络安全技术，它是信息生存技术的重要组成部分，卡耐基梅隆大学的学者给这种生存技术下了一个定义：所谓“生存技术”就是系统在攻击、故障和意外事故已发生的情况下，在限定时间内完成使命的能力。它假设我们不能完全正确地检测对系统的入侵行为，当入侵和故障突然发生时，能够利用“容忍”技术来解决系统的“生存”问题，以确保信息系统的保密性、完整性、真实性、可用性和不可否认性。

国际性该领域主要以美国DARPA（美国国防高级研究计划署）与NSF（美国国家自然科学基金）研究

1.3 基于虚拟化的云计算容忍入侵方案的提出

国内外许多研究机构和组织对容忍入侵的研究大多采用门限密码学、秘密共享以及分布式冗余复制等技术，并取得了丰硕的成果。

随着计算机技术的发展，2007年出现了一种新的计算模型­­——云计算，不失一般性，云计算仍然存在可生存的问题，于是我们考虑将容忍入侵思想应用到平台之上，而云计算是在虚拟化技术基础上发展而来的。然而国内外许多研究机构和组织对容忍入侵的研究大多是在真实硬件平台上采用门限密码学、秘密共享以及分布式冗余复制等技术；都需要系统具有高计算能力和高存储能力，并且这些资源一般分割，是针对特定应用而设计，因此资源利用率低，难以管理，通用性差；另外系统对冗余组件的数目与质量也有较高要求，并且冗余组件恢复时严重影响服务

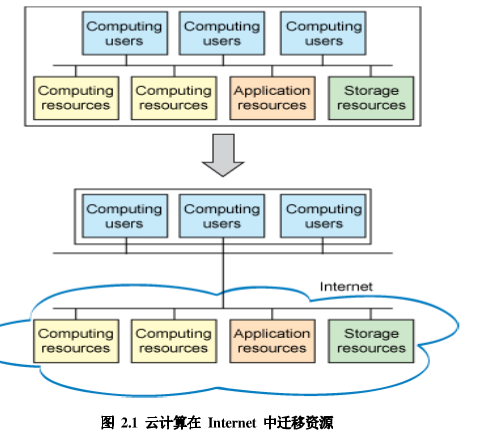
1.5 本文主要工作和组织结构

本文在现有的虚拟化技术Xen之上构建云计算容忍入侵系统，并基于CC-VIT模型提出了一种基于虚拟化攻击结构的云计算新企业数据中心敏感数据容忍入侵保护方案。该系统能在N=2F+1个副本中容忍F个副本的错误，并保证系统在无入侵时只有F+1个活动

2 云计算技术简介

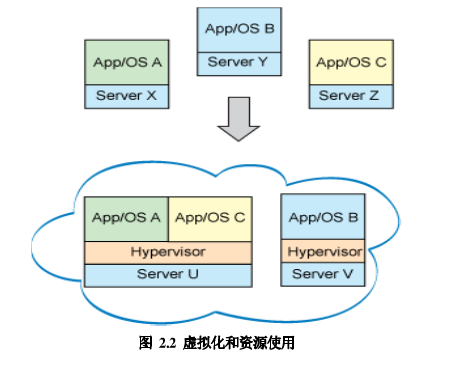
云计算最重要的特点之一就是可伸缩性，而实现它的关键技术则是虚拟化（Virtualization）。虚拟化允许在一台物理机节点上同时运行多个系统和应用实例，从而有效的提高服务器的利用率，节省大量的物理资源。虚拟化另外一个有趣也很有用的特点就是在线迁移，当某台物理服务器节点负载过重时，我们可以通过相应操作将某台运行中的虚拟机动态迁移到一台新的负载较轻的物理机节点，从而达到均衡负载的目的。

从宏观上看，云计算只是将计算和存储资源从企业自身的数据中心迁移到公共的云端。用户可以选择并向云服务提供商提交自己的资源需求，包括计算资源、存储资源、网络带宽等等，云服务提供商再从公共云端基础设施中划分出相应的资源给用户。



通过虚拟化划分资源和在线提供服务，云计算能够实现提供廉价资源的目的，并且能够将用户从管理数据中心的大量繁琐的工作中解脱出来，另外云计算提供者还可以轻松地扩展虚拟环境，以通过服务提供者的虚拟基础设施提供更大的带宽或计算资源。

在云中，可以在多个操作系统和应用程序之间共享（虚拟化）服务器，从而减少服务器的数量。更少的服务器意味着需要更少的空间（减少数据中心占用的空间）和更少用于制冷的电力（减少碳污染）。



2.1 云计算的概念及演进

云计算是这个商业计算模型。它将计算任务分布在大量计算机构成的资源池上，使各种应用系统能够根据需要获取计算能力、存储空间和各种软件服务。

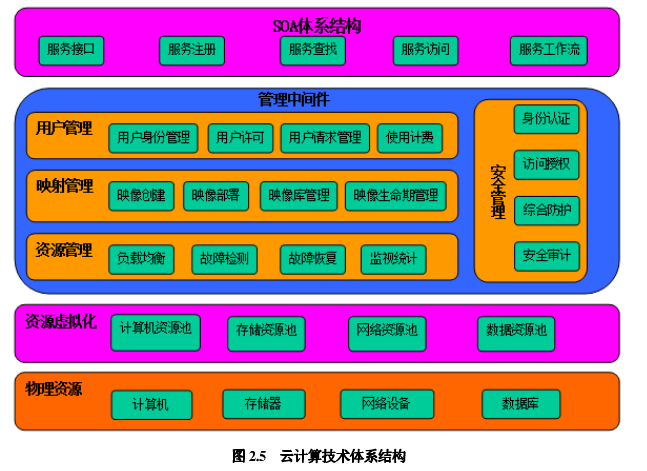
云计算通过互联网将超大规模的计算存储资源整合起来，形成一个虚拟的计算资源池，并以服务的形式按需提供给用户。云计算带来的服务整合与按需供给能够显著提高计算资源的利用率，降低每服务的能耗量，并且有效屏蔽计算资源的出错问题。

云计算并不是革命性的新发展，而是历经十载不断演进的结果，上世纪80年代末，开始出现应用大量系统来解决单一问题（通常是科学问题）的情况，这就是网格计算的概念，而这种概念又导致向云计算的发展。云计算是并行计算、分布式计算和网格计算的发展，或者说是这些计算机科学概念的商业实现。云计算使虚拟化、效用计算、IaaS、PaaS、SaaS等概念混合演进并跃升的结果，因此业界也有很多学者将云计算简单定义为Everything as a Service。

“云”计算产业可以归纳为一个金字塔模型。处于金字塔基座位置的是基础设施层，提供计算、存储、带宽等按需的IaaS云基础设施服务，这是所有应用和平台的基础，也是虚拟化、自动化等云计算关键技术的集中体现层，如Amazon的AWS，世纪互联的CloudEx；基于基础设施之上的是为应用开发提供接口和软件运行环境的平台层的PaaS服务，如微软的Azure和Google的App Engine；处于金字塔顶端的是应用层，提供在线的软件服务即SaaS服务，最知名的就是SalesForce了，另外还有NetSuite、Google Apps、微软的Live Office等。这三个层面合起来构成了一条完整的云计算产业链。

2.2 云计算技术体系结构

云计算技术体系结构分为4层：物理资源层、资源池层、管理中间层和面向服务架构SOA（Service-Oriented Architecture）构建层。

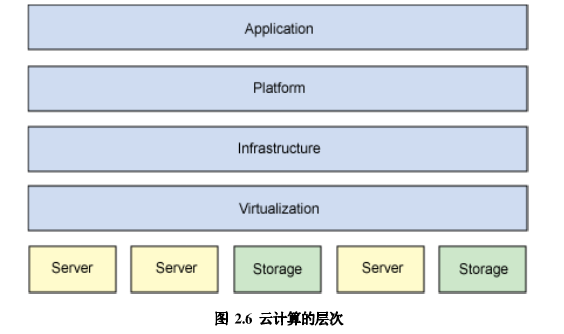


物理资源层包括计算机、存储器、网络设施、数据库和软件等；资源池层是将大量相同类型的资源构成同构的资源池，如计算资源池、数据资源池等，构建资源池更多是物理资源的集成和管理工作，例如研究一个标准集装箱的空间如何装下2000个服务器、解决散热和故障节点替换的问题并降低能耗；管理中间件负责对云计算的资源进行管理，并对众多应用任务进行调度，使资源能够高效、安全地为应用提供服务；SOA构建层将云计算能力封装为注册、查找、访问和构建服务工作流等。其中管理中间件和资源池是云计算技术最关键部分，SOA构建层的功能是更多依靠外部设施提供。

云计算的管理中间件负责资源管理、任务管理、用户管理和安全管理等工作。资源管理负责均衡地使用云资源节点，检测节点的故障并试图恢复或屏蔽之，并对资源的使用情况进行监视统计：任务（映射）管理负责执行用户或应用提交的任务，包括完成用户任务映象（Image）的部署和管理、任务调度、任务执行、任务生命期管理等等；用户管理是实现云计算商业模式的一个必不可少的环节，包括提供用户交互接口、管理和识别用户身份、创建用户程序的执行环境、对用户的使用进行计费等；安全管理故障云计算设施的整体安全，包括身份认证、访问授权、综合防护和安全审计等。

2.3 云计算栈剖析

云计算从不同层次角度可以提供一系列的服务，因为云计算是一个服务的集合体。云计算层次栈从底层到高层的每一层都定义了提供的服务的级别。

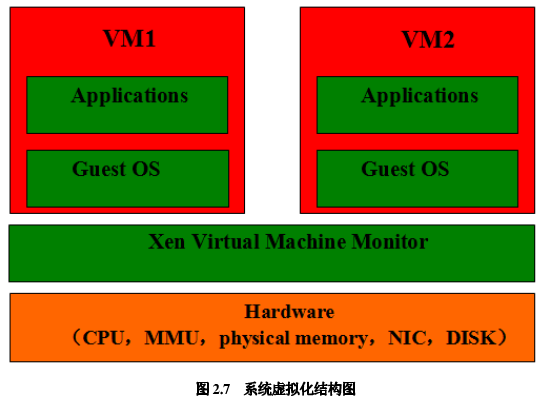


最底层的服务是基础设施即服务，就是将计算资源、存储资源和网络资源等基础设施最为服务出租。这样不仅可以保证虚拟机具有足够多的处理能力，而且也可以为存储和Internet访问预留足够的网络带宽资源。实际上，IaaS可以提供满足特定Qos条件的任何计算机或数据中心的能力，使之能执行任意操作系统和服务软件。IaaS的上一级是平台即服务层。PaaS类似于IaaS，但是它包括操作系统和围绕特定应用的必需的服务。PaaS实际上就是IaaS加上一个用于特定应用的定制软件栈。最后，在顶层是直接为用户提供应用程序服务的软件即服务层，软件即服务(SaaS)是通过互联网向用户提供相应的服务，客户根据需要定制相应的服务，并根据服务的使用情况付费。由于是计量服务，SaaS允许出租一个应用程序，并计时收费。

2.4 虚拟化系统模型简介

云计算的关键技术是虚拟化，系统虚拟化通过软硬件划分、分时服务、服务质量保证及模型与仿真执行等技术，允许在同一硬件平台上虚拟出多个隔离的硬件 执行环境。这样的虚拟执行环境又被称为虚拟机，它们是对真实计算机环境的复制，用户使用虚拟环境的感受与真实硬件环境相同。用户可以在这些虚拟硬件环境中运行操作系统和各种应用软件。

虚拟机监控器（VMM）:它是运行于虚拟机中的操作系统和底层硬件环境之间附加的一个软件层。虚拟机监控器负责管理底层的硬件资源，并且将这些资源分配给上层运行着的一个或多个虚拟机。虚拟机监控器对于上层的虚拟机和其中的操作系统有着完全的控制权，包括系统状态、占用的资源和执行状况等。操作系统则通过虚拟机监控器与各种硬件资源进行交互。



虚拟机监控器有三大典型的特征：

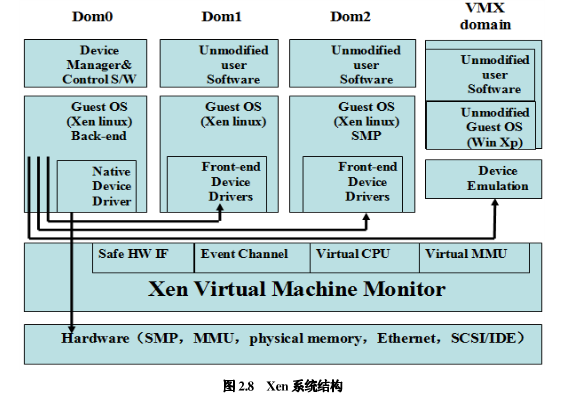
1. 虚拟机监控器提供与所虚拟的计算机环境几乎相同的执行环境。即任何在虚拟机上的执行结果与非虚拟机上的执行结果相同。
2. 虚拟机监控器必须具有对真实系统资源的完全控制。所有运行在虚拟机监控器上的程序访问任何系统资源前，都必须获得由虚拟机监控器显式分配的该系统资源的控制权。
3. 效率的高低也是虚拟机监控器的一个重要特征。指令可以在真实的处理器上执行，也可以在虚拟机监控器模拟的虚拟处理器上执行，但是模拟执行往往要比直接执行慢一些。

Xen最初是由英国Cambridge大学开发，后来得到了Citrix公司的大力支持。Xen是当前比较流行的虚拟化软件，它是基于x86平台的虚拟化实现，能够支持当前流行的通用操作系统。对于上层应用，Xen保证提供与运行在真实硬件环境上的操作系统一致的应用二进制接口（ABI）。

在Xen的整个虚拟化实现架构中存在一个特权管理虚拟机（Privileged Management Domain）,简称为Dom0或Domain 0或VM0，它与其他运行在Xen上的虚拟机不同，可以看作是虚拟机监控器Xen的延伸。

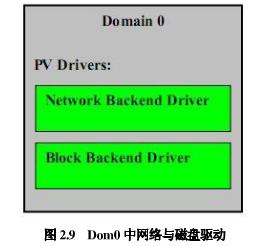
2.5 Xen虚拟化模型

Xen的系统模型如图所示，Dom0是一个修改过的Linux kernel，是唯一运行在Xen Hypervisor之上的虚拟机，它拥有访问物理I/O资源的权限，同时和系统上运行的其他虚拟机进行交互。



Xen Hypervisor是一个介于硬件和操作系统之间的软件层，它负责在各虚拟机之间进行CPU调度和内存分配，Xen Hypervsior不仅抽象出硬件层，同时控制虚拟机的执行，因为这些虚拟机共享同一个处理环境。Xen Hypervsior不会处理网络、存储设备、视频以及其他I/O。

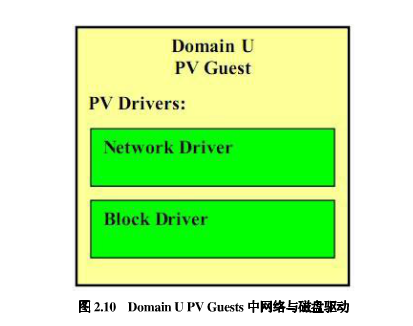
Dom0需要在其他Domain启动之前。Dom0中包含两个驱动：Network Backend Driver 和Block Backend Driver，分别负责处理来自Domain U的网络和本地磁盘请求。Network Backend Driver 直接和本地网络硬件进行通信以处理所有来自Domain U上客户操作系统的网络请求。Block Backend Driver和本地存储设备进行通信以处理来自Domain U的读写请求。



运行在Xen Hypervsior上的所有半虚拟化虚拟机被称为“Domain U PV Guests”，其上运行着被修改过内核的操作系统，如Linux、Solaris、FreeBSD等其他UNIX操作系统。所有的全虚拟化虚拟机被称为“Domain U HVM Guests”，其上运行着不用修改内核的操作系统，如Windows等。

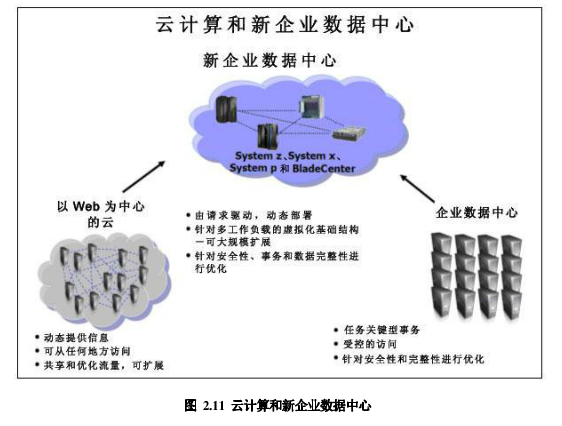
Domain U PV Guests的内核知道它自己不能直接访问物理硬件，因为它知道有其他的虚拟机也运行在同一个环境中。而Domain U HVM Guests的内核不知道它正在和其他虚拟机共享一个处理环境，它认为它运转在物理硬件上。 Domain U PV Guests 包含两个驱动：PV Network 和PV Block Driver。

Domain U HVM Guests虚拟机内没有PV Driver，而是在Dom0 里为每一个HVM Guest启动一个特殊守护进程：Qemu-dm，由Qemu-dm负责客户操作系统的网络和磁盘请求。Domain U HVM Guests必须进行初始化为某类机器，所以要在Domain U 上附加一个软件：Xen虚拟固件，来模拟BIOS。



2.6 云计算与新企业数据中心

新企业数据中心是虚拟化、高效管理的数据中心，它将使用以Web为中心的云所采用的某些工具和技术，并进行一般化以便可由范围更广的客户采用，另外还进行增强以支持安全的事务性工作负载，如图所示。



通过高效且共享的基础架构，企业能够对新的业务需求迅速作出反应，实时解析大量信息，而且还能根据实时数据作出明智的业务决策。新企业数据中心是一种演进的新模型，能提供有助于IT和业务目标保持一致的高效且动态的新方法。从高级别的架构角度来看，新企业数据中心的基础架构服务在逻辑上可以分为不同的层次。物理硬件层之上的两层是虚拟化环境层和管理层，它们是新企业数据中心基础架构服务的关键。通过将这两层结合起来，可以确保数据中心内的资源得到有效的管理，并可以快速部署和配置。新企业数据中心旨在处理混合模式的工作负载。

2.7 云计算的现状和服务案例

最近一年，对云计算和相关基础设施的投资呈爆炸式增长。这样巨大的投资表明，对云中资源的虚拟化存在着巨大的需求。过去一年已经有了很多新的服务，下图展示了其中的一部分。



SaaS就是以服务的形式访问Internet上的软件。一种早期的SaaS方法是Application Service Provider(ASP)。ASP提供对Internet上存放或交付

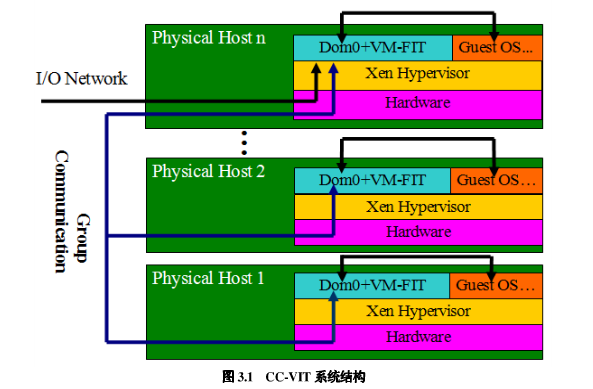
3 云计算平台的容忍入侵技术

3.1 CC-VIT系统模型简介

3.1.1 CC-VIT系统结构

本文采用虚拟化技术解决了云计算平台的容忍入侵性问题，其系统模型如图所示。

Hypervsior是运行于硬件之上的一个极小软件层，也就是虚拟机监控器；在Hypervisor之上划分出不同域，其中服务实例在客户域（Guest Domains）DomU中运行，客户域包括被动模式和活动模式；特权域Dom0控制着客户域的创建、执行与销毁。



副本管理器包括以下部分：

1. 接受客户请求的基础组件；
2. 为所有副本连续转发请求的组通信组件；
3. 实现积极恢复功能的复制逻辑；
4. 实现按需激活被动副本的表决组件；
5. 实现状态更新的相关机制；
6. 实现副本域克隆的组件；

CC-VIT通过Hyervisor来提供通信，复制逻辑运行于特权域Dom0中，负责服务实例的具体工作；而具体的服务副本则运行于隔离的客户域中。客户和服务的交互是通过运行于系统域Dom0中的副本管理来实现。

CC-VIT架构的合理性设计基于以下假设：

1. 客户与远程服务的交互是基于Request/Reply网络层消息，并在网络层被截取。
2. 远程的服务实例能被模拟为确定状态机。
3. 服务副本包括操作系统和执行环境，可能会因为Byzantine错误而失败；在单个恢复周期内允许f<[(n-1)/2]个副本失败。
4. 系统的其他部分，包括Hypervisor和信任的系统域仅仅因为主机的突然宕机而失败。

**3.1.2 CC-VIT系统服务的工作流程**

在CC-VIT中有一个物理机节点接受客户的请求，该节点中副本管理器截取客户的连接请求并通过通信系统分发给副本组中所有的副本。每一个副本执行客户的请求，然后将执行结果返回给一个接受客户连接请求的物理机节点。

当这个物理机收到大量的返回结果后，运行于系统域中的副本管理器通过执行多数表决算法确定一个正确的结果，并将结果返回给用户，整个过程的实现对于用户来说是透明的。复制逻辑具体实现服务副本的实例化和初始化机制。

**3.2 云计算平台容忍入侵关键技术**

**3.2.1 异构混合容错模型**

相互隔离的执行环境允许实施异构的容错容侵模型，也就是说可以在应用域中容忍对应用实例的恶意攻击，如Byzantine错误；在系统域和Hypervisor中可以容忍物理机本身的宕机。在这个异构模型中，运行于隔离的应用域中的服务副本，包括操作系统、中间件结构以及服务的实施可以完全的异构。

单一主机冗余执行RESH(redundant execution on a single host)是CC-VIT模型的一种应用，它允许在单一物理机上冗余执行服务的多个副本，这种冗余执行需要虚拟机监控器创建多个应用域来实现隔离运行副本实例。RESH能够容忍副本中非良性的随机错误，比如无法检测的内存为错误；以及能够容忍N版本编程所产生的软件错误。

多主机冗余执行REMH是CC-VIT模型的另一种应用，它允许相同核心架构的副本可以在多个物理机上冗余执行。与RESH不同的是，在REMH系统域的复制逻辑中集成了一个组通信模块，组通信模块负责一组物理机上副本之间的通信与协调。REMH能够容忍部分运行副本的物理主机突然的宕机。

**3.2.2 活动副本与被动副本**

传统的BFT复制方案需要在每一个副本中执行客户请求。为了减少资源的消耗，我们向CC-VIT系统中引入被动副本的概念。在业务正常运行时至少需要F+1个活动副本发生错误或返回超时，另一个被动副本就接替。要使这样的方法可行，系统架构必须能够允许快速激活被动副本，不然服务对于客户来说将不可用。

在本系统中，驻有被动副本的虚拟机被置于挂起状态，它们仅仅分配了内存、磁盘，并不分配CPU和带宽。当集合被动副本时，副本管理器只需要激活相应的虚拟机就行，这个过程仅需要几百毫秒。一般情况下，本系统容忍F个副本的错误仅需要2F+1个副本，但为了使系统在整个生命周期中容忍无限的副本的错误，则需要在错误副本数突破门限值F时采用积极恢复策略来不断恢复错误副本，因此在实际运行时需要2F+2个副本，多余的一个副本为被动副本，它将用于副本域的克隆和积极恢复。

**3.2.3 状态更新与状态转移策略**

在无状态的复制系统中，从旧副本到新副本的转移过程是相当迅速的；然而在有状态的复制系统中，当创建好一个新的操作系统、中间件以及服务的实例后，新的副本需要根据服务状态进行初始化，这个过程需要状态转移的支持。状态转移增加了副本在积极恢复过程中的服务不可用时间，因为在创建状态检验点时服务的执行会短暂的中断；另一方面状态也会对网络通信资源和CPU资源造成一定的影响。

虚拟化技术可以通过虚拟机监控器来实现磁盘的存储管理。在虚拟化磁盘管理中，虚拟化的永久状态存储可以用来实现高效的多个域之间的状态的重映射机制，这种机制可以实现低成本低开销的状态转移策略。状态转移必须处理不同副本之间的异构性，异构性一般是通过多样性引入的。状态转移时将副本初始状态转移到本地特定副本形式状态需要副本具体实施的支持。

在本系统中，激活被动副本后，为了处理挂起的请求，被动副本也必须具有最近的可用应用状态。为了解决状态的转移和更新问题，本系统采用定期更新的方法来更新被动副本。通常情况下，当执行完一个修改请求后，副本管理器会检索到状态更新，并在本地日志列表中存储每一个状态更新的校验版本。为了保证校验版本的正确性，可以在副本中设置分布式校验点，并且采用检查本地状态检验和的方法来判断状态的有效性。只有本地状态无效时才采用耗费一定资源的远程状态更新方法。当此日志列表的大小达到一定的限制后，副本管理器暂时激活本地被动副本以执行累积的状态更新。当执行完更新后再将此副本置于挂起状态，副本管理器同时裁剪掉相应的日志信息。可以看出，这种定期更新被动副本的方法减少了发生错误时突然更新大量状态的负载，因为当激活被动副本来容忍错误时，只有最近一次状态更新之后的状态才需要更新执行。

3.2.4 积极恢复策略

传统的容忍入侵系统一般采用Byzantine错误容忍复制算法来实现在N个副本中容忍有限的F个副本的错误。这种方法存在一个潜在的问题，就是只要给定了足够的时间，攻击者就能够攻破F个副本，从而突破系统能够容忍的门限值F，最终使容忍入侵系统失效。解决这个问题的办法就是使副本周期性地从一个安全的代码库重新初始化，也就是一种积极恢复策略。积极恢复策略能够移除一些潜在的攻击。如果采用解决恢复策略，只要在恢复周期内错误的副本不会突破门限值F，系统就能够在整个系统生命周期中容忍无限的副本的错误。

在传统的积极恢复策略中只允许部分副本的同时重启以避免服务的不可用性；如果采用虚拟化技术，可以实现所有副本同时的恢复而对服务的执行几乎不产生什么的影响。由于任何虚拟化技术都提供了一种创建和销毁域的核心功能，因此通过这种机制可以实现一种高效的积极恢复策略。周期性恢复策略可以在隔离的系统域中通过恢复服务来触发。为了减少系统服务的不可用时间，可以采用一种并行的恢复策略，也就是在创建新域的同时，允许另外一个先前的域继续执行。

在采用虚拟化技术时，复制逻辑运行于一个隔离、无入侵的系统域里，因此可以将复制逻辑当成一个可信实体来重新初始化应用域，而不需要任何特殊硬件的支持。恢复策略将一个完整副本初始化到一种“干净”状态，然后通过错误容忍状态转移协议安全地获得其他副本的服务状态，随后转移到相应的服务状态执行系统分配的任务。Hypervisor能够关闭和启动运行于虚拟机中的副本，在实现恢复策略时，新的副本实例可以并行于当前执行的副本启动，采用的方法是在另外一个虚拟机中启动。在积极恢复时这种方法可以大量地降低服务不可用时间，因为重启进程不会影响到副本的可用性。当初始化完一个新的副本后，副本管理器就可以关闭先前运行的副本，并同时激活刚初始化完毕的新的副本。

然而基于虚拟化的积极恢复也具有以下不足之处：积极恢复通常对服务有一定的负面影响，因为本地硬件资源有限，但是这些有限的资源被所有的副本域共享。在进行积极恢复时难免会涉及到新域的创建和旧域的销毁，因此对资源的消耗是难免的。但在另一方面来说，这种积极恢复策略显著降低了副本在新旧状态间转移时服务的不可用时间。

实际上在本系统中，由于被动副本具有最近的可用应用状态，因此被动副本可以用作积极恢复的基础。当错误副本数即将突破门限值F时积极恢复被触发，在本系统中，可以通过克隆被动副本的方式快捷创建新的具有最近可用状态的副本域，然后通过状态转移获得最近一次状态更新之后的状态。状态更新和状态转移完成后，在副本管理器的控制下，当前激活的更新副本开始执行先前挂起的请求，并且关闭先前错误的活动副本可以用做积极恢复的基础。这样，干净状态的副本就能高效而连续的接替错误副本的运行，并能有效降低积极恢复对系统平台的突发负面影响。

3.2.5