RSA CONFERENCE CHINA 2011
2011 信息安全国际论坛

the adventures of

云计算环境下的数据安全问题

演讲人：李舟军

职务：信息安全系主任

单位：北京航空航天大学

报告提纲

- 网络应用现状及其安全问题
- 云计算及其安全问题
- 云计算环境下的数据安全问题
- 可搜索加密方案
- 外包数据库的访问控制
- 全同态加密方案
- 结束语

网络应用现状及发展趋势

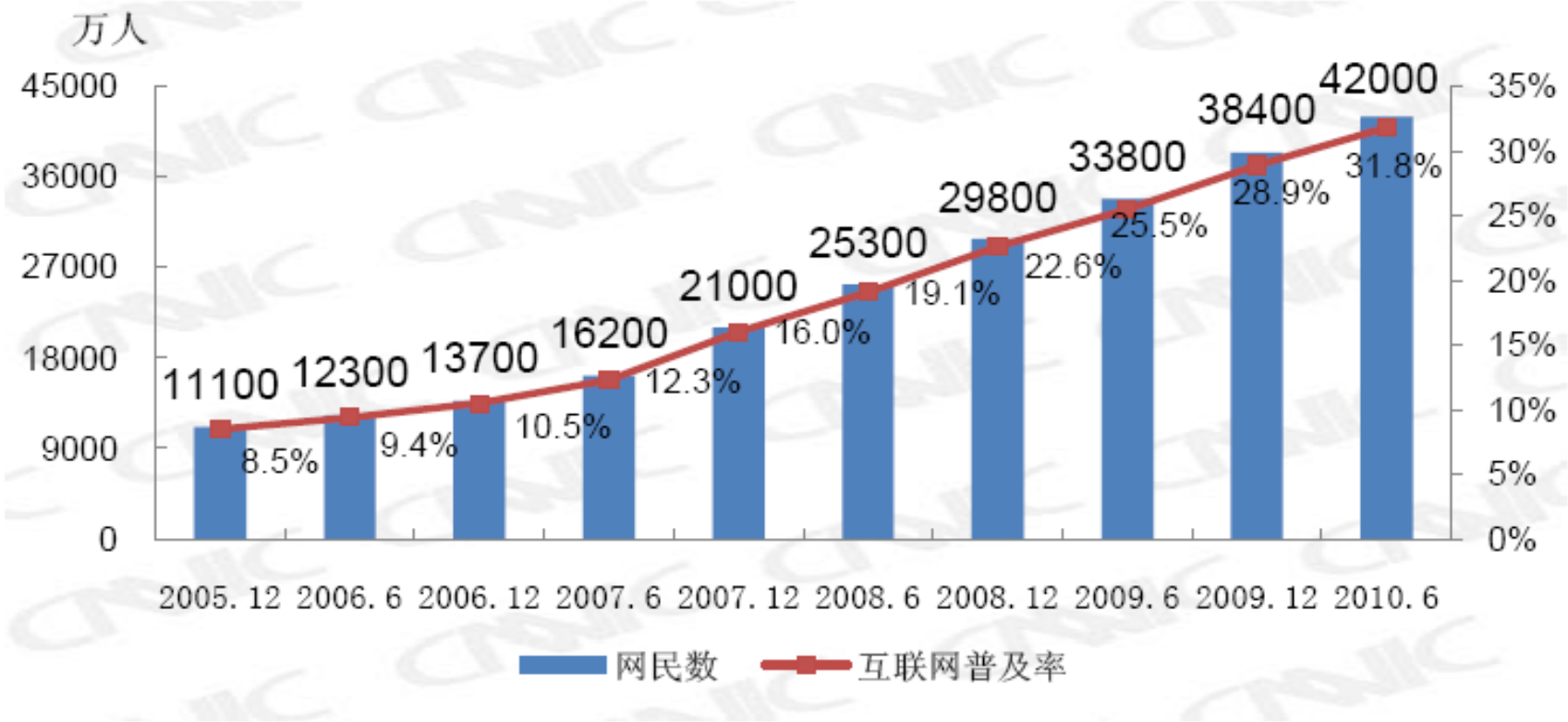
- 互联网是人类智慧的结晶，20世纪的重大科技发明，当代先进生产力的重要标志。它深刻影响着世界经济、政治、文化和社会的发展，促进了社会生产生活和信息传播的变革。
- 互联网在中国的发展速度超乎人们想象，中国已成为世界第二大IPv4地址拥有国，互联网发展与普及水平居发展中国家前列。
- 互联网全面影响并改变着现代国人的生活，已成为人们日常生活中不可缺少的工具，融入到学习、工作和生活的每个细节，让现代人的生活变得更加精彩纷呈：

  看新闻、听音乐、学知识、玩游戏、看影视，聊天、偷菜、淘宝、购物，写博客、写微博，公开表达民意……

  真可谓： 随心所欲， 一网打尽

网络应用现状及发展趋势

- 截至2010年6月，国内网民规模达到4.2亿，突破了4亿关口，较2009年底增加3600万人。互联网普及率升至31.8%，较2009年底提高2.9个百分点。



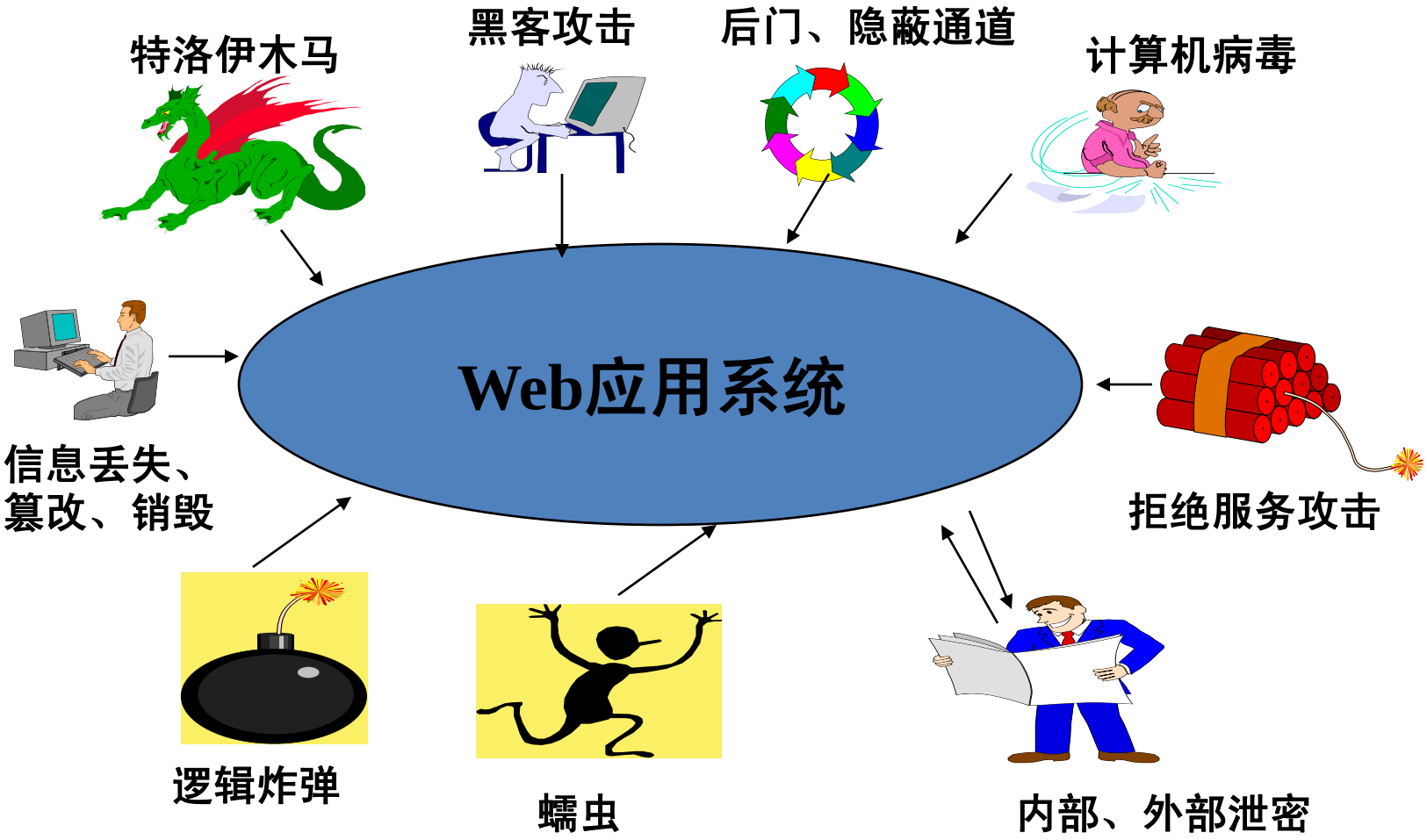
商务应用快速发展

2010年上半年，我国网民的互联网应用表现出如下特点：

- 商务化程度迅速提高
- 娱乐化倾向继续保持
- 沟通和信息工具价值加深
- 商务类应用表现尤其突出：

截至2010年6月底：网络购物、网上支付和网上银行的使用率分别为 33.8%、30.5%和29.1%，用户规模分别达到1.42亿、1.28亿、1.22亿，半年用户规模增幅分别为31.4%，36.2%和 29.9%，增速在各类网络应用中排名前三。

网络应用的安全问题


特洛伊木马
黑客攻击
后门、隐蔽通道
计算机病毒
Web应用系统
信息丢失、
篡改、销毁
拒绝服务攻击
逻辑炸弹
蠕虫
内部、外部泄密

网络应用的安全问题

- 网络安全成为网络应用的最大制约因素。

仅2010年上半年：

- 有59.2%的网民在使用互联网过程中遇到病毒或木马攻击；
- 30.9%的网民账号或密码被盗过；
- 电子商务网站访问者中89.2%的人担心假冒网站，其中，86.9%的人表示如果无法获得该网站进一步的确认信息，将会选择退出交易。
- 网络安全和信任问题严重影响和阻碍网络应用和电子商务向更广和更深层次发展（安全协议并不一定安全!）。

网络应用安全威胁总体大趋势

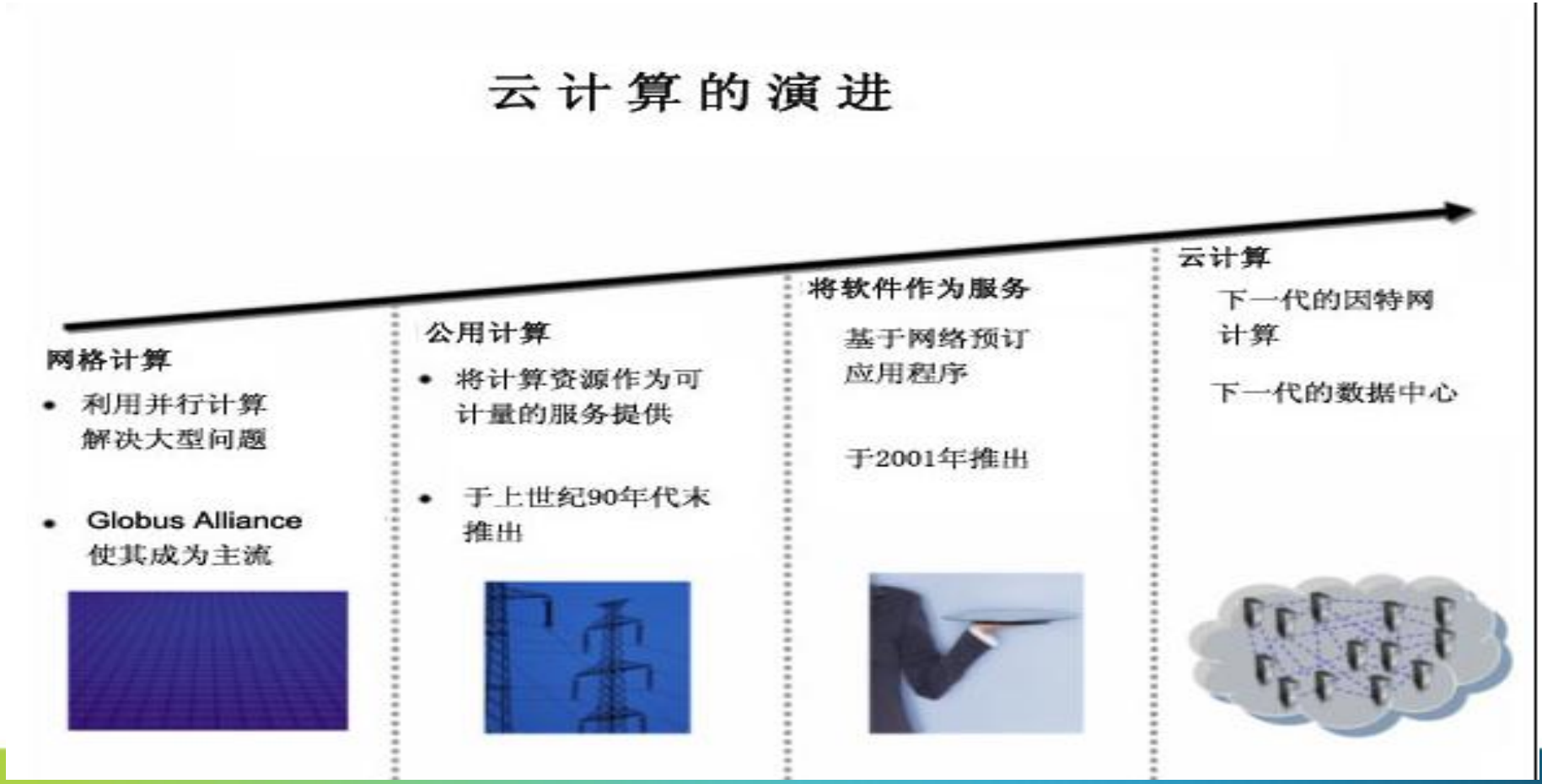
- 安全威胁正从网络层向应用层全面发展
- 攻击者从炫耀技术（损坏电脑）转向追求经济利益
- 攻击转向基础设施以窃取重要数据（用户资料、密码、金钱）
- Web2.0、社交网络已迅速成为网络犯罪的温床，而网络钓鱼成为其安全的第一杀手
- 互联网上的几乎所有安全威胁将以新的面貌、甚至更严重的程度重现于移动互联网（手机）上
- 被病毒”感染“的木马和蠕虫数量呈现上升趋势。这些”混血战士”同时具备病毒的高感染性、“抗击打能力”和木马的战术特性
- 控制、物理层面（震网病毒肆虐-->工业控制系统的安全性）

报告提纲

- 网络应用现状及其安全问题
- 云计算及其安全问题
- 云计算环境下的数据安全问题
- 可搜索加密方案
- 外包数据库的访问控制
- 全同态加密方案
- 结束语

云计算

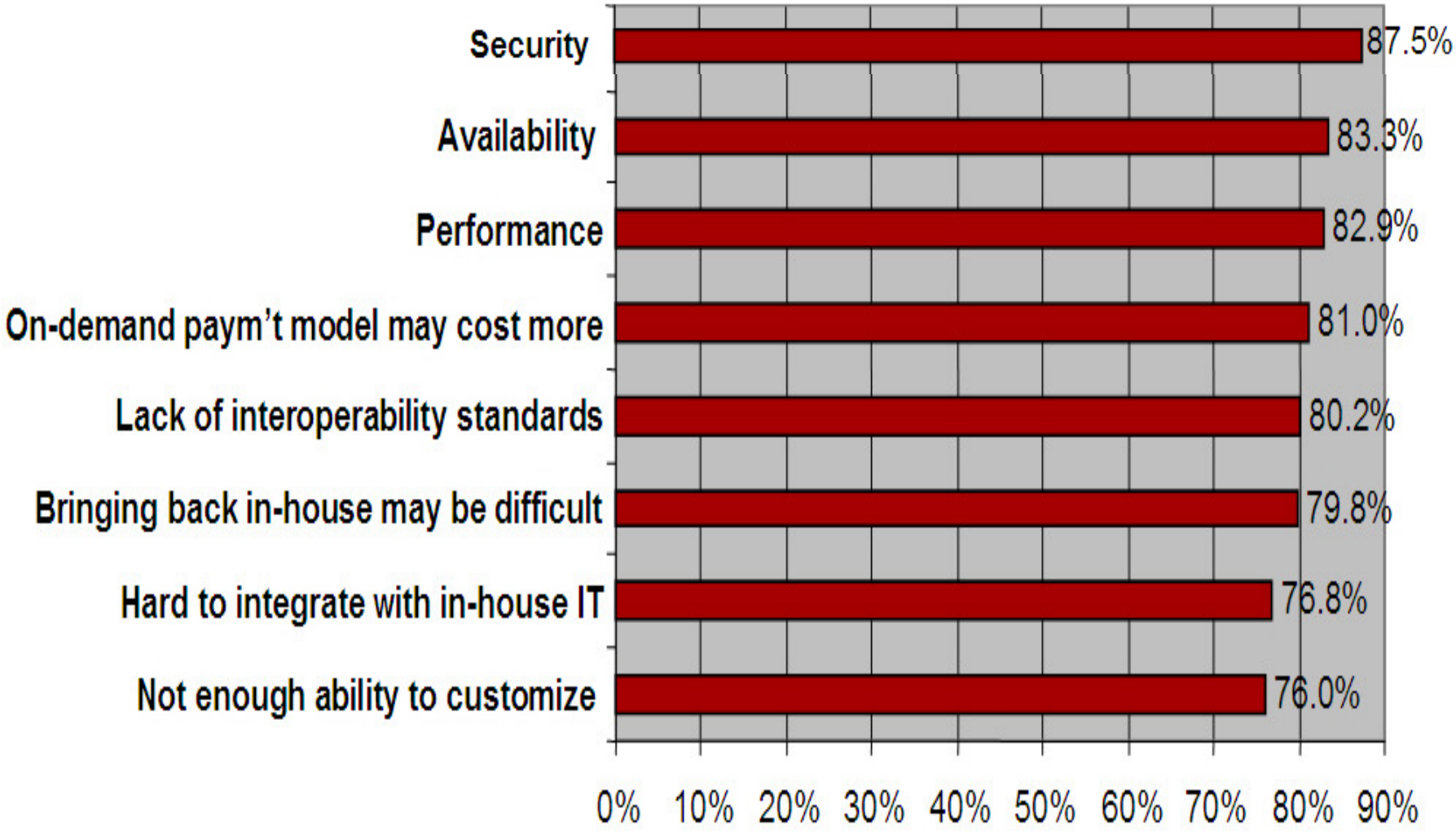
- 云计算（Cloud Computing）是网格计算、分布式计算、并行计算、效用计算、网络存储、虚拟化、负载均衡等传统计算机技术和网络技术发展融合的产物。



云计算

- 云计算旨在通过网络把多个成本相对较低的计算实体整合成一个具有强大计算能力的完美系统，并借助SaaS(软件即服务)、PaaS(平台即服务)、IaaS(基础设施即服务)、MSP（管理服务提供商）等先进商业模式把这些强大的计算能力分布到终端用户手中。
- 云计算的核心理念：通过不断提高“云”的处理能力，进而减少用户终端的处理负担，最终使用户终端简化成一个单纯的输入输出设备，并能按需享受“云”的强大计算处理能力。
- 云计算以其便利、经济、高可扩展性等优势得到产业界、学术界、政府等各界的广泛关注和高度重视，被认为是信息技术领域的下一个重要增长点，具有巨大的市场前景。

云计算的最大问题是安全



云计算技术存在的7大风险

- 研究机构Gartner近日发布一份名为《云计算安全风险评估》的报告，列出了云计算技术存在的7大风险：
- 1.特权用户的接入

  在公司外的场所处理敏感信息可能会带来风险，因为这将绕过企业IT部门对这些信息“物理、逻辑和人工的控制”。企业需要对处理这些信息的管理员进行充分了解，并要求服务提供商提供详尽的管理员信息。
- 2.可审查性

  用户对自己数据的完整性和安全性负有最终的责任。传统服务提供商需要通过外部审计和安全认证，但一些云计算提供商却拒绝接受这样的审查。面对这样的提供商，用户只能用他们的服务做一些琐碎的工作。

云计算技术存在的7大风险

- 3.数据位置

在使用云计算服务时，用户并不清楚自己的数据储存在哪里、甚至哪个国家。用户应当询问服务提供商数据是否存储在专门管辖的位置，以及他们是否遵循当地的隐私协议。

- 4.数据隔离

云计算将所有用户的数据置于共享环境之中。用户应当了解云计算提供商是否将一些数据与另一些隔离开，加密服务是否可靠。如加密系统出问题，则所有数据都不能再用。

- 5.数据恢复

云计算提供商应当告诉用户：发生灾难时，用户数据和服务将会面临什么情况，服务提供商是否有能力恢复数据，以及需要多长时间。任何没有经过备份的数据和应用程序都将出现问题。

云计算技术存在的7大风险

- 6.调查支持

  在云计算环境下，调查不恰当的或是非法的活动将难以实现，因为来自多个用户的数据可能会存放在一起，并且有可能会在多台主机或数据中心之间转移。如果服务提供商没有这方面的措施，那么在有违法行为发生时，用户将难以调查。

- 7.长期生存性

  理想情况下，云计算提供商将不会破产或是被大公司收购。但是用户仍需要确认，在发生这类问题的情况下，自己的数据会不会受到影响。用户需要询问服务提供商如何拿回自己的数据，以及拿回的数据是否能够被导入到替代的应用程序中。

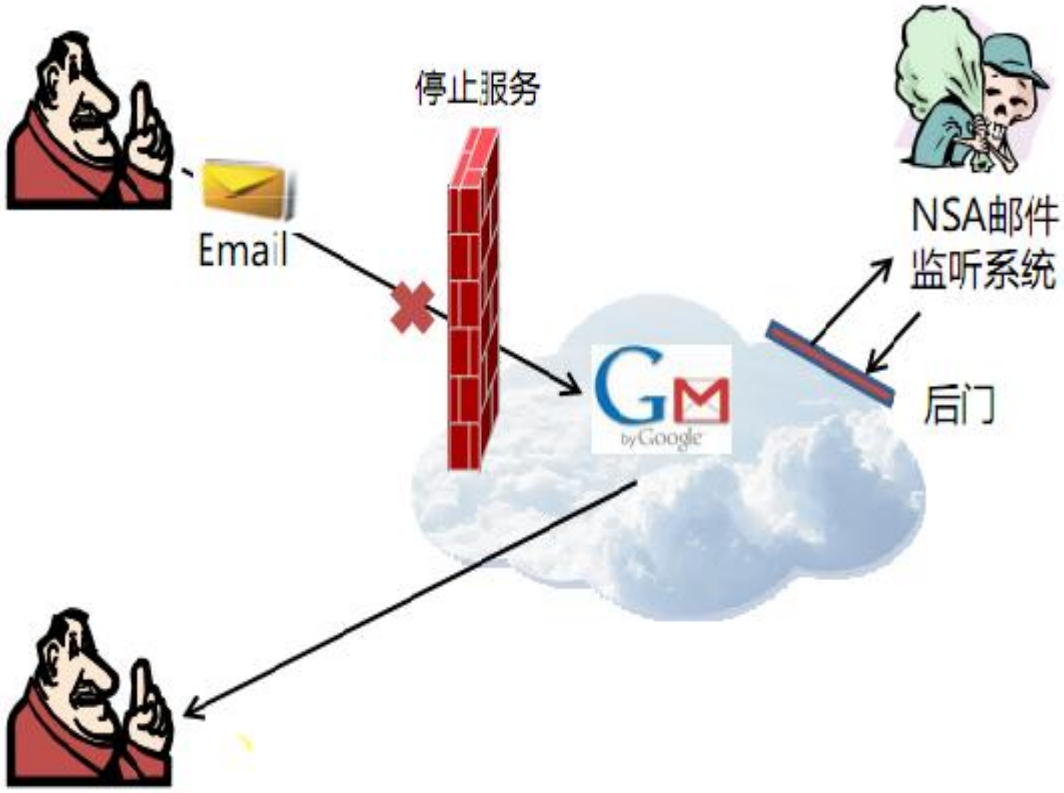
云计算安全事件

随着云计算技术推广和应用，其安全问题逐渐暴露。云计算的倡导者和领军者Google，Amazon，Google 等相继爆出各种安全事故：

- 2009年2月，Google Gmail爆发全球性故障，服务中断长达4小时；
- 2009年3月，Google发生大批用户文件外泄事件，因Google的疏忽导致用户保存在Google Docs的部分文档会在用户不知晓的情况下被共享；
- 2009年3月，微软的云计算平台Azure停止运行约22个小时；
- 2009年2月和7月，亚马逊的“简单存储服务(simple storage service)”两次中断（ 7月断网长达6小时），导致依赖于网络单一存储服务的网站被迫瘫痪。
- Gartner 2009 年调查结果显示：74%受访企业的CTO表示近期不会采用云计算，其首要原因在于：担忧云计算环境中数据的安全性与隐私性问题。

云计算安全隐患

截止2008年底，仅Google的邮件服务Gmail，在中国大约拥有10%的市场占有率，拥有2000万左右的中国用户。



美国国家安全局根据《外国情报监视法》要求，在Gmail邮件系统中安置后门，以便其对往来的电子邮件进行监视，甚至能够强制对某些特定群体用户停止服务，或者拦截重要邮件！

报告提纲

- 网络应用现状及其安全问题
- 云计算及其安全问题
- 云计算环境下的数据安全问题
- 可搜索加密方案
- 外包数据库的访问控制
- 全同态加密方案
- 结束语

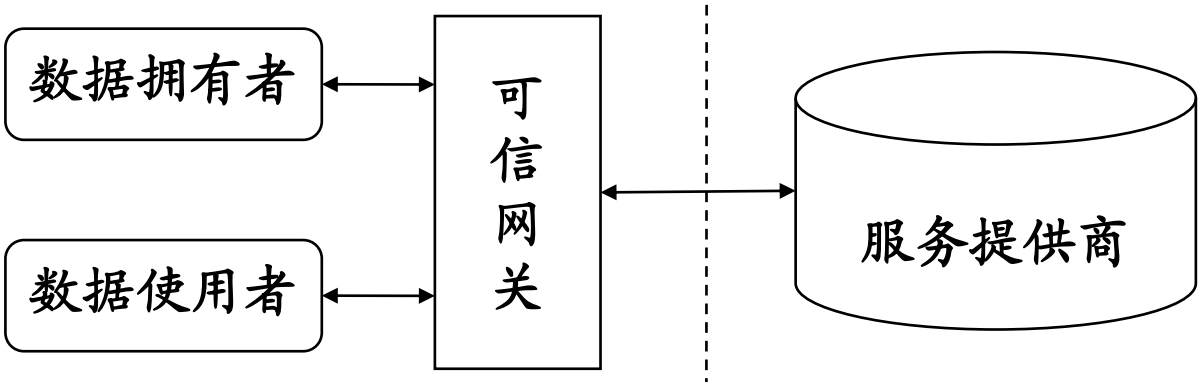
云计算环境下的数据安全问题

- 综上所述，云计算的数据安全问题已成为制约其发展的首要因素。因此，要让企业和个人用户普遍使用云计算技术与平台，放心地将自己的数据交付于云服务提供商管理，就必须首先致力于解决云计算所面临的各种安全问题。
- 但云计算的服务计算模式、动态虚拟化管理方式、多租户共享运营模式等给数据的安全与隐私保护提出了严重的挑战，用户数据的安全与隐私保护需求是云计算产业发展无法回避的核心问题。

云计算环境下的数据安全问题

- 云计算环境中，企业和个人用户一般将自己的数据存放在云计算平台，委托云服务提供商完成数据的存储和管理等复杂任务，而用户只需通过网络获得相应的服务。
- Hakan Hacigumus于2002年在云计算模式下首次提出了“外包数据库（Databases as a Service，DaaS）”的概念。通过外包数据库的方式，企业降低了总成本，且可以把精力放在其核心业务上。
- 对外包数据库的使用者来说，DaaS 如同一个支持逻辑数据运算和逻辑数据存储的黑箱，而顾客只能看到组织中的数据。

云计算环境下的数据安全问题



数据库服务模型

1）数据拥有者：数据拥有者是一个实体，它产生并拥有数据。通常假设数据拥有者的计算资源和存储能力比服务器要弱很多。

2）服务提供商：服务提供商即外包数据库，是一个远程服务器，用于存储和管理由数据拥有者产生的数据。

3）数据使用者（用户）：数据使用者也可以是数据拥有者；如果数据拥有者是一个组织，那么数据使用者也可以是员工或客户。

云计算环境下的数据安全问题

- 在DaaS模型中：

  数据拥有者 将其数据存储在服务提供商端

  数据使用者 可以查询其需要的信息

  服务提供商 负责数据存储和管理、响应用户查询

  可信网关　负责数据加密、解密和查询完整性检验等工作

- 由于外包服务器是不可信第三方，数据拥有者的敏感数据不仅要对外包服务器保密，其解密密钥也不再和外包服务器共享。数据拥有者必须先将敏感数据加密，然后再存储在外包服务器上。
- 这样，存储在外包服务器上的加密数据不仅不能被外部未授权用户访问，也不能被服务提供商未授权访问，这就要求加密的数据不能在外包服务器端进行解密处理。

云计算环境下的数据安全问题

- 因此，直接对存储在外包服务器中的加密数据进行查询、修改等操作，就显得尤为重要。
- 例如，对google 和Yahoo！之类为用户提供免费电子邮箱等服务的在线服务提供商，应该能够在不获取任何有关原文信息的前提下，为用户提供邮件收取、邮件检索、垃圾邮件过滤和邮件删除等功能。
- 如果允许对密文操作，那么就可以在不需解密的情况下处理对搜索引擎所做的加密查询，从而在访问信息时有效地保证了数据的隐私性。

云计算环境下的数据安全关键问题：

1）在公钥加密体制下，研究加密外包数据库中对密文数据的操作问题以及密钥的分发等问题，使得可以在保障数据安全性的前提下，可对数据进行一些常见的查询、修改等操作。

2）外包数据库的访问控制与密钥管理问题，需要一个具有全安全性的基于属性的分层加密方案，不仅支持细粒度的访问控制，而且还满足高性能、完全授权、可扩展等需求。

3）如何提高全同态加密方案的效率，使得无须对加密数据进行解密，即可利用全同态加密方法直接对存储在外包服务器中的加密数据进行数据运算以及查询、修改等操作。

报告提纲

- 网络应用现状及其安全问题
- 云计算及其安全问题
- 云计算环境下的数据安全问题
- 可搜索加密方案
- 外包数据库的访问控制
- 全同态加密方案
- 结束语

可搜索加密方案

对密文进行操作的方法主要有：带关键字搜索的公钥加密、公钥混淆加密、以及私有信息检索等方法。

1） 带关键字搜索的公钥加密方案

- 带关键字搜索的对称加密方案虽然具有简单、高效的优点，但需要在数据使用者和服务器之间共享一个私钥，因此无法实现对第三方数据的秘密搜索。
- 第一个带关键字搜索的公钥加密体制（PEKS scheme）由 Boneh 等人提出，最初被用于加密的电子邮件系统中。该方案的检索效率很低。此外，该方案由于需要数据使用者和服务器之间共享一个安全的通道，其实用性也较差。

可搜索加密方案

- 2005 年，Baek 等提出一种不需要安全信道的PEKS系统（Secure-Channel-Free PEKS），该系统通常称为 指定验证方的PEKS。
- 2007 年，C. Gu等对其进行了改进，用PEKS 加密关键词时省去了对运算，在一定程度上提高了效率。
- 2009 年，Rhee 等人进一步提出了一个增强的安全模型。之后，Byun 等人提出了关键字猜测攻击的概念，他们将如何构造一个可以抵抗关键字猜测攻击的指定验证方的PEKS 作为一个公开问题。
- Rhee等提出了一个安全模型并给出了构造，从而解决了这个公开问题。
- 但之前所有的PEKS 方案的安全性均在随机预言机（Random Oracle，RO）模型下证明，不能反映这类方案在现实生活中的安全性。Zhang等人构造一个指定验证方的PEKS 方案，该方案的安全性证明不再依赖于RO 模型。

可搜索加密方案

2）公钥混淆加密

- Ostrovsky 和Skeith 提出了公钥混淆（Public-Key Obfuscation）的概念，通过加密程序来达到混淆的目的，即通过运行一个加密程序来处理一个加密的输出，之后再对加密输出进行解密。
- 由于云服务平台上可以执行加密的程序，而云服务器无法获知关于程序本身的任何信息，因此公钥混淆密码学可用于解决云计算环境下“平台即服务（Platform-As-a-Service）”中存在的安全问题。但只有在可以执行高效实用的密文处理运算时，上述方法才可用于解决“数据库即服务”中存在的安全问题。

3） 私有信息检索

- 私有信息检索（Private Information Retrieval，PIR）允许一个用户在保持询问隐私的情况下从数据库中检索信息，但这种方案的通信代价比较大。因此，只有在数据使用者执行很多计算，通信复杂性仍然保持很小时，私有信息检索才被认为是可接受的。

高效的密钥分发方案

- 如何高效地分发解密密钥是外包数据库密钥管理问题中的一个研究热点和难点。
- 一个基本的解决方法是：在数据拥有者发送多个解密密钥之前，先将他们压缩为一个解密密钥，之后通过一个安全信道将压缩的解密密钥发送给授权的数据使用者。数据使用者收到后，使用该解密密钥及其他信息恢复所需的明文。
- 目前关于数据压缩的研究工作有很多，如：聚合签名中的签名规模、广播加密中的密文规模、对称密码学体制下的“多用途密钥” 等。
- 相关的研究工作还有：

1） 预先定义层次的密码学密钥分配方案

2） 对称密码学加密中的压缩

3） IBE 中的压缩密钥

报告提纲

- 网络应用现状及其安全问题
- 云计算及其安全问题
- 云计算环境下的数据安全问题
- 可搜索加密方案
- 外包数据库的访问控制
- 全同态加密方案
- 结束语

外包数据库的访问控制问题

- 云计算的访问控制是在外包服务器端不可信的情况下保障数据机密性的解决方案。
- 数据拥有者在存储数据之前预先对其进行加密，通过控制用户对解密密钥的获取权限来实现访问控制的目标。
- 当用户在外包服务器上共享机密数据时，所采用的加密系统不仅要支持细粒度的访问控制，而且还要满足高性能、完全授权、可扩展等需求，从而可以随时随地为使用不同设备的客户提供最佳访问数据的服务，并且能够在企业内部独自进行访问授权、用户访问权限撤销等操作。

外包数据库的访问控制问题

云计算中访问控制的两种方案：

- 分层访问控制方法是一种经典的密文访问控制方法，云计算中的很多访问控制技术都是基于分层访问控制方法的。但分层访问控制方法中，Token表完全由数据拥有者维护且具有复杂的结构，每次用户权限的变更都必须对相应的Token表进行复杂的操作，因此效率低下。
- 另一种方法是双层加密的技术，在外包服务器端进行一层额外的加密保护，通过更改外包服务器层的密钥以实现不同的访问控制策略。实际上，云计算环境中的外包服务器完全可能因为利益的关系与非法的用户共谋，造成数据的泄露。

外包数据库的访问控制问题

- 随着用户对个人隐私的关注，访问控制系统需要对客户的隐私加以保护，例如，在某些环境中，访问策略本身就是敏感信息。

- 因此在云计算环境中，还需要为客服提供匿名的访问控制，以满足客户对个人隐私保护的需求。

报告提纲

- 网络应用现状及其安全问题
- 云计算及其安全问题
- 云计算环境下的数据安全问题
- 可搜索加密方案
- 外包数据库的访问控制
- 全同态加密方案
- 结束语

全同态加密

能否直接对存储在外包服务器中的加密数据进行运算和操作？

- 在云计算环境下，为了保护用户的隐私及数据安全，需要先对数据加密，再把加密后的数据放到云服务器端。
- 用户对外包数据处理时必须频繁地存取和加解密数据，这不仅极大地增加了外包服务器端和用户端的通信和计算开销，而且也难以保证数据在处理过程中的安全性。
- 使用全同态加密算法，使得在不暴露数据的情况下，数据使用者通过云服务器不仅可以直接对加密数据进行各种运算，而且也可对加密数据进行查询或修改等操作，并将符合条件的加密数据返还给数据使用者，之后数据使用者即可使用对应的解密密钥对收到的加密数据进行解密。

全同态加密

设加密操作为 E，明文为 m，相应密文为 e，即e = E(m)。

若对明文操作 f，可构造操作F，满足F(e) = E(f(m))，即：

F(E(m)) = E(f(m))，则称 E 为一个针对 f 的同态加密算法。

若对任意复杂的明文操作f，都能构造出相应的 F，则称E为全同态加密（Fully Homomorphic Encryption）算法。

目的是找到一种能在加密的数据上进行任意数量的加法和乘法运算的加密算法,使得对加密数据进行某种操作所得到的结果恰好等于对加密前的数据进行预期的操作后所得到的密文。

全同态加密

- 早在1978 年, R. Rivest、L. Adleman 和M.Dertouzos就提出了“全同态加密” 的概念。
- 随后，密码学家们对此进行了艰苦的探索，但只能部分解决这一问题。例如，RSA 算法对于乘法运算是同态的，但对别的运算（如加法）就无法构造出对应的F。而Paillier 算法则是对加法运算是同态的。
- 能实现全同态加密的理论上的解决方案于2009 年被C.Gentry设计出来。该方案是基于理想格构造的，可看作一种特殊的公钥密码体制。但该方案计算效率很低，离真正实用还有很大一段距离。

全同态加密

- 2010 年, N. P. Smart 等 借鉴C. Gentry构造全同态加密方案的思想，选定两个大整数组成公钥和私钥，一个大整数组成密文，给出了基于相对小的密钥和密文规模的全同态加密方案。
- M. van Dijk等人用整数集代替理想，设计了全同态加密方案，把此方案的安全性问题归结到找一个近似的最大公约数－即给出一系列是某个隐整数的近似倍数的整数，找出此隐整数。与C. Gentry设计的用多项式环和其中的理想格设计的方案相比，其优点是更简洁，但效率依然很低。
- 2011 年，J. Loftus 等人对N.P. Smart 等构造的全同态加密方案进行改进，得到了满足不可区分非适应性选择密文攻击（IND-CCA1）安全性的全同态加密方案。
- 目前，设计不可区分适应性选择密文攻击（IND-CCA2）安全性的全同态加密方案是很有意义的工作。

全同态加密

- 全同态加密为直接对外包服务器中的加密数据进行运算和操作提供了可能。用户对外包数据处理时就不必频繁地存取和加解密数据，极大地减少外包服务器端和用户端的通信和计算开销，也保证了数据处理过程中的安全性。
- 利用全同态加密技术，对加密数据的分析能得到与原始数据同样细致的结果。在保持用户数据隐私性的同时，为分析和挖掘服务提供商所存储的海量数据开辟了无限的商机。

例如：

1）云计算服务提供商就能接受用户委托，在不暴露原始数据的前提下充分分析其他用户的数据；

2）可把加密了的个人工资和支出情况交给网上纳税服务系统来处理；

3）能帮助计算机用户从搜索引擎中提取信息，而不必让搜索引擎了解具体请求内容；

4）也用于对加密的电子医疗档案进行分析

结束语

虽然目前该领域的研究已取得重要进展，但仍然存在以下一些问题：

1）已有的带连接关键词的可搜索公钥加密方案，需要服务器和用户之间存在一个安全的信道，不能用于外包数据库的场景。

2）目前已有的可搜索公钥加密方案只能在随机预言机模型下证明其安全性，期望设计的方案能在标准模型下是可证明安全的。

3）云计算环境下海量数据的加密存储需要相对应规模的解密密钥，因此传统方法需传送大量的私钥,并需建立相应规模的安全通道来传递私钥。已有的压缩方法，要么只能在对称密码体制下实现，要么计算效率和通信效率很低。

4）已有的基于属性的分层加密方案仅在随机预言机模型下证明其语义安全性，期望设计的方案能在随机预言机模型下具有全安全性。

5）目前的全同态加密方案计算量巨大，难以在现有计算技术条件下实现。如何提高全同态加密方案的加解密效率，降低密钥存储空间；如何设计满足IND-CCA2 安全性的全同态加密方案，是研究难点。

谢谢！