目录

1　大数据隐私保护技术

2　云安全概述

3　云安全技术

大数据的特点（IBM观点）

- 量大(Volume)：TB→PB→EB→ZB
- 样多(Variety)：类型多样，非结构化为主
- 价值(Value)：价值密度低，价值总量高
- 高速(Velocity)：生成和处理速度快
- 真实(Veracity)：数据的质量

大数据的用途

从大数据中寻找到背后隐藏的意义，发现事先未能想到的关系、有意思的联系。迅速分析大量，复杂的非结构化数据，来解决个人、产业，非商业组织，政府之间的非线性关系。

12306网站

- 2014年，12306网站被爆泄露大量用户信息；
- 2018年12月，互联网安全新媒体Freebuf爆料12306疑似再次发生数据泄露；
- 数十万用户姓名、密码、电话、邮箱、家庭住址等高度敏感的信息泄露，数据流入黑市（一份包括60万账户信息的文件标价仅为20美元）。

酒店信息泄露

- 2013年, 7天酒店系统漏洞，管理权限被获取，内部数据泄露；
- 2014年, 乌云平台披露多个酒店网站系统漏洞，2000万用户入住数据在网上疯传；
- 2015年，乌云平台披露携程机票系统信息泄露，可导致系统瘫痪；
- 2018年，万豪酒店被曝发生史上最严重的数据泄露事件之一，市值蒸发超250亿美元。

棱镜门事件

- **棱镜计划**

棱镜计划（英语：PRISM）是一项由美国国家安全局（NSA）自2007年开始实施的绝密级网络监控监听计划。

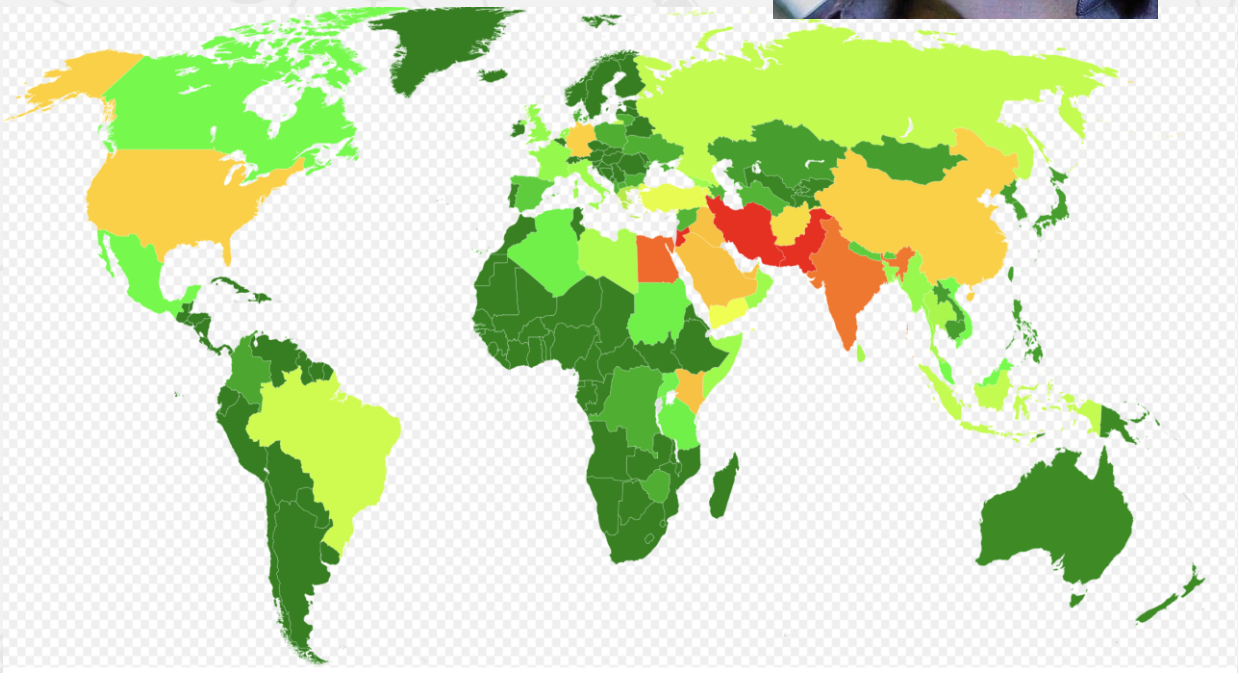
- **实施细节**

泄露的文件中描述PRISM计划能够对即时通信和既存资料进行深度的监听。许可的监听对象包括任何在美国以外地区使用参与计划公司服务的客户，或是任何与国外人士通信的美国公民。国家安全局在PRISM计划中可以获得电子邮件、视频和语音交谈、影片、照片、VoIP交谈内容、文件传输、登录通知的数据，以及社交网络细节，并透过各种联网设备，如智能手机、电子式手表等各式联网设备对特定目标进行攻击。

- **爱德华•斯诺登**

泄露这些绝密文件的是国家安全局合约外包商员工爱德华·斯诺登，于2013年6月6日在英国《卫报》和美国《华盛顿邮报》公开。


PRISM



西工大遭受邮件攻击事件

- **攻击事件源起**

  2022年4月12日该校电子邮件系统发现一批以科研评审，答辩邀请和出国通知等为主题的钓鱼邮件，内含木马程序，引诱部分师生点击链接，非法获取师生电子邮箱登录权限，致使相关邮件数据出现被窃取风险。同时，部分教职工的个人上网电脑中也发现遭受网络攻击的痕迹。

- **攻击实施主体TAO（特定入侵行动办公室）**

1. NSA旗下，成立于1998年，是目前美国政府专门从事对他国实施大规模网络攻击窃密活动的战术实施单位；
2. NSA针对西北工业大学的攻击行动代号为“阻击XXXX”（shotXXXX）。

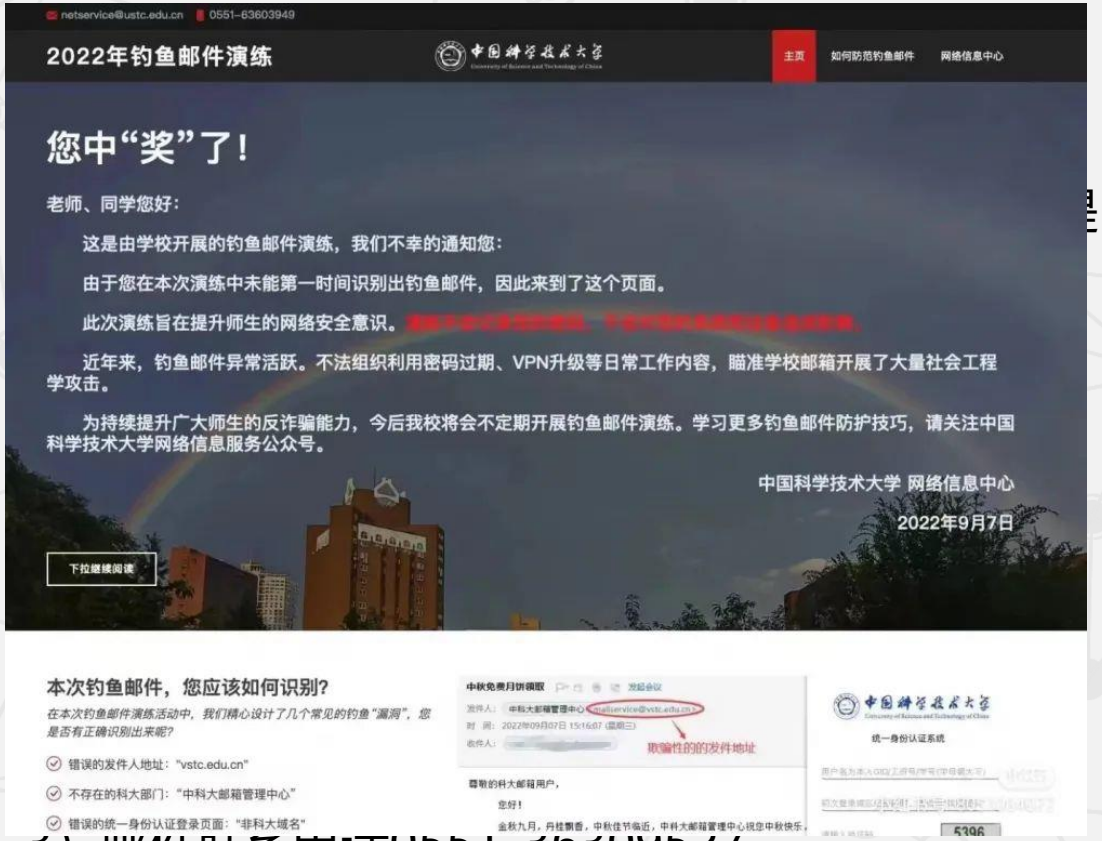
- **此次攻击特点分析**

1. 掩盖真实 IP 精心伪装网络攻击痕迹
2. 使用了种类繁多功能各异的专用网络攻击武器装备

中科大发4万封“免费送月饼”钓鱼邮件

- **中科大反诈演练**

在一个小时发送了4万多邮件，这些邮件全部针对



3）邮件联系电话0551-36309527

为假，中科大电话是以6360开头。

勒索病毒?

- 勒索病毒是一种黑客通过技术手段将受害者机器内的重要数据文件进行加密，最终迫使受害者向黑客缴纳文件解密赎金，黑客收到赎金后，进一步协助受害者恢复被加密数据，从而达到病毒非法牟利勒索钱财的目的。

勒索病毒传播方式

内网传播

病毒会根据用户计算机内网IP，生成覆盖整个局域网网段表，然后循环依次尝试攻击。

外网传播

勒索病毒会随机生成IP地址，尝试发送攻击代码。

勒索病毒主要传播方式

弱口令攻击

口令爆破攻击依然是当前最为流行的攻击手段，使用过于简单的口令或者已经泄露的口令是造成设备被攻陷的最常见原因。

钓鱼和垃圾邮件

“钓鱼邮件”攻击是最常见的一类攻击手段，在勒索病毒传播中也被大量采用。通过具有诱惑力的邮件标题、内容、附件名称等，诱骗用户打开木马站点或者带毒附件，从而攻击用户计算机。

利用系统与软件漏洞攻击

漏洞攻防一直是安全攻防的最前沿阵地，利用漏洞发起攻击也是最常见的安全问题之一。“永恒之蓝”工具就是其中利用漏洞的一个典型代表，其被用来传播 WannaCry勒索病毒。

网站挂马攻击

挂马攻击一直以来是黑客们热衷的一种攻击方式，常见的有通过攻击正常站点，插入恶意代码实施挂马，也有自己搭建恶意站点诱骗用户访问的。

破解软件与激活工具

破解软件与激活工具通常都涉及到知识产权侵权问题，一般是由个人开发者开发与发布，缺少有效的管理，其中鱼龙混杂，也是夹带木马病毒的重灾区。

通过U盘感染

U盘随意使用U盘拷备文件，内外网混用等，易于传播病毒

勒索：WannaCry 索要比特币

事件概况

2017年 蠕虫病毒攻击

影

- 勒索 11200 家受
- 我国 大领域，超30

事

- 201 对打补丁的 版本，无法获取补丁，因此在全球造成大范围传播。

勒索：哥斯达黎加被勒索攻击

"FOR COSTA RICA"

https://www.hacienda.go.cr/
https://www.mtss.go.cr
https://fodesaf.go.cr
https://siua.ac.cr

It is impossible to look at the decisions of the administration of the President of Costa Rica without irony, all this could have been avoided by paying you would have made your country really safe, but you will turn to Bid0n and his henchmen, this old fool will soon die. You also need to know that no organized team was created for this attack, no government of other countries has finalised this attack, everything was carried out by me with a successful affiliate, my name is unc1756. The purpose of this attack was to earn money, in the future I will definitely carry out attacks of a more serious form at with a larger team, Costa Rica is a demo version.
Pedir un Servicio privado de destrucción y destrucción, muy caro, prepago, garante
exp//profile/126771-unc1756/

It is impossible to look at the decisions of the administration of the President of Costa Rica without irony, all this could have been avoided by paying you would have made your country really safe, but you will turn to Bid0n and his henchmen, this old fool will soon die. You also need to know that no organized team was created for this attack, no government of other countries has finalised this attack, everything was carried out by me with a successful affiliate, my name is unc1756. The purpose of this attack was to earn money, in the future I will definitely carry out attacks of a more serious format with a larger team, Costa Rica is a demo version.
Pedir un Servicio privado de destrucción y destrucción, muy caro, prepago, garante
exp//profile/126771-unc1756/

PUBLISHED 975

08/05/2022 | 27870 | 54 [672.19 GB]

/ ROOT

(mtss desaf)2021.rar	1.94 GB
2.zip	252.40 MB
2022.rar	42.94 MB
3.zip	10.75 GB
4.zip	7.35 GB
5.zip	156.08 MB
6.zip	18.53 GB
9.zip	7.77 GB

全国200多家三甲医院中招勒索病毒

全国200多家三甲医院中招勒索病毒：2018年9月，在一份报告中爆出，在全国三甲医院中，有247家医院检出了勒索病毒，以广东、湖北、江苏等地区检出勒索病毒最多。互联网时代，随着移动医疗、AI医疗影像、电子病历等等数字化程序的普及，医疗数据被泄露屡见不鲜。

网络攻击：DDoS攻击

事件1

mirai僵尸网络发动的DDoS攻击-美国断

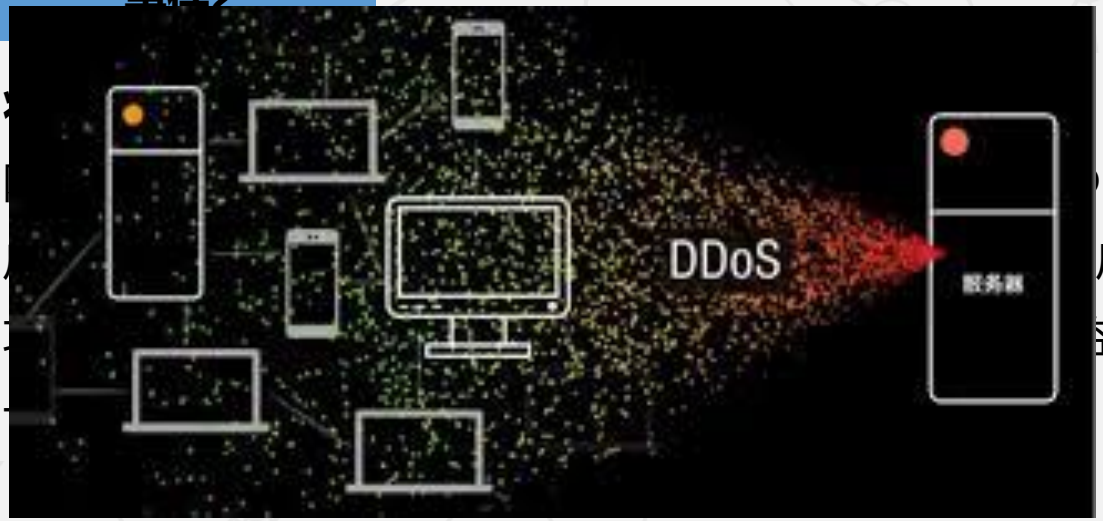
2016年10月21 日美国大面积断网事件：

问，美国主要公共服务、社交平台、民众

事件原因是攻击者对美国互联网域名解析

球约89万台智能设备，就是攻击流量的主

事件2



DoS攻击：这些攻击事件源于大量用

应用的恶意APP，该APP在动态接收到

监测的数据显示，近两个月，已经有五

网络攻击：北京健康宝遭受攻击

事件概况

2022年4 月 28 日，在北京召开的第 318 场疫情防控新闻发布会上，北京健康宝使用高峰期遭受网络攻击，经初步分析，网络攻击源头来自境外，北京健康宝保障团队进行及时有效应对，受攻击期间，北京健康宝相关服务未受影响。经分析，这是一起典型的网络拒绝服务攻击（DDoS 攻击）事件。

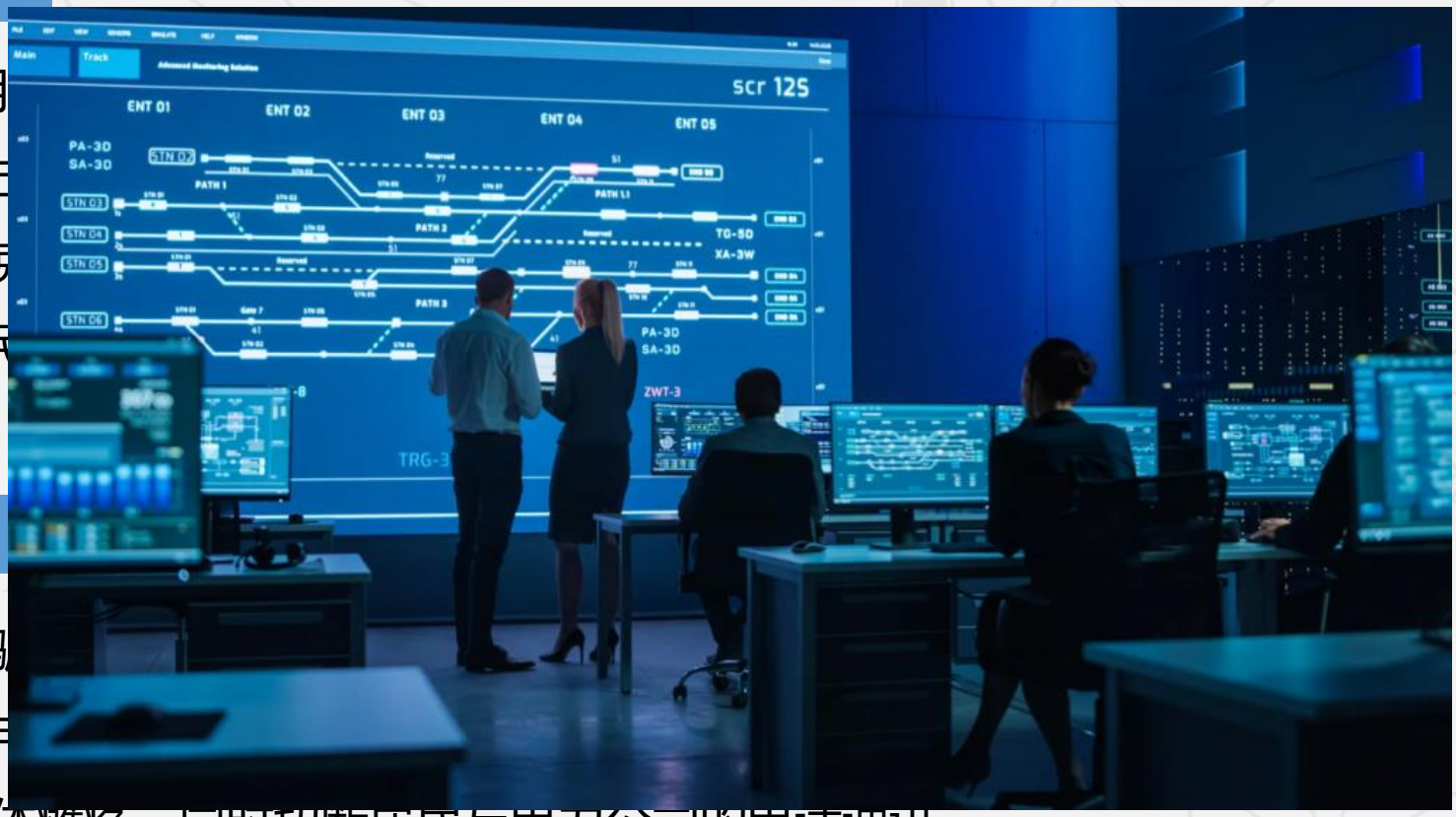
事件总结

这次事件的发起方是其内部命名为 Rippr 的团伙。该团伙使用了已经披露过的恶意代码家族 Fbot 作为攻击武器。此次事件的 Fbot 变种，最早发现于 2 月 10 日，自被发现以来就异常活跃地参与到 DDoS 攻击中。

破坏国家关键基础设施：乌克兰

事件概况

2015年12月
恶意代码攻击
弗兰科夫斯克
140万名居民



事件总结

黑客利用欺骗
用此软件将电
毒让电脑全体瘫痪，同时切断居民与电力公司的电话通讯。

破坏国家关键基础设施：俄乌网络战

事件概况

2022年2月24日，俄罗斯针对乌克
物理战场之外，是以俄乌为主的多
上的激烈较量。在俄乌网络战中，
毁，还有网络信息战对舆论的影响

事件总结

数字战争带来的思考和启示：

网络战成为融入现代战争不可或缺的一部分；

破坏性APT攻击和DDoS攻击成为国家级网络战重要手段；

针对关键基础设施的攻防对抗成为网络战的关键。

网络舆情攻击：突发公共卫生事件

事件概况

2020年春节前夕，新型冠状病毒“COVID-19”引起的肺炎疫情从武汉席卷至全国，导致全国陷入紧张的气氛之中。疫情发生初期，由于病毒的未知性、病毒的突发性和强传染性，导致大量病人没有及时受到良好的医疗照护，这不仅造成了患病当事人及家属的恐慌，还产生了许多“医疗瘫痪”“政府瞒避”等负面话题和视频在各大社交平台上快速传播蔓延，一时间引发了社会的巨大恐慌，并使得群众对政府公信力产生了强烈质疑。

事件分析

澳大利亚研究院2020年发布的一份报告显示，数以千计的推特账号曾在10天内大量转发新冠病毒相关信息，“步调一致、有组织地散播新冠病毒是中国‘生化武器’的阴谋论”，而这些“网络水军”都与美国有关。

网络舆论攻击：个人盈利违法行为

事件1

广东公安机关网安部门查明，某“网络水军”团伙利
门为特定网络直播间提供有偿代刷虚假评论、点赞、
8000万条，代刷直播间浏览量1亿余次、点赞量7亿余
名犯罪嫌疑人抓获归案，捣毁作案窝点3个，查扣作案

事件2

广西公安机关网安部门查明，刘某利用配音和剪辑软
社会事件虚假信息后在短视频平台发布，博人眼球、
后，再以数千元不等的价格进行出售。刘某通过“造
并发布虚假视频信息79个，累计播放量3153万次、点
出售账号获利3.8万元。目前，刘某已被属地公安机关

信息泄露：万豪酒店被APT入侵拖库

事件概况

2018年11月，万豪国际集团表示其公司旗下喜达屋酒店的一个客房预订数据库被黑客入侵，多达5亿人次的详细信息可能遭到泄露。泄露数据库中包含约5亿名客人信息，其中高达3.27亿人次的泄露信息包括名字、邮寄地址、电话号码、护照号码、生日、到达和离店信息等。

自2014年起，即存在第三方对其旗下喜达屋网络未经授权的访问。黑客入侵后不破坏数据，只潜伏，在服务器里安置“后门”，达到源源不断获取最新数据的目的。针对企业数据库的攻击手段很多，简单的如弱口令暴力破解、SQL注入等。

事件总结

1）酒店集团信息泄露事件三大主因：

未经授权的第三方组织窃取数据、特权账号被公开至Github导致泄露、POS机被恶意软件感染。

2）僵尸的感染对象已经从服务器、PC，扩展向智能手机APP、摄像头、路由器、家居安防系统、智能电视、智能穿戴设备等。

信息泄露：高铁数据被泄露

事件概况

2022年4月，我国国家安全机关破获一起为境外刺探、非法提供高铁数据的重要案件。上海某科技公司为牟取利益，持续采集、传递数据给某境外公司。这起案件是《中华人民共和国数据安全法》实施以来，首例涉案数据被鉴定为情报的案件，也是我国首例涉及高铁运行安全的危害国家安全类案件。

事件总结

国家基础信息、国家核心数据事关国家安全、国计民生和重大公共利益，是数据安全保护工作的重中之重。希望全社会进一步增加国家安全意识，坚持总体国家安全观，共同建立健全数据安全治理体系，提高数据安全保障能力，筑牢维护国家安全的钢铁长城。

信息泄露：NetFlix与差分隐私

事件概况

2006 年 10 月，Netflix 举办了一项大奖赛（Netflix Prize Competition）。比赛中，Netflix 提供了一个数据集，包含了从 1988 年到 2005 年间，超过 48 万个随机选择的匿名用户对于一万七千多部电影的评分。数据集中包括的具体数据有：用户 ID（随机分配，无法推出真实 ID）、电影信息（ID、年份、标题、用户对电影的评分等。Netflix 希望参赛队提出一种算法，使得在相同的训练集上，新算法预测用户喜好的性能比现有算法的性能高出 10%，并宣布对效果最好的队伍给予 50000 美元的奖励。这场比赛持续了 5 年之久，引发了各国参赛队伍激烈的竞争。然而，2008 年，来自德州大学奥斯汀分校的两名研究人员 Arvind Narayanan 和 Vitaly Shmatikov 发表了名为《Robust De-anonymization of Large Sparse Datasets》的论文，详细描述了如何对 Netflix 提供的数据进行隐私攻击。为此Netflix被发起了集体诉讼，最终以600W美金了结此事。

事件总结

如果你要公布一个数据集，仅仅粗暴的移除其中的ID这类敏感信息是完全不足以保护隐私的。于是在该事件发生之后的同一年，微软的C. Dwork提出了一个概念，叫做Differential Privacy，也就是差分隐私。

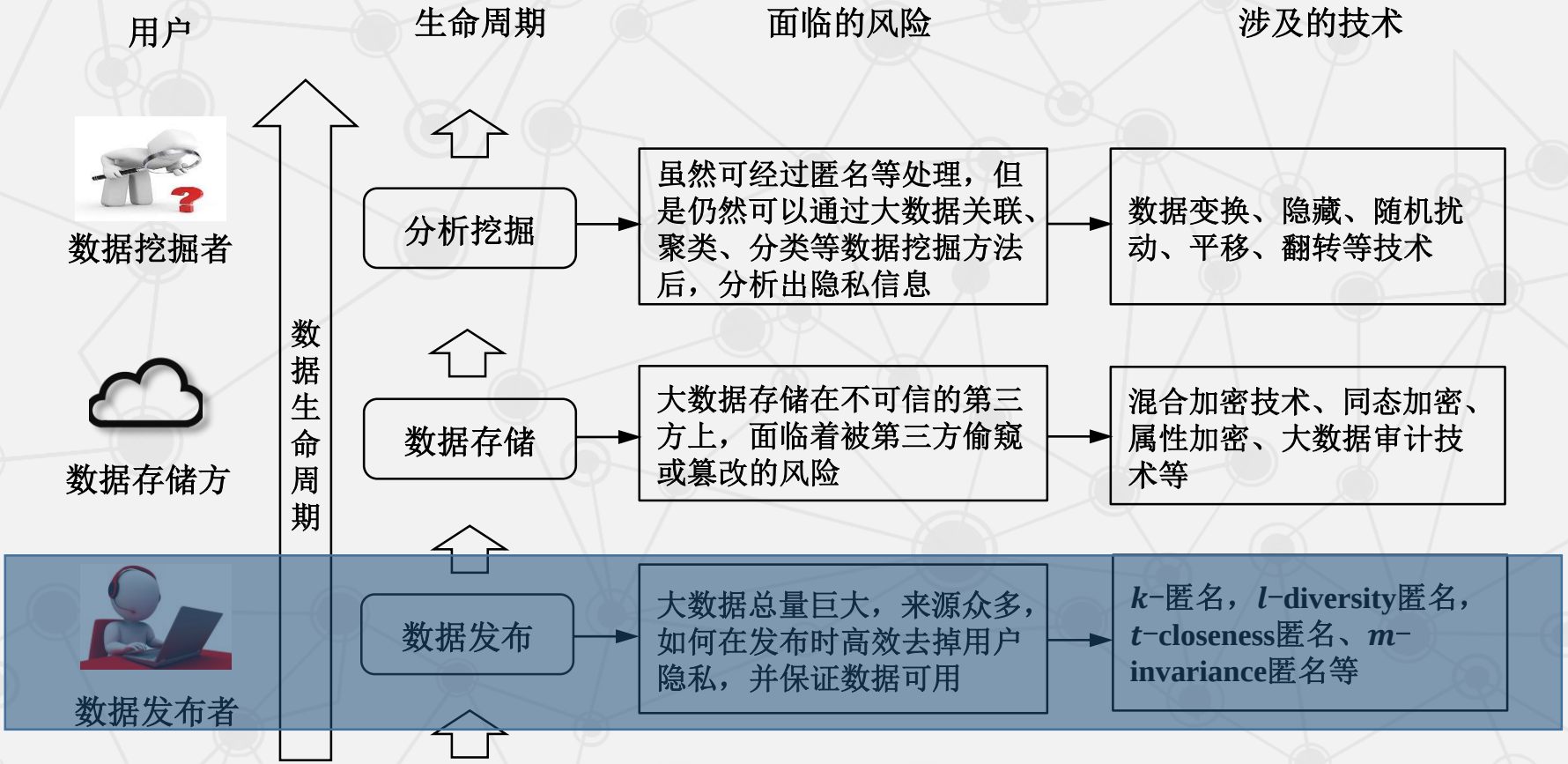
大数据隐私的定义及量化

- 维基百科：Privacy is the ability of an individual or group to seclude themselves or information about themselves, and thereby express themselves selectively.

- 百度百科：与公共利益、群体利益无关，不愿告人或不愿公开的个人的私事。

- 隐私 =（信息本体+属性）×时间×地点×使用对象

大数据生命周期包括：发布、存储、挖掘和使用等过程，涉及发布者、存储方、挖掘者和使用者等数据用户。



简单匿名化

- 简单匿名化(即删除姓名、地址等敏感信息)是否安全?
- 为了推动公共医学研究，麻省保险委员会发布了政府雇员的医疗数据。为防止用户隐私泄露，删除了姓名、身份证号和家庭住址等敏感信息。然而，MIT的Sweeney成功破解了这份匿名化处理后的医疗数据，能够具体确定某人的医疗记录。因为虽然删除了敏感信息，但仍然保留了3个关键字段：性别、出生日期和邮编。Sweeney同时有一份公开的该州投票人名单(属性包括姓名、性别、出生日期、住址和邮编等)。她将两份数据进行匹配，发现匿名医疗数据中与被攻击者生日相同的人有限，而其中与被攻击者性别和邮编都相同的人更是少之又少。由此，Sweeney就能确定被攻击者的医疗记录。

链接攻击技术

攻击者能从其他渠道获得包含了用户标识符的数据集，并根据准标识符连接多个数据集，重新建立用户标识符与数据记录的对应关系。这种攻击称为链接攻击（linking attack）。

同一信息在不同数据集中的敏感程度不同，隐私数据很容易通过多数据集的交叉匹配被挖掘出来。

tId		Province	age	sex	disease	Zip
1		安徽	32	女	肺炎	100870
2		安徽	32	女	流感	100871
3		安徽	32	男	哮喘	100876
4		安徽	32	男	肺结核	100875
5		重庆	37	女	脂肪肝	101230
6		重庆	37	女	脂肪肝	101231
8		重庆	37	女	甲亢	101231

Name	Sex	Age	Province	Zip	Date	Party
李青	女	32	安徽	100870	2016.07	共产党
刘然	女	37	重庆	101231	2016.07	共产党
马方	男	34	湖南	436570	2016.07	共产党
沈思	女	37	重庆	101231	2016.07	共产党
张佳	男	33	天津	106470	2016.07	共产党
方淮	男	38	河北	104670	2016.07	共产党

K匿名技术

- 使多条记录具有相同的准标识码组合，这些具有相同准标识码组合的记录集合被称为等价组。
- K-匿名：要求发布的数据中存在一定数量的在准标识符上不可区分的记录，即每个等价组中的记录个数为k个，使攻击者不能判断出隐私信息所属的具体个体。

Age	Sex	Zip	Disease
(20,30]	M	1211**	肺炎
(20,30]	M	1211**	流感
[30,40]	F	1315**	糖尿病
[30,40]	F	1315**	糖尿病
(40,50]	*	1526**	心脏病
(40,50]	*	1526**	高血压

L多样性匿名

Age	Sex	Zip	Disease
(20,30]	M	1211**	肺炎
(20,30]	M	1211**	流感
[30,40]	F	1315**	糖尿病
[30,40]	F	1315**	糖尿病
(40,50]	*	1526**	心脏病
(40,50]	*	1526**	高血压

- K-匿名中的k个不可区分的记录可能在隐私属性上的取值相同，此时同样会泄露隐私。
- L多样性要求k个不可区分的记录在隐私属性上至少有L个不同的取值，即L-diversity保证每一个等价类的敏感属性至少有L个不同的值。

针对L多样性匿名的攻击

- 假设原始数据只有一个敏感属性：特定病毒的测试结果。它有两个值：阳性和阴性。进一步假设有10000条记录，其中99%为阴性，只有1%为阳性。这两个值的灵敏度是非常不同的。
- 由于敏感属性只有两个值，即只能生成2-多元表，所以最多可以有10000×1%=100个等价组。

- 针对某一个满足2多样性的等价组：49个阳性和1个阴性。问题来了，该等价组中的任何人都将被认为是98%的阳性，而不是整体数据上的1%，这是用户不希望的。

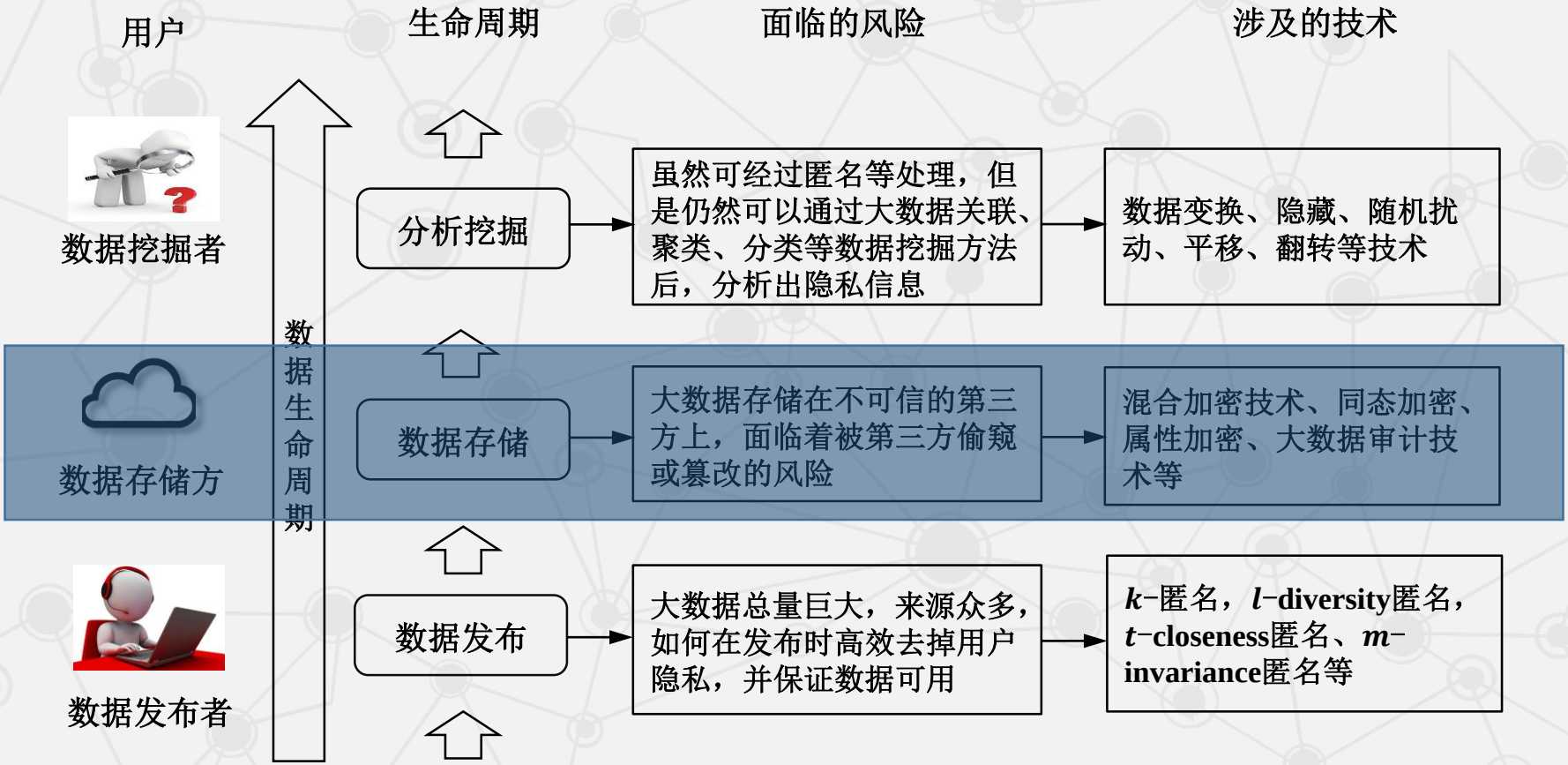
T相近匿名

- 若等价组中敏感值的分布与整个数据集中敏感值的分布具有明显的差别，攻击者可以以一定概率猜测目标用户的敏感属性值。
- t-相近要求每个k-匿名组中敏感属性值的统计分布与该属性在整个数据集中的总体分布“接近”。
- 若一个等价组的敏感属性取值分布与整张表中该敏感属性的取值分布的距离不超过阈值t，则称该等价组具有t相近性。若一个表中所有等价组都有t相近性，则该表也有t相近性。
- t-closeness匿名以EMD（earth mover’s distance，陆地移动距离）衡量敏感属性值之间的距离。

m-invariance匿名

- 数据发布匿名机制最初只考虑了发布后不再变化的静态数据，但大数据的一大特点是其会不断动态更新。
- 对于任意一条记录，只要此记录所在的等价组在前后两个发布版本中具有相同的敏感属性值集合，不同发布版本之间的推理通道就可以被消除。
- 为了保证这种约束，研究者引入虚假的用户记录，这些用户记录不对应任何原始数据记录，只是为了消除不同数据版本间的推理通道而存在，同时引入了额外的辅助表标识等价类中的虚假记录数目，以保证数据使用时的有效性。

大数据生命周期包括：发布、存储、挖掘和使用等过程，涉及发布者、存储方、挖掘者和使用者等数据用户。



同态加密技术

- 同态加密是基于数学难题的计算复杂性理论的密码学技术。对经过同态加密的数据进行处理得到一个输出，将这一输出进行解密，其结果与用同一方法处理未加密的原始数据得到的输出结果是一样的。

$$E(\mathrm{msg}_1)E(\mathrm{msg}_2) = E(\mathrm{msg}_1 + \mathrm{msg}_2)$$

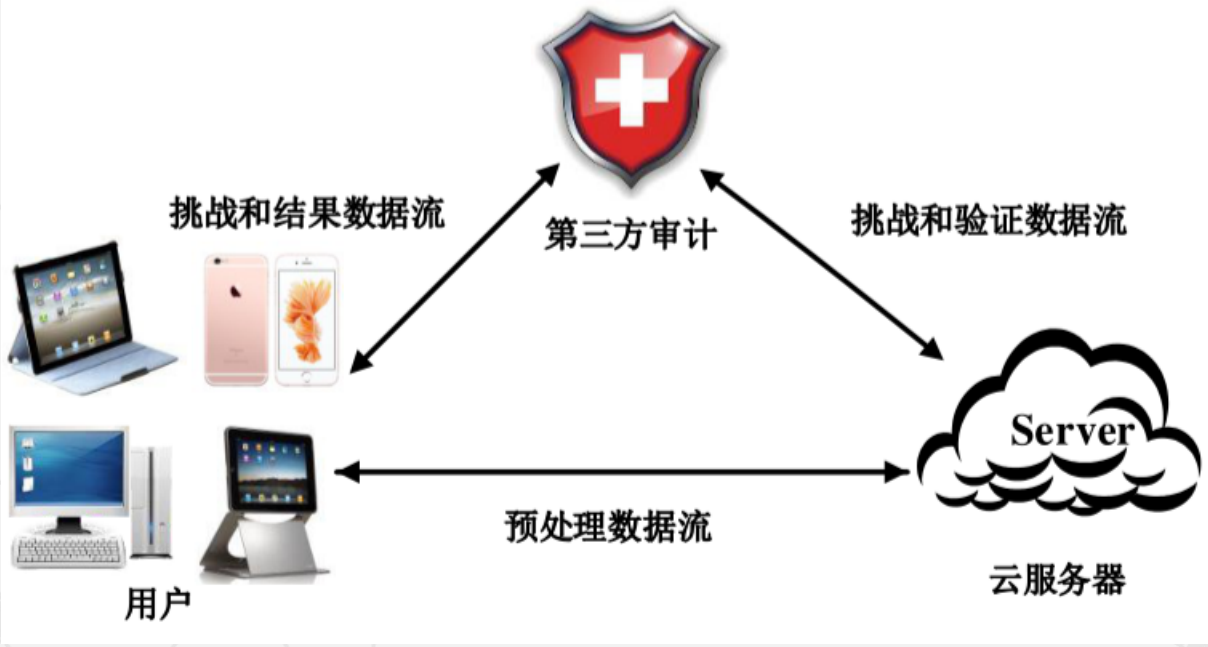
$$E(\mathrm{msg}_1)^{\mathrm{msg}_2} = E(\mathrm{msg}_1\mathrm{msg}_2)$$

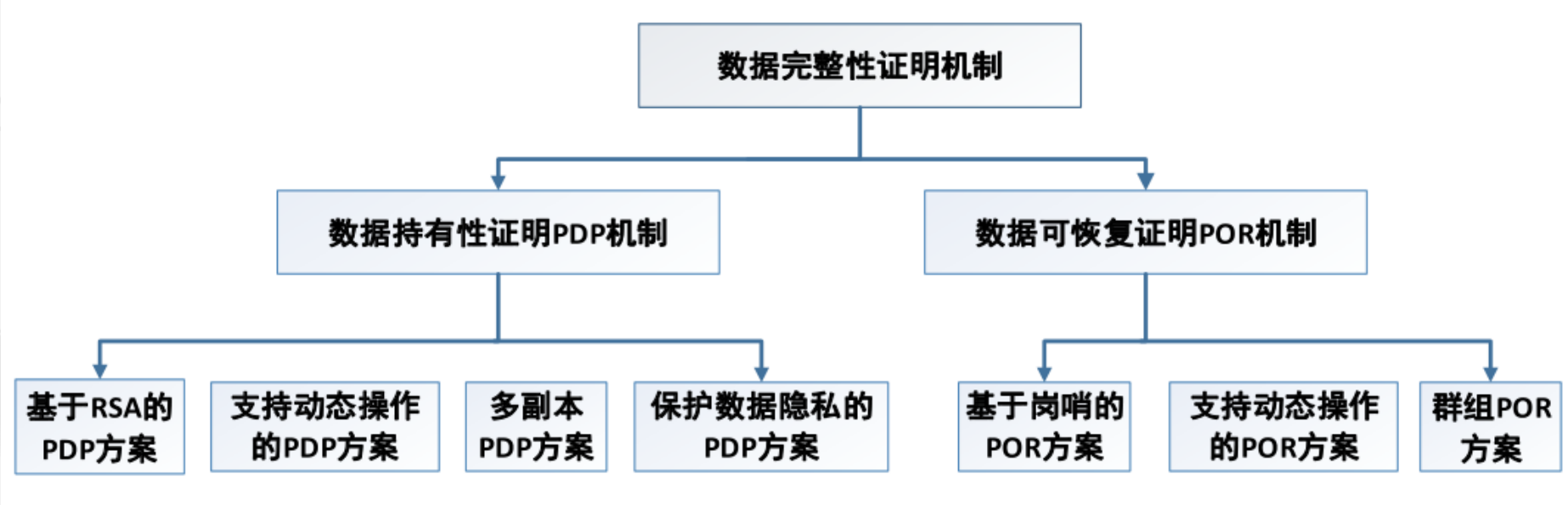
大数据审计技术

- 云存储审计是指数据拥有者或者第三方机构对云上的数据完整性进行审计，通过对数据进行审计，确保数据不会被云服务提供商篡改、丢弃，并且在审计的过程中用户的隐私不会被泄露。
- 可证明的数据持有（provable data possession，PDP）模型：先从服务器上随机采样相应的数据块，并生成持有数据的概率证据，客户端维持着一定数量的元数据，并利用元数据来对证据进行验证。
- 可恢复证明（proof of retrievability，POR）模型：原始文件首先被纠错码编码并产生对应标签，编码后的文件及标签被存储在服务器上。当用户选择服务器上的某个文件块时，可以采用纠错码解码算法来恢复原始文件。

云安全


预处理数据流
Server
挑战和验证数据流
用户
云服务器


挑战和结果数据流
第三方审计
挑战和验证数据流
Server
预处理数据流
用户
云服务器



- 数据持有性证明：本地保存哈希函数的结果，随机存取一小块数据操作验证。

- 数据可恢复性证明：已经编码的数据块中加入“岗哨”的数据块，抽取岗哨比对证明完整性，并通过判断出错岗哨的数量判断是否可回复。

基于属性的加密技术

- 制定数据的访问策略(对属性的要求)
- 数据按照访问策略进行加密
- 按照用户具有的属性为用户发放数据

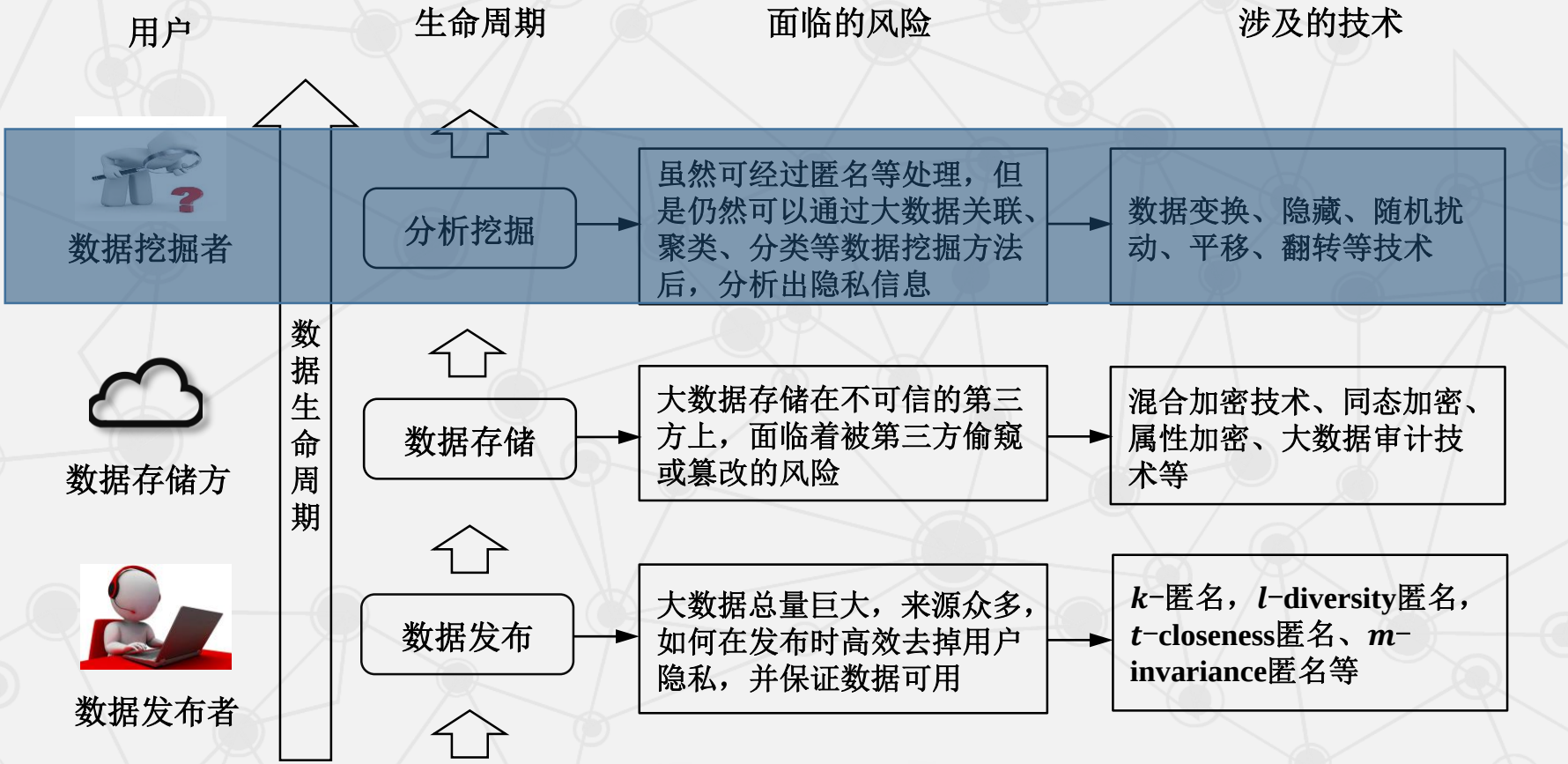
- KP-ABE (Key-Policy Attribute-Based Encryption)

用户(查询方)规定访问密文的策略，因此适合长效查询类应用，如付费电视、视频点播、数据库访问等。

- CP-ABE(Ciphertext-Policy Attribute-Based Encryption)

发送者(发布方)规定访问密文的策略，因此适合灵活访问控制类应用，如网络论坛访问、征婚广告等。

大数据生命周期包括：发布、存储、挖掘和使用等过程，涉及发布者、存储方、挖掘者和使用者等数据用户。



2006 年 10 月，Netflix 举办了一项大奖赛（Netflix Prize Competition）。比赛中，Netflix 提供了一个数据集，包含了从 1988 年到 2005 年间，超过 48 万个随机选择的匿名用户对于一万七千多部电影的评分。数据集中包括的具体数据有：用户 ID（随机分配，无法推出真实 ID）、电影信息（ID、年份、标题、用户对电影的评分等。Netflix 希望参赛队提出一种算法，使得在相同的训练集上，新算法预测用户喜好的性能比现有算法的性能高出 10%，并宣布对效果最好的队伍给予 50000 美元的奖励。这场比赛持续了 5 年之久，引发了各国参赛队伍激烈的竞争。然而，2008 年，来自德州大学奥斯汀分校的两名研究人员 Arvind Narayanan 和 Vitaly Shmatikov 发表了名为《Robust De-anonymization of Large Sparse Datasets》的论文，详细描述了如何对 Netflix 提供的数据进行隐私攻击。为此Netflix被发起了集体诉讼，最终以600W美金了结此事。

差分隐私技术

- 在2016年WWDC主题演讲中，苹果工程副总裁Craig Federighi宣布苹果使用本地化差分隐私技术来保护iOS/MacOS用户隐私。根据其官网披露的消息，苹果将该技术应用于Emoji、QuickType输入建议、查找提示等领域。

- Google利用本地化差分隐私保护技术从Chrome浏览器每天采集超过1400万用户行为统计数据。

差分隐私技术

定义 1 （ε-差分隐私）对于任意一对相邻数据库（定义为差别最多有一个记录的两个数据库）B_1和 B_2，任意一个可能的查询结果 S，一个提供ε-差分隐私保护的算法 M 必须满足：

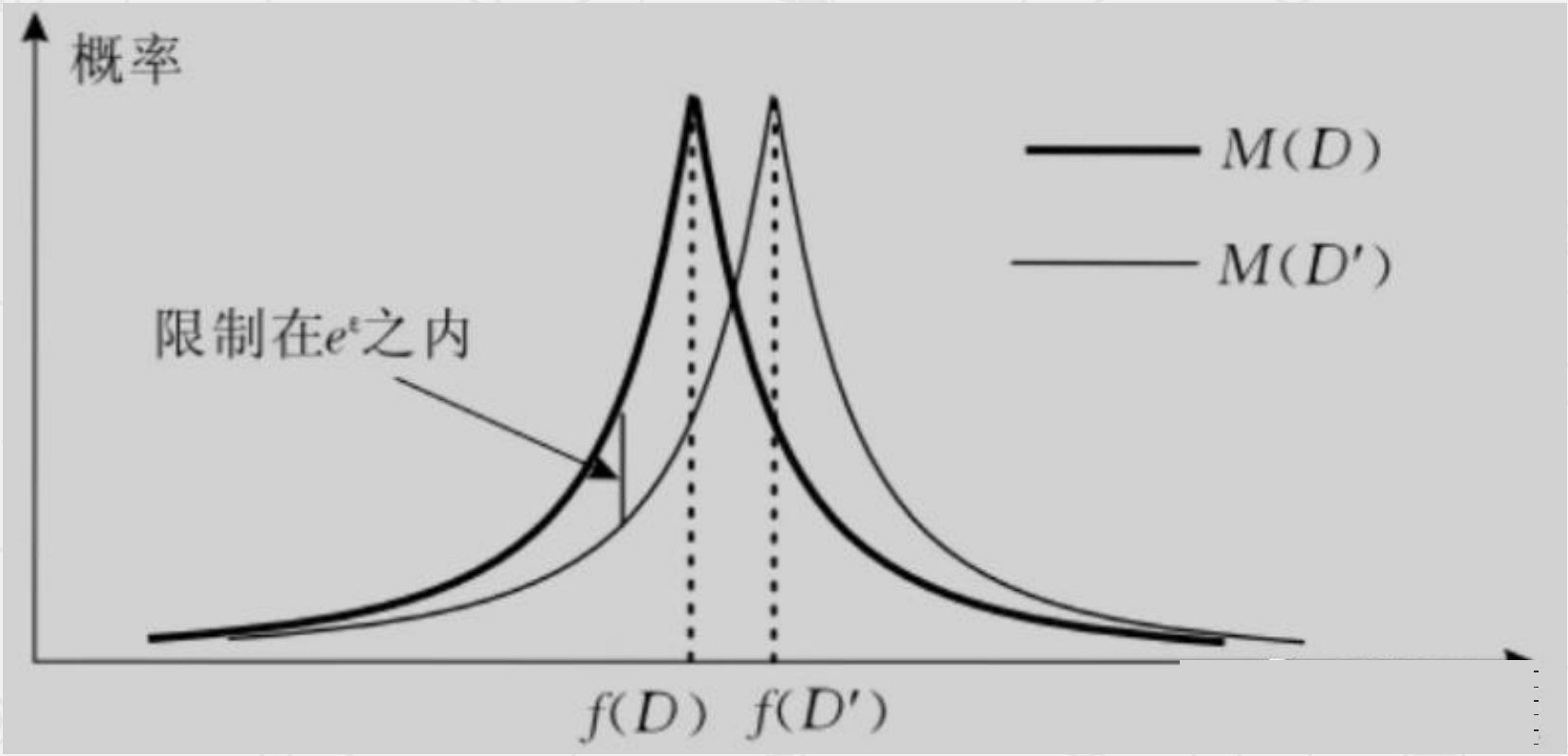
$$\Pr[M(B_1) = S] \leq \exp(\varepsilon) \cdot \Pr[M(B_2) = S].$$

姓名	诊断结果
Tom	0
Jack	1
Henry	1
Diego	0
Alice	1

差分隐私技术-拉普拉斯机制

- 如果案例中的f是一个提供ε-差分隐私保护的查询函数。例如$f(i)=count(i)+noise$,其中$noise$是服从某种随机分布的噪声。
- 这种针对统计输出的随机化方式使得攻击者无法得到查询结果间的差异，从而能保证数据集中每个个体的安全。
- $noise$噪声值由Laplace机制提供。由于Laplace机制仅适用于数值型查询结果，而在许多实际应用中，查询结果为实体对象(例如一种方案或者一种选择)，为此，提供了指数机制来针对实体对象提供查分隐私保护 。

差分隐私技术-拉普拉斯机制



差分隐私技术-拉普拉斯机制

- 随机算法M在两个数据集上的概率分区越接近，输出的结果就越难区分，可区分性差，隐私保护强度越高。
- 特殊情况下，当为0的时候，两个数据集的分布重合，输出结果完全不可区分，隐私强度是高了，但是这种情况下，原始数据的可用性也就丧失了。
- 我们要做的是数据信息的隐私保护与可用性之间的平衡，所以隐私系数的设置至关重要。

差分隐私技术-指数机制

The Exponential Mechanism: The exponential mechanism $\mathcal{M}_E(x, u, \mathcal{R})$ selects and outputs an element $r \in \mathcal{R}$ with probability proportional to $\exp(\frac{\epsilon \cdot u(x,r)}{2\triangle u})$

- 设随机化算法M输入为数据集 X，输出为一个实体对象r∈R,μ(X,R)为可用性函数，△μ为函数μ(x,R)的敏感度，若以正比于exp(ϵ·u(x,r)2△u)的概率从输入中选择并输出r，则算法M是 ϵ-DP的。

差分隐私技术-指数机制

- 假如拟举办一场体育比赛，可供选择的项目来自集合｛足球，排球，篮球，网球｝，参与者们为此进行了投票，现要从中确定一个项目，并保证整个决策过程满足ε - 差分隐私保护要求。

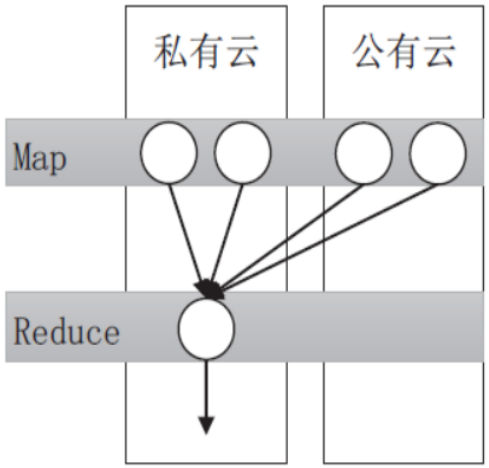
项目	可用性	概率		
	$\Delta q = 1$	$\varepsilon = 0$	$\varepsilon = 0.1$	$\varepsilon = 1$
足球	30	0. 25	0. 424	0. 942
排球	25	0. 25	0. 330	0. 075
篮球	8	0. 25	0. 141	1. 5E−05
网球	2	0. 25	0. 105	7. 7E−07

基于数据分离的隐私保护

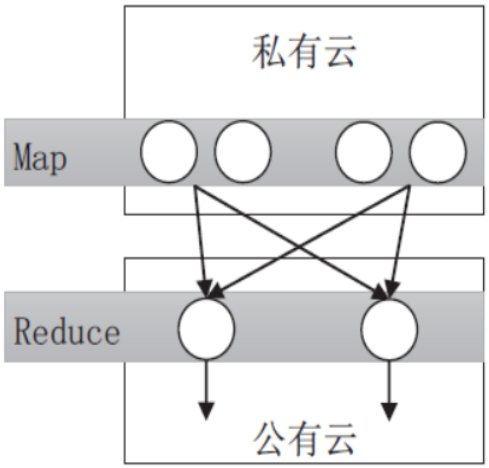
- 混合云分离执行：

1）对原始数据中包含的敏感数据进行标记，将数据划分为敏感数据集和非敏感数据集；

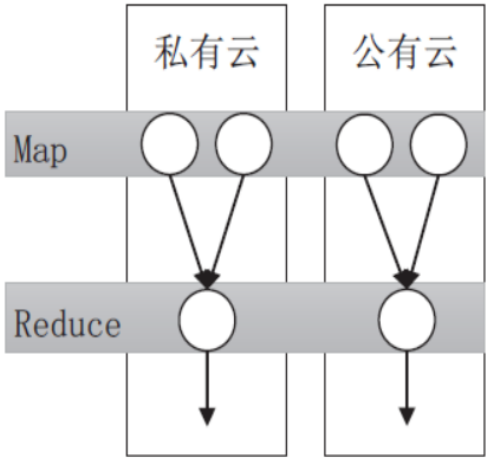
2）将在不同数据集上的相关计算任务也进行划分，并将非敏感数据及其相关的计算任务外包到公有云存储并计算，而小规模的敏感数据及其相关的计算任务保留在本地或者安全的私有云执行。



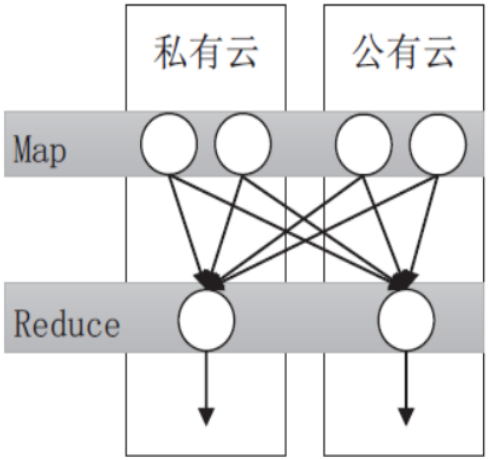
(a) Map混合



(b) 水平分割



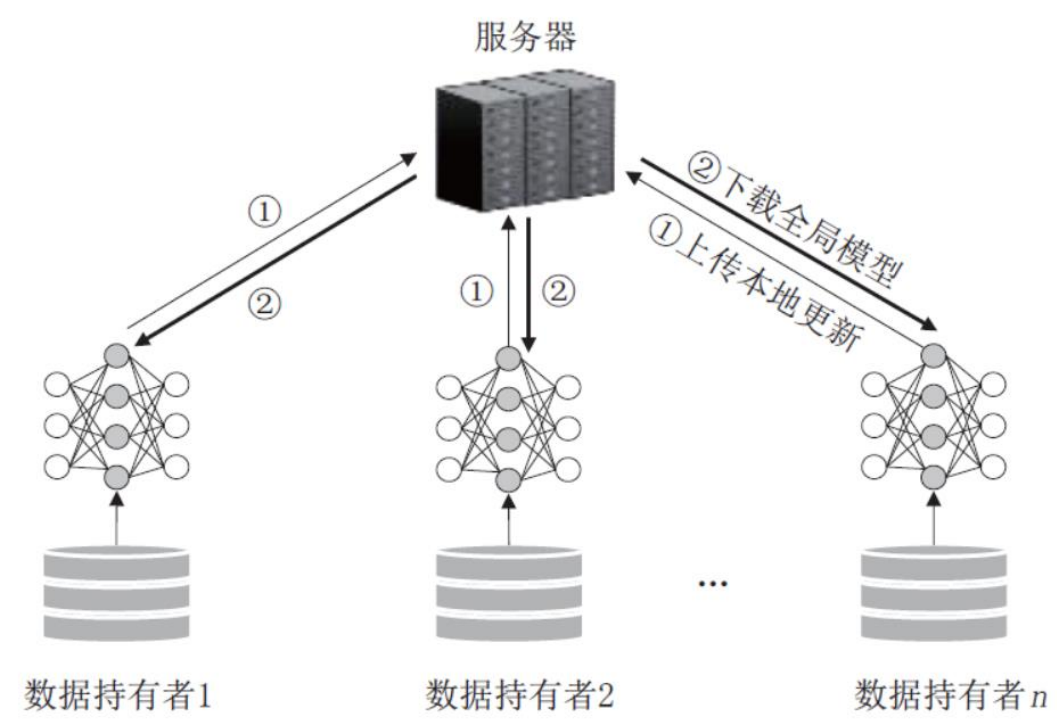
(c) 垂直分割



(d) 混合

基于数据分离的隐私保护

- 联邦学习：原始数据全部在本地存储及计算的思路，特别是对于敏感信息比较密集，且不太容易被标记和划分的原始数据集。



基于安全多方计算的隐私保护

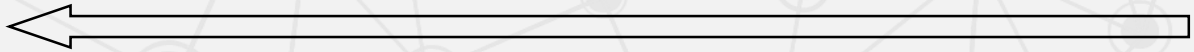
- 姚氏混淆电路：使用布尔电路（boolean circuit）表述待计算函数，结合不经意传输（Oblivious Transfer）技术设计安全多方计算协议

- 秘密分享：如果使用一个（k，n）门限秘密共享模型，n表示参与方总数，k表示门限，那么至少需要用k个秘密共享来重构敏感值，每个秘密共享都不能泄露任何有关原始值的信息。

百万富翁问题

(1) Bob选择一个大整数x，用Alice的公钥加密，得到k=Enc(x)；

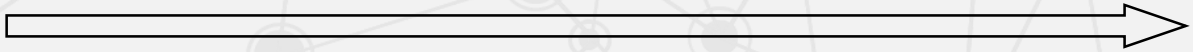
(2) Bob把 y=k-j+1 发送给Alice；



Alice

Bob

金额 i



金额 j

(3) Alice生成一组数据：{y, y+1, …. y+9}；

(4) Alice计算 Dec{y, y+1, …. y+9}；//Dec表示用自己的私钥解密

1<=i, j <=10

百万富翁问题

(5) Alice将结果除以一个素数p后，取余，即z1 = Dec(y) mod p、z2 = Dec(y+1) mod p、 …… z10 = Dec(y+9) mod p ；

(6) Alice的金额是i，她从z1到z10选择第i个至第10个数据加1（即zi+1），然后将10个数据发送给Bob；

(7) Bob找到第j个数据zj，如果zj = x mod p，那么i>j，否则，i<=j；

(8) Bob将结果返回给Alice。

Alice不相信怎么办？

百万富翁问题

- Bob选择了一个非常大的数x，然后加密得到k。 然后把k-j+1发给了Alice。这个时候Bob的信息已经藏在这个数据里面了，但是因为x很大，所以Alice没办法从k-j+1之中推导出j，对Alice来说就是个随机数。

- 然后Alice 计算了10个数，Dec(k-j+{1，2，3...10})。对Alice来说k-j+{1，2，3....10}都是没有意义的随机数，那么在这些数上解密，当然也只会得到一些没有意义的随机数。但是其中一个数是有意义的，那就是Dec{k-j+j}，为什么呢？因为k-j+j=k，而k是什么呢？k=Enc（x）那么理所当然Dec（k）=x。所以说这十个数对Alice没有意义，但是对Bob是有意义的，因为他知道x。

百万富翁问题

- Alice把这十个数发给Bob，在Bob看来，这是十个数就是9个毫无意义的随机数和一个x，而x正好在第j个位置上
- Alice接下来在这个十个数从第i个数开始都+1，再给Bob。
- Bob的视角要不就是9个看不懂的数和 x 或者是 9个看不懂的数和 x+1。
- 如果是x+1，那么他知道，自己的数j肯定不小于Alice的i，反之则小于。

目录

1　大数据隐私保护技术

2　云安全概述

3　云安全技术

2 云安全概述

传统数据中心网络架构

保护云计算本身的安全

云计算安全

安全云

安全成为云计算的一种服务

云安全

- 云计算特有的数据和服务外包、虚拟化、多租户和跨域共享等特点，带来了前所未有的安全挑战。安全和隐私问题已成为阻碍云计算普及和推广的主要因素之一；

- 由于云计算环境下的数据对网络和服务器的依赖，隐私问题尤其是服务器隐私的问题比网络环境下更加突出。存在客户对云计算的安全性和隐私保密性存疑，企业数据无法安全方便地转移到云计算环境等一系列问题；

云安全

云计算的特点	安全威胁
数据和服务外包	（1）隐私泄露 （2）代码被盗
多租户和跨域共享	（1）信任关系的建立、管理和维护更加困难； （2）服务授权和访问控制变得更加复杂； （3）反动、黄色、钓鱼欺诈等不良信息的云缓冲 （4）恶意SaaS应用
虚拟化	（1）用户通过租用大量的虚拟服务使得协同攻击变得更加容易，隐蔽性更强； （2）资源虚拟化支持不同租户的虚拟资源部署在相同的物理资源上，方便了恶意用户借助共享资源实施侧通道攻击。

云安全定义

云计算安全性常被称为：云安全。

云安全由策略、程序、技术和控制组成，它们共同保护云中的系统、基础设施和数据。

云安全侧重于确保与云交互的网络的各个方面都受到保护。它还确保云基础设施免受网络攻击和复杂的威胁。

云计算安全挑战

拒绝服务攻击：分布式拒绝服务攻击 (DDoS) 越来越常见。攻击者使用拒绝服务攻击使网站的服务器不堪重负，以至于无法响应用户请求。这可能会导致网站长时间无法使用。

数据丢失：由于各种原因，数据可能会从云中丢失。即使没有人主动尝试获取数据，数据也可能由于事故或自然灾害而丢失。云安全解决方案不仅有助于防止窃贼，还可以包括保护数据免受意外事件影响的措施。

数据泄露：如果数据不受保护，则云上的数据对于黑客而言可能是缓慢移动的靶标。一些黑客窃取数据以敲诈组织的成员。其他黑客则将数据出售给希望了解公司机密的实体。如果没有正确的云安全解决方案，公司的数据可能会泄露。

云计算安全挑战

易受攻击的接入端：基于云的系统提供无与伦比的访问，但用于与云交互的设备通常不受保护或保护不足。因此，几乎所有访问云系统的手机、平板电脑、笔记本电脑或其他移动设备都可能为心怀不轨的人员或软件提供攻击机会。然而，借助适当的防火墙，可以限制对适当类型流量的访问。

警报和通知：当存在安全漏洞时，在威胁造成重大损害前予以阻止只是工作的一部分。一个全面的系统将确保向重要的利益相关者通报情况。通常，即使在攻击被阻止后，攻击的损害也来自于 IT 团队需要过长的时间以做出反应并提醒其他人。迅速式云安全系统可在需要时向需要了解情况的人提供警报和通知。

大云安全威胁

- 数据外泄：最具破坏性的外泄是针对敏感数据的，包括金融和健康信息、商业机密和知识产权。最终的责任在于在云中维护数据的组织。
- 受损的凭证和中断的身份认证：由于身份认证松懈、弱密码、密钥或者证书管理不善造成的。
- 入侵接口和API：当第三方依赖API时，就会泄露更多的服务和凭证。
- 利用系统脆弱性：资源共享和多租户造成了新的攻击面，但是发现和修复漏洞的成本要比潜在的损害要低。
- 账户劫持：所有账户都应该收到监管，以便每笔交易都可以追溯到强求他的个人。

大云安全威胁

- 恶意的内部人员：很难检测，系统管理员的错误有时可能被错误地诊断为威胁，较好的策略是职责分离并强制执行活动。
- 其他的威胁是：高级持久威胁、永久性数据丢失、不尽责调查、云服务滥用、DoS攻击、共享技术、硬件故障、自然灾害、与云有关的恶意软件、基础设施设计和规划不足、销售点入侵和支付卡窃取器、犯罪软件和网络空间间谍、内幕和特权滥用、web应用攻击和物理失窃/损失。

2010年云安全联盟

滥用云计算：利用云计算进行不法活动，比如使用IaaS支持的多个AWS实例或者应用程序发起DDoS攻击或发垃圾邮件和恶意软件。

不安全的API可能无法在从身份认证和访问控制开始到运行期间监管和控制应用程序的一系列活动中保护用户。

恶意内部人员：这种特殊形式的攻击所造成的潜在危害巨大。

共享技术：考虑虚拟化支持的多租户访问带来的威胁。

账户劫持：云用户必须意识到并防范所有窃取凭证的方法。

数据丢失或泄露是对使用云服务的个人或者组织具有破坏性的两种风险。

恶意使用云计算

- 说明：攻击者使用云的理由与合法消费者相同：低成本进行巨量处理。
- 影响：密码破解、DDoS、恶意存取、垃圾邮件、CAPTCHA破解等。
- 举例：Amazon的EC2出现了垃圾邮件和恶意软件的问题。

数据丢失数据泄露

- 说明：由于不合适的访问控制或者弱加密造成数据破解；因为多租户结构，不够安全的数据风险较高。
- 影响：数据完整性和数据保密性。
- 举例：喂，别碰我的云：可查询第三方电脑中的数据泄露。

恶意业内人士

- 说明：云提供商的员工可能滥用权力访问客户数据/功能；减少内部进程的可视性可能会妨碍探测这种违法行为。
- 影响：数据完整性和数据保密性、名誉损失、法律后果。
- 举例：根据Verizon的2010数据泄露报告，48%的数据泄露是业内人士造成的，在云计算系统中会更加严重。

流量拦截或劫持

- 说明：对客户或者云进行流量拦截和改到发送；偷取凭证以窃取或者控制账户信息/服务。
- 影响：数据完整性和数据保密性、声誉影响、资源恶意使用造成的后果。
- 举例：Twitter账户盗用等。

共享技术潜在风险

- 说明：公共的硬件、运行系统、中间件、应用栈、网络组件可能有着潜在的风险。
- 影响：成功使用会影响多个用户。
- 举例：VMWare Workstation、VMware Player、VMware Server和VMware ESX中的脆弱部件。

云安全技术挑战

- 对于云计算的安全保护，通过单一的手段是远远不够的，需要有一个安备的体系，涉及多个层面，需要从法律、技术、监管三个方面进行；

- 传统安全技术，如加密机制、安全认证机制、访问控制策略通过集成创新，可以为隐私安全提供一定支撑，但不能完全解决云计算的隐私安全问题；

- 需要进一步研究多层次的隐私安全体系（模型）、全同态加密算法、动态服务授权协议、虚拟机隔离与病毒防护策略等，为云计算隐私保护提供全方位的技术支持。

目录

1　大数据隐私保护技术

2　云安全概述

3　云安全技术

云数据加密

- RSA非对称加密机制：经典加密算法（公钥和私钥）
- 同态加密：直接对加密数据进行操作，不需要解密
- 保序加密：本来有顺序的明文，加密之后依然保持顺序，但是除此之外不知道任何信息。比如要加密2，4，5三个数字，我们把它们分别换成145，4545，2524252，这样就保证了别人不知道我原来的数字，但是又保持了原来的顺序，但是我们必须保证能把后面三个数字复原回去才行（解密）

数据库服务安全

- 云用户通常将数据的控制权委托给几乎所有的云服务提供商CSP支持的数据库服务，并关注DBaaS的安全方面。用户评估DBaaS安全性的模型包括几个实体组：数据所有者、数据用户、云服务提供商CSP、第三方代理或第三方审计人员。

- 数据所有者和DBaaS用户担心数据的完整性和机密性受损，以及数据的不可用。在DBaaS中，数据丢失的主要原因包括授权、认证和会计机制不充分、加密密钥和技术使用不一致，在不进行备份的情况下更改或者删除记录，以及操作失败。

操作系统安全

- 操作系统允许多个应用程序根据一组策略共享物理系统的硬件资源。操作系统的一个关键功能是保护应用程序免受各种恶意攻击，例如未经授权访问特权信息、篡改可执行代码和欺骗；
- 操作系统的强制安全性认为：对于任何安全策略，其中策略逻辑的定义和安全属性的分配由系统安全管理员严格控制。访问控制、认证用法和加密用法策略都是强制操作系统安全要素；
- 访问控制策略是指OS如何控制对不同系统对象的访问，认证用法定义OS用于认证某主体的认证机制，而密码用法策略指用于保护数据的密码机制。

虚拟机安全

- 虚拟机逃逸的定义是指虚拟机里运行的程序利用虚拟机的漏洞突破虚拟机管理器（Hypervisor），获得宿主机操作系统管理权限，并控制宿主机上运行的其他虚拟机，导致安全环境架构的彻底破坏。

- 虚拟化环境中，上层虚拟机与底层虚拟机监控器进行交互，实现对底层物理资源的访问，而这种交互过程就是潜在的引入虚拟机逃逸漏洞的关键。虚拟机逃逸的发生来源于虚拟化技术自身的安全问题，而虚拟化自身的安全问题往往由虚拟化技术的机制产生。

虚拟化安全

- 采用虚拟化技术之前，用户可以在防火墙设备上建立多个隔离区，进行严格的访问控制，对不同的服务器采取不同的规则进行管理，即使有服务器遭受到攻击，危害仅局限在一个隔离区内，影响范围不会太大。而采用虚拟化技术后，所有GuestOS会集中连接到同一台虚拟交换机或实体交换机与外部网络通信，使得原来通过防火墙采取的防护措施就会失效，如果一台GuestOS发生问题，安全问题就会通过网络扩散到其他的GuestOS。
- 在虚拟机迁移中，常见有静态迁移和动态迁移，无论那种形式，虚拟机在不同服务器之间迁移并且这种迁移经常会是自动完成，这可能会让一些重要的虚拟机迁移到不安全的物理服务器上。

虚拟化安全 cont.

- 虚拟化存储的动态迁移过程，无法对数据进行分类，存在泄漏重要数据的风险。
- 许多用户会保持重要的VM镜像，这些标准VM镜像可以制作出新的虚拟机，从而快速部署或灾难恢复，但是这样一来，标准VM镜像的脆弱性就会被复制传播，如杀毒软件病毒库未更新、系统补丁未升级等。
- 虚拟化灾备的安全管理，由于虚拟化设备需要定期的自动备份，对这些备份的镜像的管理和镜像备份的手段的安全性，如备份服务器的安全性以及备份数据的安全性等。
- 虚拟化的数据销毁机制，残存数据是否能够完全粉碎，防止数据恢复窃取机密信息。

数据集中后的安全问题

- 用户的数据存储、处理、网络传输等都与云计算系统有关。如果发生关键或者隐私信息丢失、窃取，对用户来说无疑是致命的：

✓ 如何保证云服务提供商内部的安全管理和访问控制机制符合客户的安全需求。
✓ 如何实施有效的安全审计，对数据操作进行安全监控。
✓ 如何避免云计算环境中多用户共享带来的潜在风险。