上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

大数据安全及隐私保护

李建华 院长，教授，博导
上海交通大学信息安全工程学院
信息内容分析技术国家工程实验室
2016年4月25日

上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

提纲

一、大数据机遇和网络安全挑战

二、大数据带来的网络安全和用户隐私问题

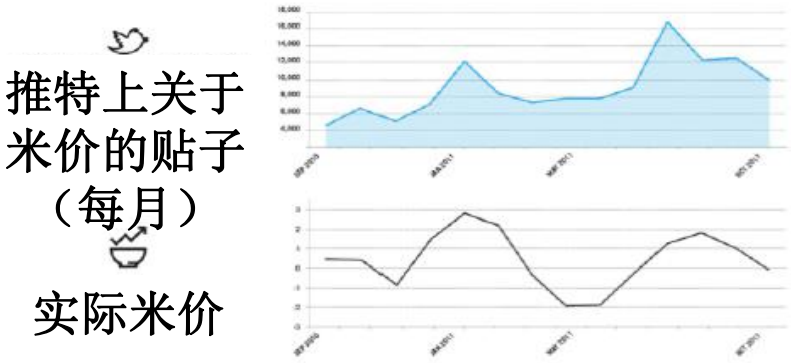
三、大数据带来的网络安全和用户隐私问题对策

大数据分析挖掘的价值

- 2010年《Science》上刊文指出，能够根据个体之前的行为轨迹预测他/她未来行踪的可能性，**即93%的人类行为可预测**。
- 大数定理告诉我们，在试验不变的条件下，重复试验多次，随机事件的频率近似于它概率。“有规律的随机事件”在大量重复出现的条件下，往往呈现几乎必然的统计特性。从“数据”到“大数据”，不仅仅是数量上的差别，更是数据质量上的提升，即从量变到质变。
- 随着计算机的处理能力的日益强大，你能获得的数据量越大，你能挖掘到的价值就越多。
- **实验的不断反复、大数据的日渐积累让人类发现规律，预测未来不再是科幻电影里的读心术。**

大数据分析挖掘的价值

- 大数据分析挖掘和数据融合的异同：大数据分析挖掘因为有极其丰富的数据作为基础，可让“有规律的随机事件”在大量重复出现的条件下，呈现几乎必然的统计特性。
  - 数据融合其实是在一定的数据量条件下，通过多源传感器的协同，改进测量和预测的结果，在发现规律、预测未来的准确性方面和大数据分析挖掘不在一个数量级上。
  - **我们认为，**从数据融合到大数据分析挖掘，其实是**从小智能到大智慧，这是大数据分析挖掘的核心价值**。



http://www.unglobalpulse.org/projects/twitter-and-perceptions-crisis-related-stress

在印尼的推特上，讨论米价的信息和实际米价的关系

大数据国家发展战略机遇

- 2015年是中国大数据发展高峰期，我国政府部门颁布了大数据开放行动的战略。
- 2015年底，《中共中央关于制定国民经济和社会发展**第十三个五年规划**的建议》**通过并提出了发展“互联网+”、分享经济和大数据等创新战略，更是将大数据开放、开发提到了国家战略高度**。
- 大数据作为社会的又一个基础性资源，将给社会进步、经济发展带来强大的驱动力。**大数据代表了先进生产力方向，已经成为不可阻挡的趋势**。

大数据安全现状

- 网络攻击成愈演愈烈之势。**如今的网络攻击，往往是通过各种手段获得政府、企业或者个人的私密数据。**在大数据时代，数据的收集与保护成为竞争的着力点。
- **“从个人隐私安全层面看，大数据将网络大众带入到开放透明的裸奔时代**，数据安全若保护不利，将引发民情抱怨不满”。中国信息安全测评中心研究员磨惟伟表示。

大数据安全挑战

- **DT（数据技术）时代开放与安全的二元挑战**。在大数据获得开放的同时，也带来了对数据安全的隐忧。**大数据安全是‘互联网+’时代的核心挑战**，安全问题具有线上和线下融合在一起的特征。
- **传统解决网络安全的基本思想是划分边界**，在每个边界设立网关设备和网络流量设备，用守住边界的办法来解决安全问题。但**随着移动互联网、云服务的出现，网络边界实际上已经消亡了**。
- 信息安全的危险正在进一步升级，在APT、DDos、异常风险、网络漏洞等网络威胁下，**传统防御型、检测型的安全防护措施已经力不从心**，无法适应新形势下的要求。

大数据安全挑战

- **难以用有效的方式向用户申请权限，实现角色预设；难以检测、控制开发者的访问行为，防止过度的大数据分析、预测和连接。**
- 大数据时代，很多数据在收集时并不知道用途是什么，往往是二次开发创造了价值，**公司无法事先告诉用户尚未想到的用途，而个人也无法同意这种尚是未知的用途**。

上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

提纲

一、大数据分析挖掘的价值

二、大数据带来的网络安全和用户隐私问题

三、大数据带来的网络安全和用户隐私问题对策

大数据依托的NoSQL缺乏数据安全机制

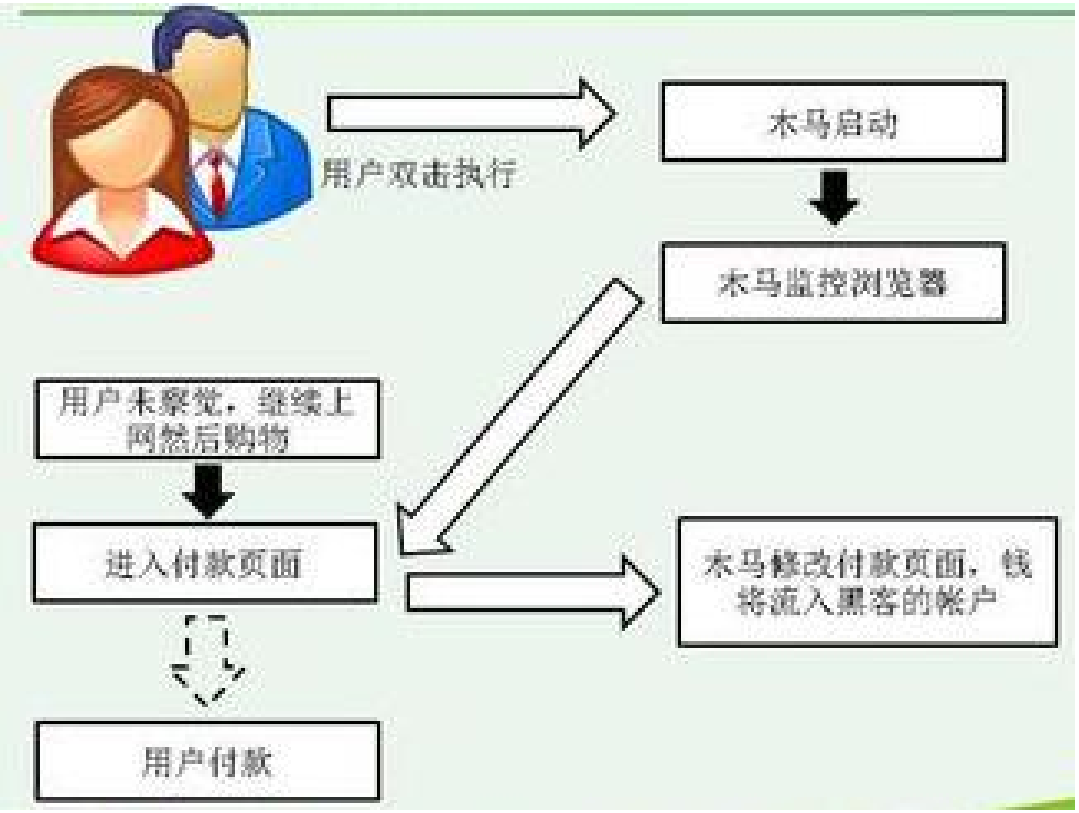
- **从基础技术角度来看，大数据依托的基础技术是NoSQL（非关系型数据库）。**
- 当前广泛应用的SQL（关系型数据库）技术，经过长期改进和完善，在维护数据安全方面已经设置严格的访问控制和隐私管理工具。而在NoSQL技术中，并没有这样的要求。
- 大数据数据来源和承载方式多种多样，如物联网、移动互联网、PC以及遍布地球各个角落的传感器，**数据分散存在的状态，使企业很难定位和保护所有这些机密数据。**
- **NoSQL允许不断对数据记录添加属性，其前瞻安全性变得非常重要，**对数据库管理员提出新的要求。

社会工程学攻击带来的安全问题

- 美国黑客凯文·米特尼克给出较全面的定义:“**社会工程学攻击是通过心理弱点、本能反应、好奇心、信任、贪婪等一些心理陷阱进行的诸如欺骗、伤害、信息盗取、利益谋取等对社会及人类带来危害的行为。**”其特点:**无技术性；成本低；效率高。**
- 该攻击与其他攻击最大的不同是其攻击手段不是利用高超的攻击技术，而是利用受害者的心理弱点进行攻击。**因为不管大数据多么庞大总也少不了人的管理，如果人的信息安全意识淡薄，那么即使技术防护手段已做到无懈可击，也无法有效保障数据安全。**由于大数据的海量性、混杂性，攻击目标不明确，因此攻击者为了提高效率，经常采用社会工程学攻击。

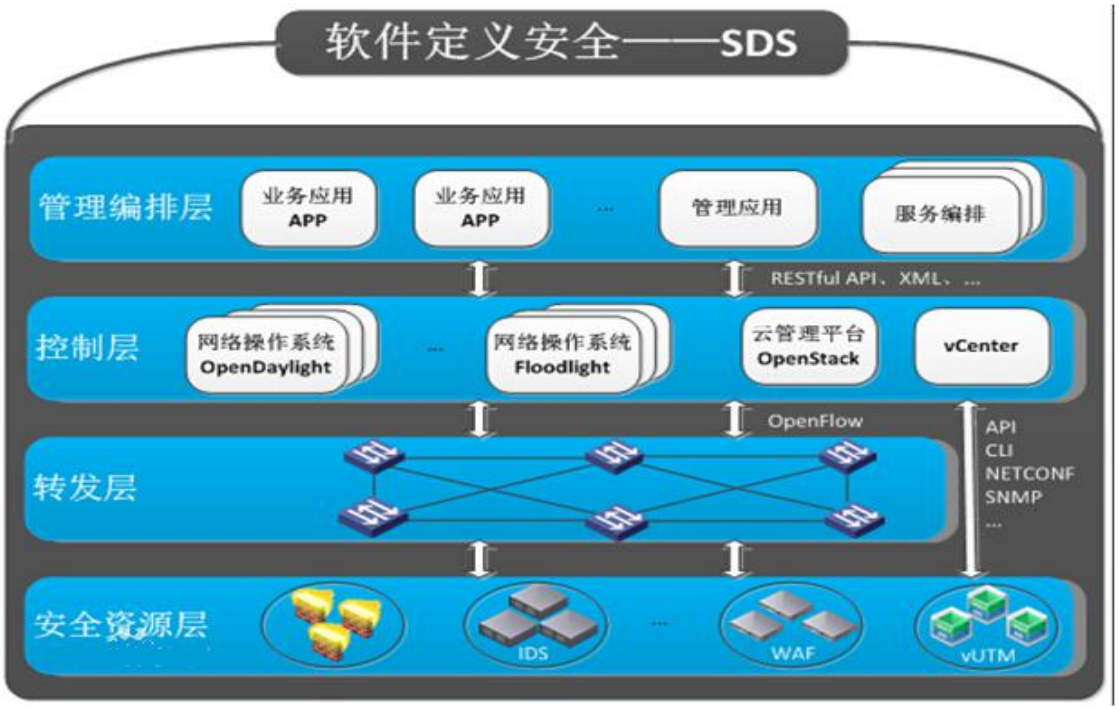
社会工程学攻击带来的安全问题

- 该类攻击的案例很多，如黑客先攻击某论坛的网站，使用户无法正常登陆。然后再假冒管理员，以维护网站名义向用户发送提醒信息，索要用户的账号和密码，一般用户此时会将密码和账号发送给黑客。此外，还有采用冒充中奖、假冒社交好友、信月卡挂失等欺诈手段获得合法用户信息。



软件后门成为大数据安全软肋

- 中科同向信息技术有限公司总经理邬玉良说:“**在这个‘软件定义世界’的时代，软件既是IT系统的核心，也是大数据的核心，几乎所有的后门都是开在软件上。**”
- 据了解，IBM、EMC等各大巨头生产制造的存储、服务器、运算设备等硬件产品，**几乎都是全球代工的，在信息安全的监听方面是很难做手脚的。**换句话说，**软件才是信息安全的软助所在。**



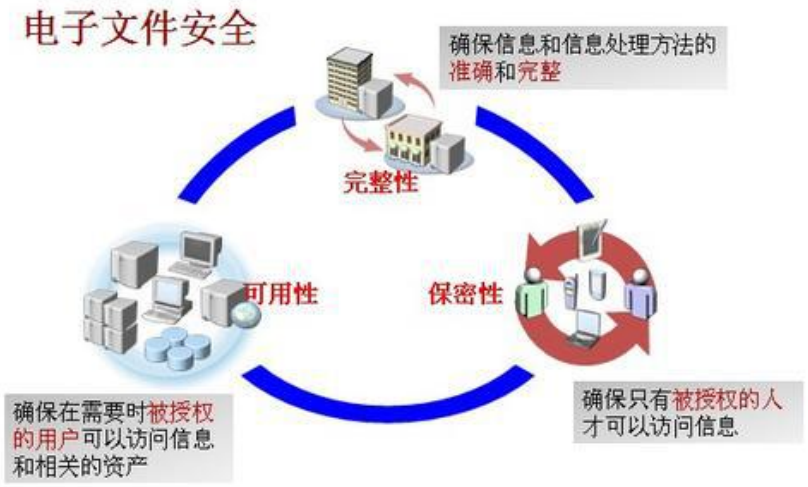
软件后门成为大数据安全软肋

- 软件供应方在主板上加入特殊的芯片，或是在软件上设计了特殊的路径处理，**检测人员只按照协议上的功能进行测试，根本就无法察觉软件预留的监听后门**。换言之，如果没有自主可控的信息安全检测方案，各种安全机制和加密措施，就都是形同虚设。
- **对于现代信息安全而言，最危险的行为是将自主控制的权力交给“他人”**。这就好比将自家的钥匙全部交到了外人手里，安全问题又从何谈起呢?

iPhone手机主要后门示意图

文件安全面临极大挑战

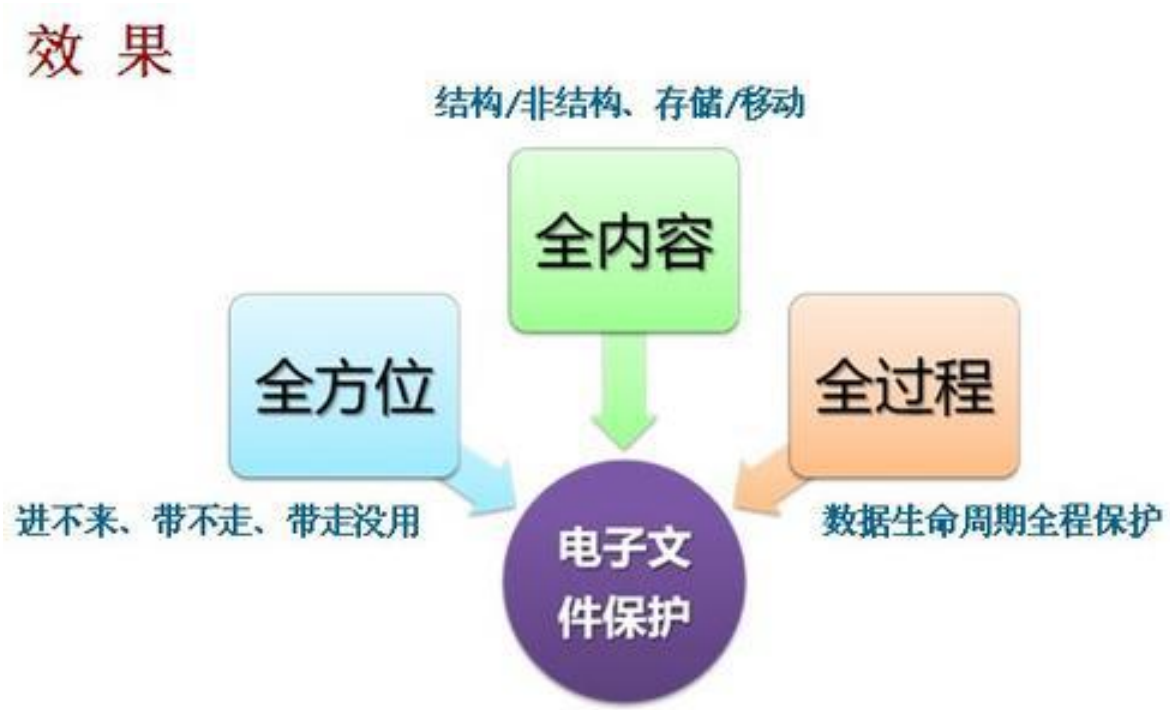
- **文件是数据处理和运行的核心**，大多的用户文件都在第三方运行平台中存储和进行处理，**这些数据文件往往包含很多部门或个人的敏感信息，其安全性和隐私性自然成为一个需要重点关注的问题**。
- **尽管文件保护提供了对文件的访问控制和授权**，例如Linux自带的文件访问控制机制，通过文件访问控制列表来限制程序对文件的操作。**然而大部分文件保护机制都存在一定程度的安全问题，它们通常使用操作系统的功能来实现完整性验证机制，因此只依赖于操作系统本身的安全性**。



文件安全面临极大挑战

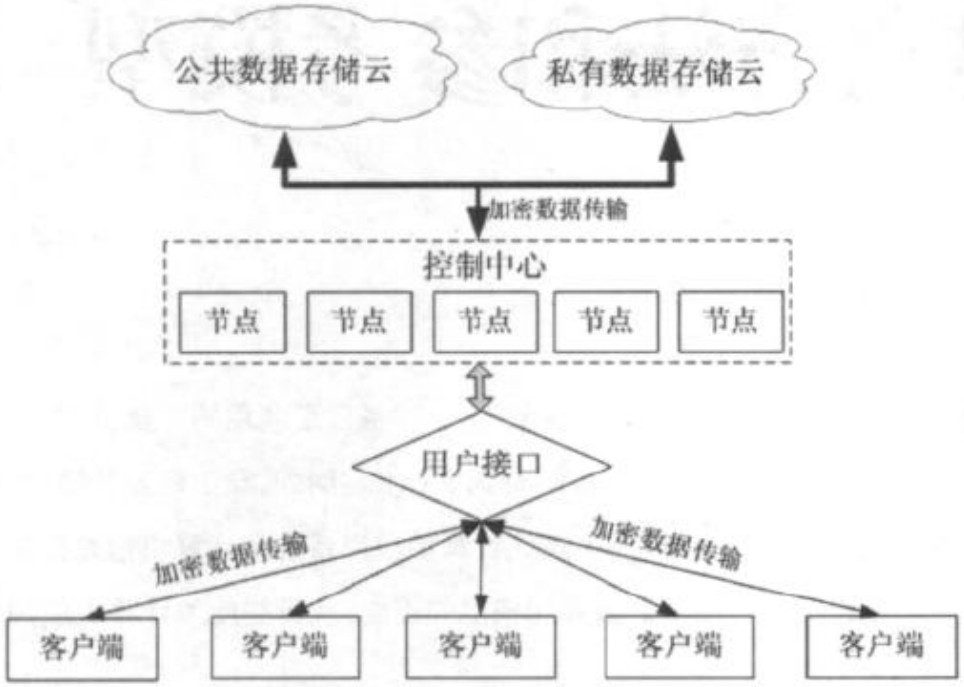
❑现代操作系统由于过于庞大，不可避免地存在安全漏洞，其本身的安全性都难以保证。

❑ 基于主机的文件完整性保护方法将自身暴露在客户机操作系统内，隔离能力差，恶意代码可以轻易发现检测系统并设法绕过检测对系统进行攻击。例如Tripwire，它本身是用户级应用程序，很容易被恶意软件篡改和绕过。



大数据存储安全问题

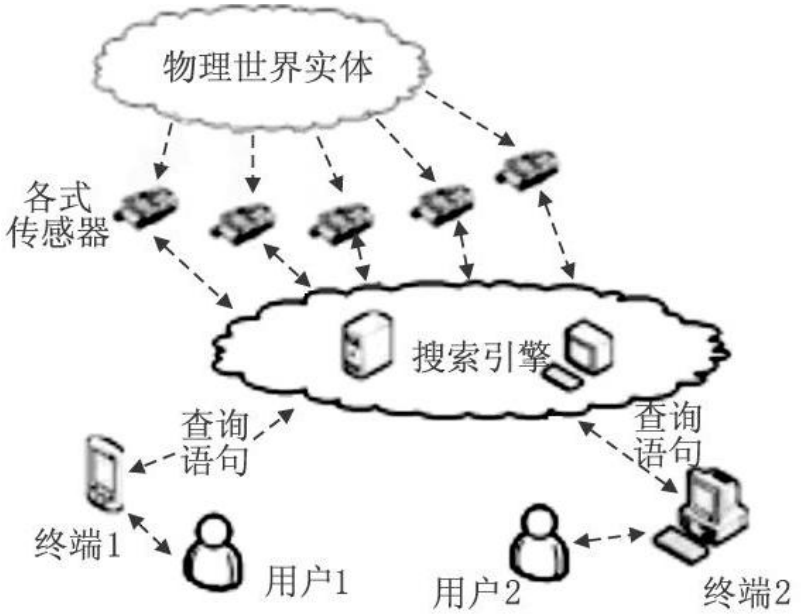
- 大数据会使数据量呈非线性增长，而复杂多样的数据集中存储在一起，多种应用的并发运行以及频繁无序的使用状况，有可能会出现数据类别存放错位的情况，**造成数据存储管理混乱或导致信息安全管理不合规范**。
- 现有的存储和安全控制措施无法满足大数据安全需求，安全防护手段如果不能与大数据存储和应用安全需求同步升级更新，就**会出现大数据存储安全防护的漏洞**。



基于云计算的安全数据存储系统体系结构

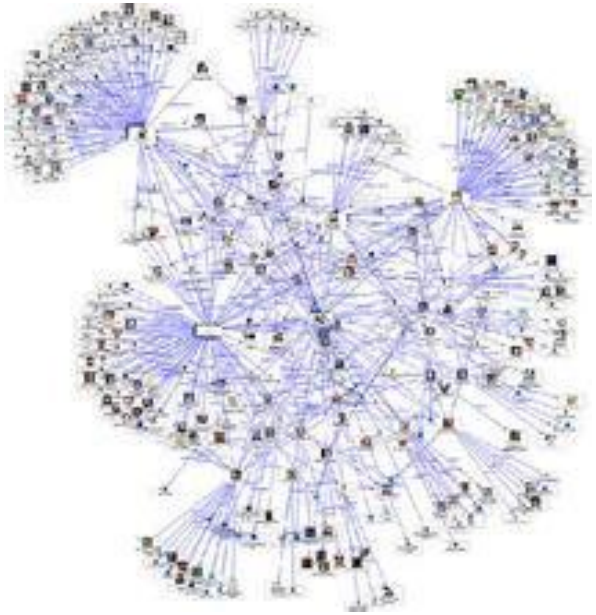
大数据安全搜索挑战和问题

- 需要更高效更智慧的分割数据，搜索、过滤和整理信息的理论与技术，以应对大数据越来越庞大的处理量，特别是实时性数据变化快，以及非结构化数据品种多。大数据安全搜索服务将上述浩瀚数据整理分类，可以帮助人们更快更高效地从中找到所需要的内容和信息。
- **大数据安全搜索挑战涉及通信网络的安全、用户兴趣模型的使用安全和私有数据的访问控制安全，**包括传统搜索过程中可能出现的网络安全威胁，比如相关信息在网络传输时被窃听以及恶意木马、钓鱼网站等，也包括服务器端利用通信网络获取用户隐私的危险。



大数据安全搜索挑战和问题

- 面向物联网实体搜索的安全和隐私挑战。当传感器和电子标签成为每个实物的附属物时，用户可能都不知道它们的存在。
- 大数据安全搜索问题包括：
  - 泛在尺寸不可见物联网实体搜索安全和隐私保护
  - 倒排表索引数据隐私安全
  - 私有数据的访问控制安全
  - 远程数据库安全搜索协议



基于大数据的威胁发现技术挑战

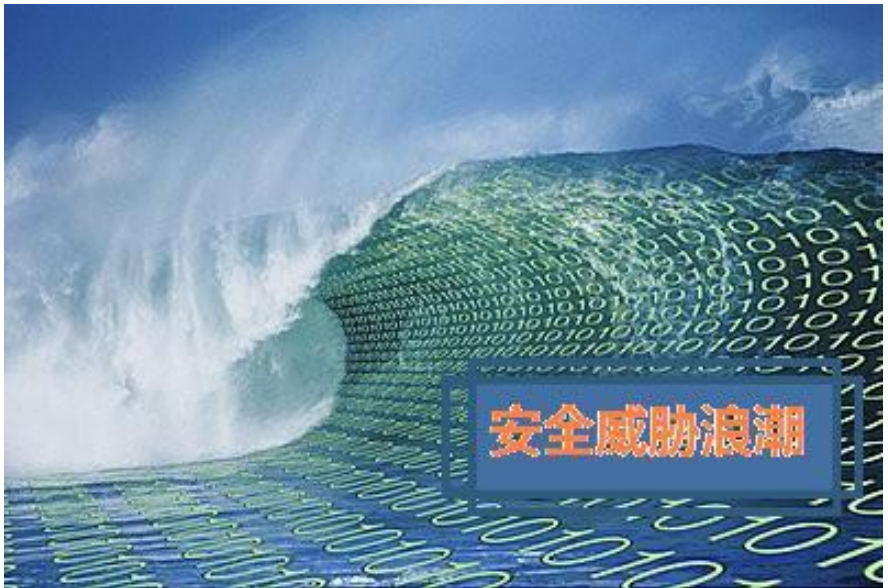
❑ 冯登国局长认为，**“棱镜”计划可被理解为应用大数据方法进行安全分析的成功故事。**通过收集各个国家各种类型的数据，利用该技术发现潜在危险局势，在攻击发生之前识别威胁。

❑ 基于大数据的威胁发现技术虽然具有上述的优点，但它目前存在一些挑战：

–一方面，**大数据的收集很难做到全面**，它的片面性会导致分析结果的偏差。为了分析企业信息资产面临的威胁，不但要全面收集企业内部的数据，还要对一些企业外的数据进行收集，这些在某种程度

上是一个大问题。

–另一方面，**大数据分析能力的不足影响威胁分析的准确性**。

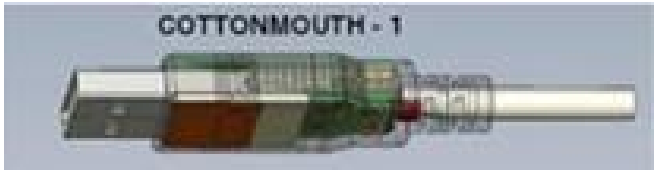


大数据带来的高级可持续攻击挑战

- 传统的检测是基于单个时间点进行的基于威胁特征的实时匹配检测，而**高级可持续攻击（APT）是一个实施过程，无法被实时检测**。
- 此外，大数据的价值低密度性，使得安全分析工具很难聚焦在价值点上，**黑客可以将攻击隐藏在大数据中，给安全服务提供商的分析制造很大困难**。黑客设置的任何一个会误导安全厂商目标信息提取和检索的攻击，都会导致安全监测偏离应有方向。

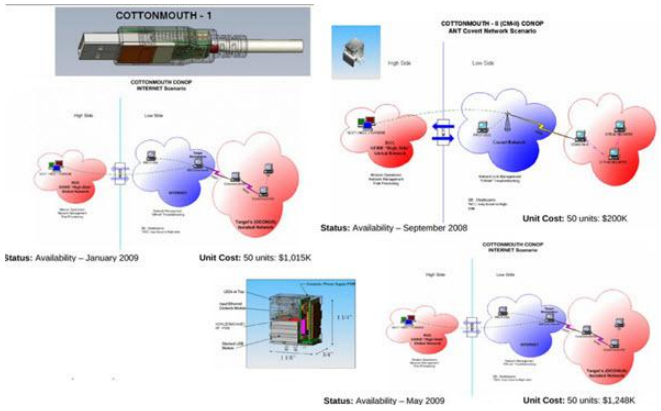
瞒天过海的APT攻击 - 量子网攻

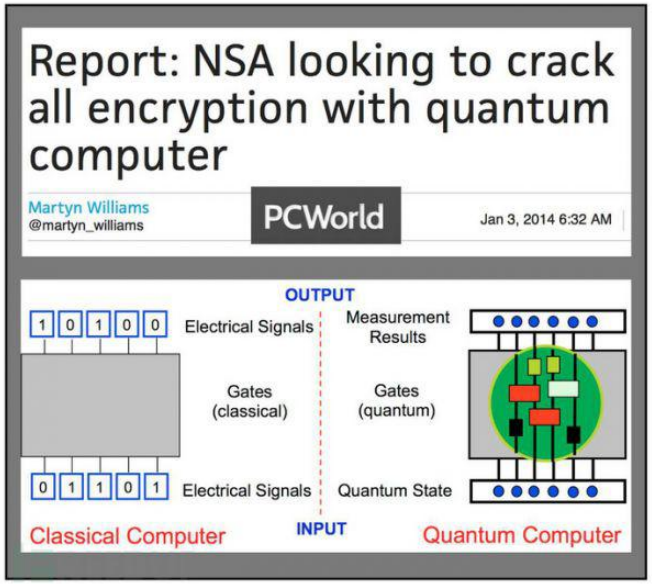
- 美国《纽约时报》2014年1月15日曝光了美国网络战新技术量子项目，**可入侵未联网的电脑、iPhone和大型网络服务器**，从2008年开始，**已操控全球10万台计算机，主要窃密中俄核心军方网络**。
- **“量子”计划的诞生，意味着个人隐私的终结，全球信息安全陷入危机。**
- 有一款代号为“棉口蛇1代”(Cottonmouth- I)的设备，它的外形象普通USB设备，但内部藏有微型无线电收发机。当它接入电脑时，能“通过秘密频道”传送从电脑中搜集的信息，从而让“数据悄然进出”。“棉口蛇1代”利用了USB设备的漏洞，**由于供应链植入是NSA监听全球的重要手段，而USB设备又无所不在，理论上任何企业和个人都难以防范。**



瞒天过海的APT攻击 - 量子网攻

- NSA“量子”项目曝光：可入侵未联网的电脑
  - 值得注意的是，在2008年至2010年夏天美国对伊朗核设施采取的网络攻击中，美国就利用了这项技术向伊朗核设施植入“震网”病毒，这也是该技术第一次参与实战
  - 量子网攻最重要的监控对象便是中国军方。美国情报机构已将设在上海的61398部队作为网络攻击目标，这一机构被认为专门负责对美方发动网络攻击。这和美起诉5名中国军官有没有关系呢？





瞒天过海的APT攻击 - 量子网攻

量子网攻原理，“无线电传输”，老技术的新用法：它利用插入计算机的微型电路板或移动存储器，通过发射某种频率的无线电波，将计算机信息传输给中继站，并最终汇总到NSA，见下图。

1. 植入微型发射器的USB接口插入被监听的计算机，也可在被监听的计算机中植入微型电路板。
2. 微型发射器将被监听计算机中的数据传输到13千米之外的国家安全局公文包大小的中继站。
3. 中继站将信息传回国家安全局操控中心。
4. 中继站可以将间谍程序植入被监听的计算机，包括攻击伊朗核设施的“震网”病毒。

瞒天过海的APT攻击 - 量子网攻

- **“量子”计划利用无线电传输技术建立的信道几乎独立于计算机自身的数据传输体系，这让传统的电脑安全机制失去了用武之地。**这种利用硬件的物理接触的方法看似传统，实际上防不胜防。因为电子设备生产、运输和销售的任一环节都有被这种微型传输器无意间接触的可能，从而加大了防范难度。
- **防御不可能一劳永逸：**不可能凭借国产化或者改善某个技术环节就可以一劳永逸地进行防御。
- **大数据时代窃密效率大增，防御难度增大：**中国互联网协会信息安全专委会委员肖新光指出，“4G带来的带宽增长，也导致了传统窃密中信息外传的速度瓶颈消失，一旦内部信息系统被突破，信息窃取的效率会大大增加。”

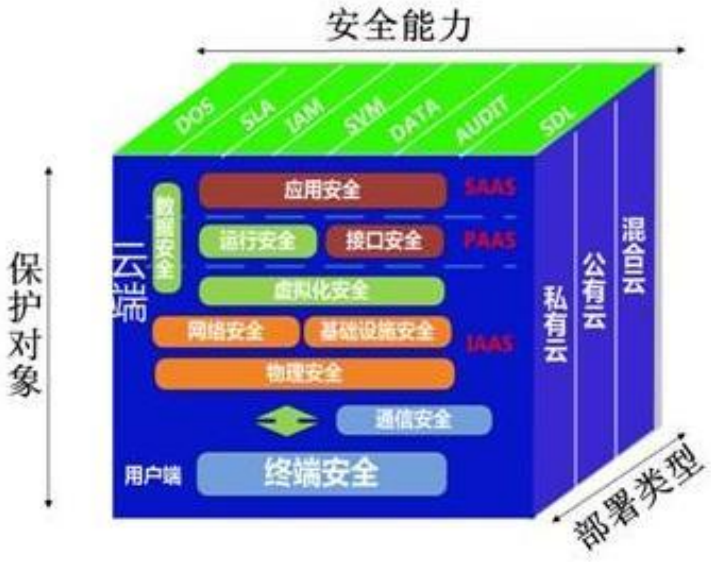
大数据支撑平台—云计算安全

- **云计算的核心安全问题是用户不再对数据和环境拥有完全控制权**
  - 云计算的出现彻底打破了地域的概念，数据不再存放在某个确定的物理节点，而是由服务商动态提供存储空间，这些空间有可能是现实的，也可能是虚拟的，还可能分布在不同国家及区域。
  - 用户对存放在云中的数据不能像从前那样具有完全的管理权，相比传统的数据存和处理方式，云计算时代的数据存储和处理，对于用而言，变得非常不可控。

大数据支撑平台—云计算安全

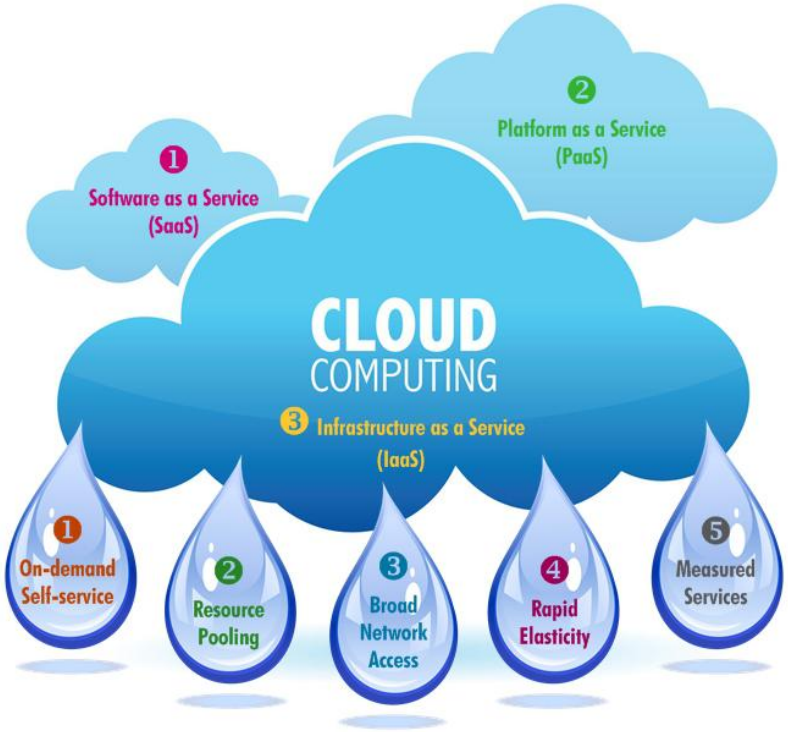
云环境中用户数据安全与隐私保护难以实现

- 在云计算环境下，各类云应用不再依靠机器或网络形成固定不变的基础设施物理边界和安全边界，数据安全由云计算提供商负责。
- 多用户应用场景中，**应用和数据库都部署在非完全可信的服务运营商端，**服务运营商可能会由于经济利益等原因,将用户的敏感客户信息泄漏给第三方,第三方可以利用这些信息进行广告投放、商品推销等活动,给客户生活、工作造成困扰。
- 使用传统的数据加密和数据混淆保护方法使得数据处理效率相对较低，违背了使用云计算的初衷，**需要采用一种新的数据隐私保护方法，实现隐私保护与数据处理性能的有效结合。**



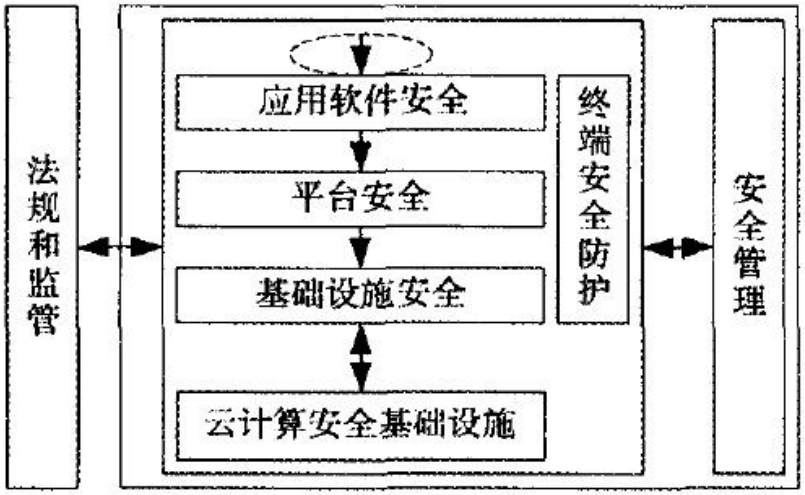
大数据支撑平台—云计算安全

- 云计算中多层服务模式同样存在安全隐患
  - 云计算发展的趋势之一是IT 服务专业化， 即云服务商在对外提供服务的同时，自身也需要购买其他云服务商所提供的服务
  - 用户所享用的云服务间接涉及到多个服务提供商，多层转包无疑极大地提高了问题的复杂性，进一步增加了安全风险



大数据支撑平台—云计算安全

- 虚拟运算平台的安全漏洞不断涌现，直接威胁云安全根基
  - 云端大量采用虚拟技术，虚拟平台的安全无疑关系到云体系的架构安全
  - 虚拟运算平台变得越来越复杂和庞大、管理难度也随之增大，如果黑客利用安全漏洞获得虚拟平台的管理控制权，后果将不堪设想



云计算安全体系架构模型（摘自云安全联盟）

大数据用户隐私保护考验

- 根据受保护对象的不同，可以将隐私保护分为三类，即位置保护、连接关系保护以及标识符保护。
- **现在很多企业认为只要将信息进行匿名处理，公布不含有用户标识符的信息，就能够实现对用户的隐私进行保护，然而事实证明，这种做法取得的保护效果并不理想。**
- 目前对用户的数据进行采集、储存、使用以及管理等工作时，**均缺乏相应的标准、规范以及监管，对企业自律性过于自信以及依赖。**此外，**用户并不会被告知其隐私信息被用于何处。**
- 数据动态性。**现有隐私保护技术主要基于静态数据集，**而在现实中数据模式和数据内容时刻都在发生着变化。因此**在这种更加复杂的环境下实现对动态数据的利用和隐私保护将更具挑战。**

大数据分析预测带来的用户隐私挑战

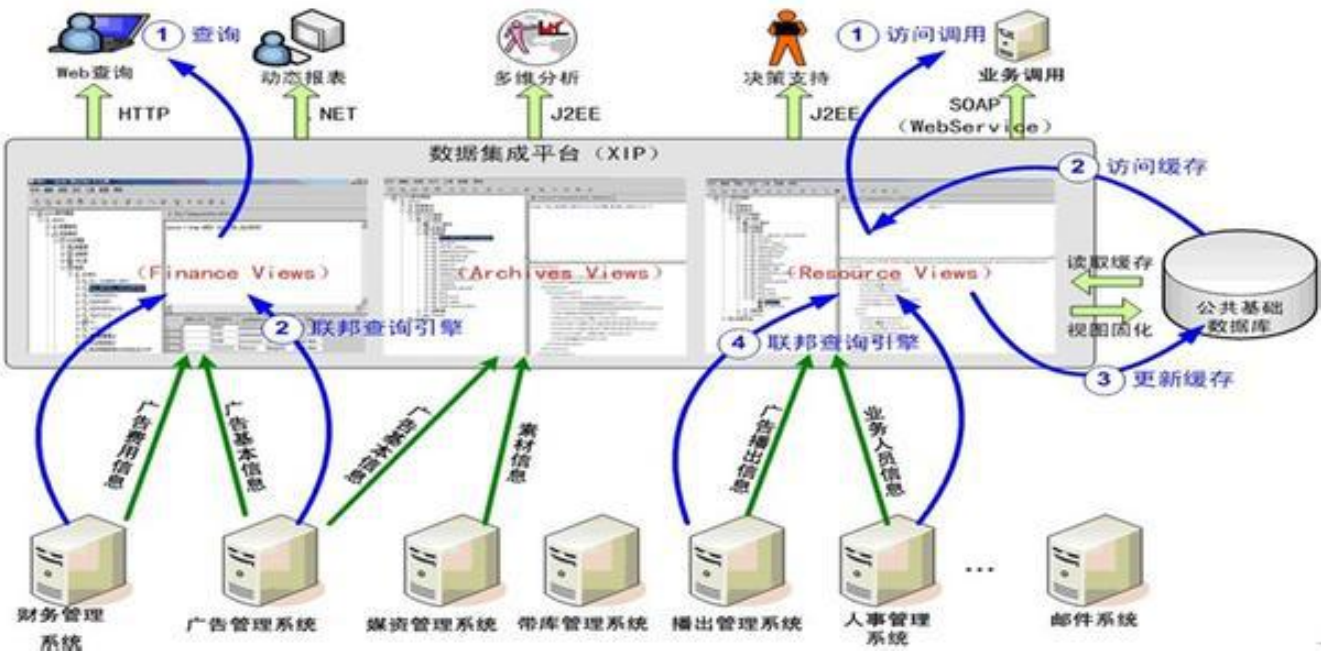
- 从核心价值角度来看，大数据关键在于数据分析和利用，但数据分析技术的发展，对用户隐私产生极大的威胁。
- 在大数据时代，想屏蔽外部数据商挖掘个人信息是不可能的。目前，各社交网站均不同程度地开放其用户所产生的实时数据，被一些数据提供商收集，还出现了一些监测数据的市场分析机构。
- 通过人们在社交网站中写入的信息、智能手机显示的位置信息等多种数据组合，已经可以以非常高的精度锁定个人，挖掘出个人信息体系，用户隐私安全问题堪忧。

安全、隐私和便利之间的冲突

- 大数据对个人信息获取渠道的拓宽需求引发了一个重要问题：**安全、隐私和便利性之间的冲突**。**消费者受惠于海量数据：更低的价格、更符合消费者需要的商品以及从改善健康状况到提高社会互动顺畅度等。同时，随着个人购买偏好、健康和财务情况的海量数据被收集，人们对隐私的担忧也在增大。**
- “棱镜门”事件爆发后，尴尬的美国总统奥巴马辩解道：“你不能在拥有100%安全的情况下，同时拥有100%隐私和 100%便利。”

大数据共享安全性问题

- **我们不知道该如何分享私人数据，才能既保证数据隐私不被泄漏，又保证数据的正常使用。**
- 真实数据不是静态的，而是越变越大，并且随着时间的变化而变化。当前没有一种技术能在这种情况下产生任何有用的结果。
- **许多在线服务要求我们共享私人信息，**但是，在记录级的访问控制之外，**我们根本不知道共享数据会意味着什么，**不知道共享后的数据会怎样被连接起来，**更不知道如何让用户对共享后的数据仍能进行细粒度控制。**



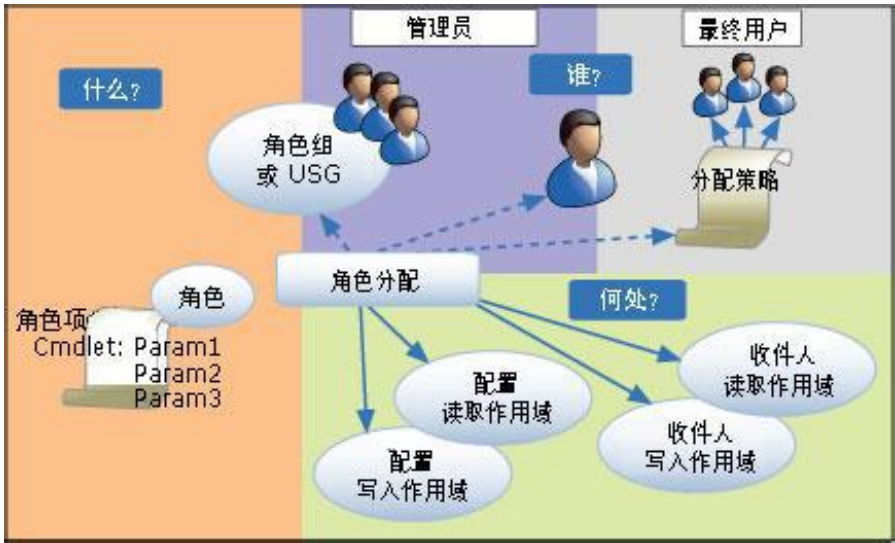
大数据访问控制难题

- **访问控制是实现数据受控共享的有效手段．由于大数据可能被用于多种不同场景，其访问控制需求十分突出。**
- 难以预设角色，实现角色划分．由于大数据应用范围广泛，它通常要为来自不同组织或部门、不同身份与目的的用户所访问，实施访问控制是基本需求。然而，在大数据的场景下，有大量的用户需要实施权限管理，且用户具体的权限要求未知。**面对未知的大量数据和用户，预先设置角色十分困难。**

大数据访问控制难题

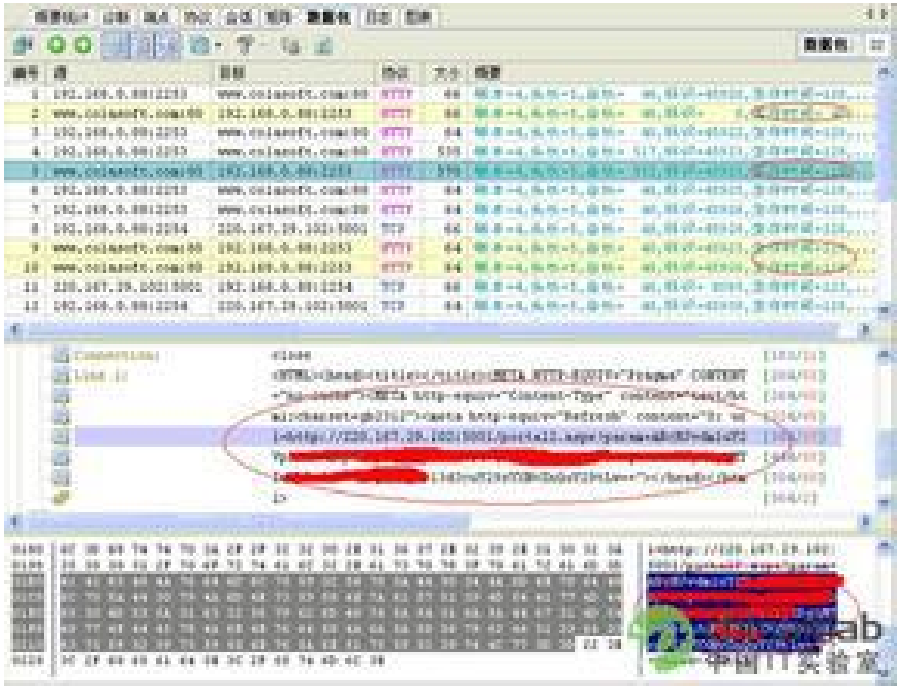
- **难以预知每个角色的实际权限。**面对大数据，安全管理员可能无法准确为用户指定其可以访问的数据范围，而且这样做效率不高。比如，医生为了完成其工作可能需要访问大量信息，但对于数据能否访问应该由医生来决定，不应该需要管理员对每个医生做特别的配置．但同时又应该能够提供对医生访问行为的检测与控制，限制医生对病患数据的过度访问。
- 不同类型的大数据存在多样化的访问控制需求．例如，在Web 2.0个人用户数据中，存在基于历史记录的访问控制；在地理地图数据中，存在基于尺度以及数据精度的访问控制需求；在流数据处理中，存在数据时间区间的访问控制需求，等。

如何统一地描述与表达访问控制需求是一个挑战性问题。



大数据的可信性难以保障

- **网络的数据并非都可信，这主要反映在伪造的数据和失真的数据2个方面。**有人可能通过伪造数据来制造假象，进而对数据分析人员进行诱导；或者数据在传播中逐步失真。**这可让大数据分析和预测得出无意义或错误的结果。冯登国局长认为，用信息安全技术手段鉴别所有数据来源的真实性是不可能的.**
- 过去往往认为“有图有真相”，事实上图片可以移花接木、时空错乱，或者照片是对的，可是文字解释是捏造的。



大数据的可信性难以保障

❑**邬贺铨院士指出，传感器收集的数据并非都是可信的，**特别是历史上该传感器的数据与同类的其他传感器报出的数据差异很大时，该数据就应弃用。

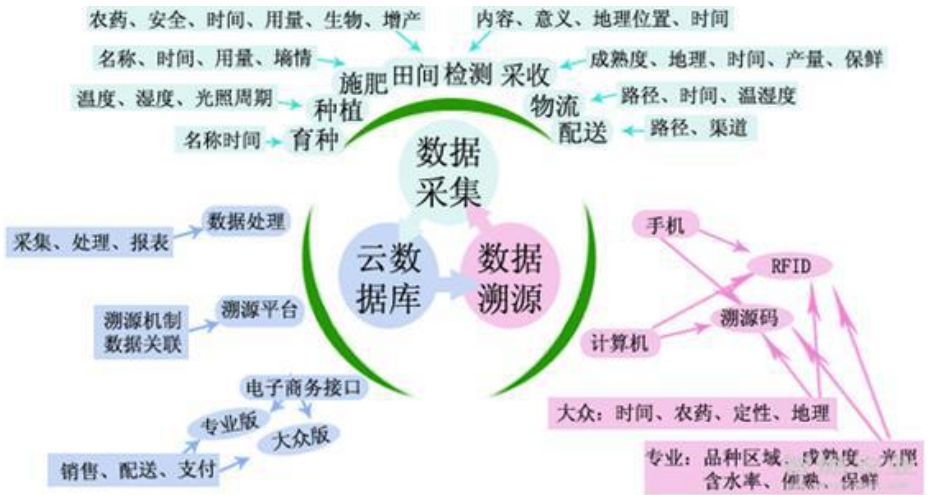
❑ **密码学中的数字签名、消息鉴别码等技术可用于验证数据的完整性，但应用于大数据的真实性时面临很大困难，主要根源在于数据粒度的差异。**例如，数据的发源方可以对整个信息签名，但是当信息分解成若干组成部分时，该签名无法验证每个部分的完整性。而数据的发源方无法事先预知哪些部分被利用、如何被利用，难以事先为其生成验证对象。

大数据溯源技术的安全应用挑战

数据溯源技术：它旨在帮助使用者确定数据的来源，进而检验分析结果是否正确，或对数据进行更新。**2009年，数据溯源技术被相关报告列为三大确保国家安全的重要技术之一，其在未来数据信息安全领域中仍具有很大的发展空间。**数据溯源技术应用于大数据安全与隐私保护中还面临如下挑战：

（1）**大数据溯源与隐私保护之间的平衡。**一方面，基于数据溯源对大数据进行安全保护首先要通过分析技术获得大数据的来源，然后才能更好地支持安全策略和安全机制的工作；另一方面，数据来源往往本身就是隐私敏感数据。用户不希望这方面的数据被分析者获得。因此，如何平衡这两者的关系是值得研究的问题之一。

（2）**大数据溯源技术自身的安全性保护**。当前数据溯源技术并没有充分考虑安全问题，例如标记自身是否正确、标记信息与数据内容之间是否安全绑定等等。而在大数据环境下，其大规模、高速性、多样性等特点使该问题更加突出。



上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

提纲

一、大数据分析挖掘的价值

二、大数据带来的网络安全和用户隐私问题

三、大数据带来的网络安全和用户隐私问题对策

基于大数据的威胁发现技术

- 利用该技术，企业可以超越以往的“保护—检测—响应—恢复”（PDRR）模式，更主动地发现潜在的安全威胁。
- 相比于传统技术，基于大数据的威胁发现技术有以下优点：

- **分析内容的范围更大。**企业信息资产包括数据资产、软件资产、实物资产、人员资产、服务资产和其它为业务提供支持的无形资产。由于传统威胁检测技术并不能覆盖这六类信息资产，因此所能发现的威胁有限。而通过在威胁检测方面引入大数据分析技术，能全面发现针对这些信息资产的攻击。

- **分析内容的时间跨度更长。**现有威胁分析技术具有内存关联性，即实时收集数据，采用分析技术发现攻击。分析窗口通常受限于内存大小，无法应对持续性和潜伏性攻击。而引入大数据分析技术后，威胁分析窗口可以横跨若干年的数据，因此威胁发现能力更强，**可以有效应对APT类攻击**。

基于大数据的威胁发现技术

❑ 相比于传统技术，基于大数据的威胁发现技术还有以下优点：

- **攻击威胁的预测性。**传统安全防护技术大多是在攻击发生后对攻击行为进行分析和归类，并做出响应。而基于大数据的威胁分析，可进行超前的预判，对未发生的攻击行为进行预防。

- **对未知威胁的检测。**传统的威胁分析常由经验丰富的专业人员根据企业需求和实际情况展开，威胁分析结果很大程度上依赖于个人经验，分析所发现的威胁是已知的。**而大数据分析的特点是侧重于**

普通的关联分析，
而不侧重因果分析，
因此通过采用恰当
的分析模型，可
发现未知威胁。

基于大数据的认证技术

❑ 基于大数据的认证技术指的是收集用户行为和设备行为数据，并对这些数据进行分析，获得用户行为和设备行为的特征，进而通过鉴别操作者行为及其设备行为来确定其身份．这与传统认证技术利用用户所知秘密，所持有凭证，或具有的生物特征来确认其身份有很大不同。该技术具有如下优点：

- **攻击者很难模拟用户行为特征来通过认证，因此更加安全。**利用大数据技术所能收集的用户行为和设备行为数据是多样的，可以包括用户使用系统的时间、经常采用的设备、设备所处物理位置，甚至是用户的操作习惯数据。通过这些数据的分析能够为用户勾画一个行为特征的轮廓。而攻击者很难在方方面面都模仿到用户行为，因此其与真正用户的行为特征轮廓必然存在一个较大偏差，无法通过认证。

基于大数据的认证技术

- 该技术还具有如下优点：

－减小了用户负担。用户行为和设备行为特征数据的采集、存储和分析都由认证系统完成。相比于传统认证技术，极大地减轻了用户负担。如，用户无需记忆复杂的口令，或随身携带硬件ＵＳＢＫｅｙ。

－可以更好地支持各系统认证机制的统一。基于大数据的认证技术可以让用户在整个网络空间采用相同的行为特征进行身份认证，而避免传统不同系统采用不同认证方式，且用户所知秘密或所持凭证各不相同而带来的种种不便。

基于大数据的数据真实性分析

- **目前，基于大数据的数据真实性分析被广泛认为是最为有效的方法。**许多企业已经开始了这方面的研究工作，如Ｙａｈｏｏ和Ｔｈｉｎｋｍａｉｌ等利用大数据分析技术来过滤垃圾邮件；Ｙｅｌｐ等社交点评网络用大数据分析来识别虚假评论；新浪微博等社交媒体利用大数据分析来鉴别各类垃圾信息等。
- 基于大数据的数据真实性分析技术能够提高垃圾信息的鉴别能力：

–一方面，引入大数据分析可以**获得更高的识别准确率。**例如，对于点评网站的虚假评论，可以通过收集评论者的大量位置信息、评论内容、评论时间等进行分析，鉴别其评论的可靠性。如果某评论者为某品牌多个同类产品都发表了恶意评论，则其评论的真实性就值得怀疑。

– 另一方面，在进行大数据分析时，通过机器学习技术，可以**发现更多具有新特征的垃圾信息**。然而**该技术仍然面临一些困难，主要是虚假信息的定义、分析模型的构建**等。

大数据带来的网络安全和用户隐私问题对策

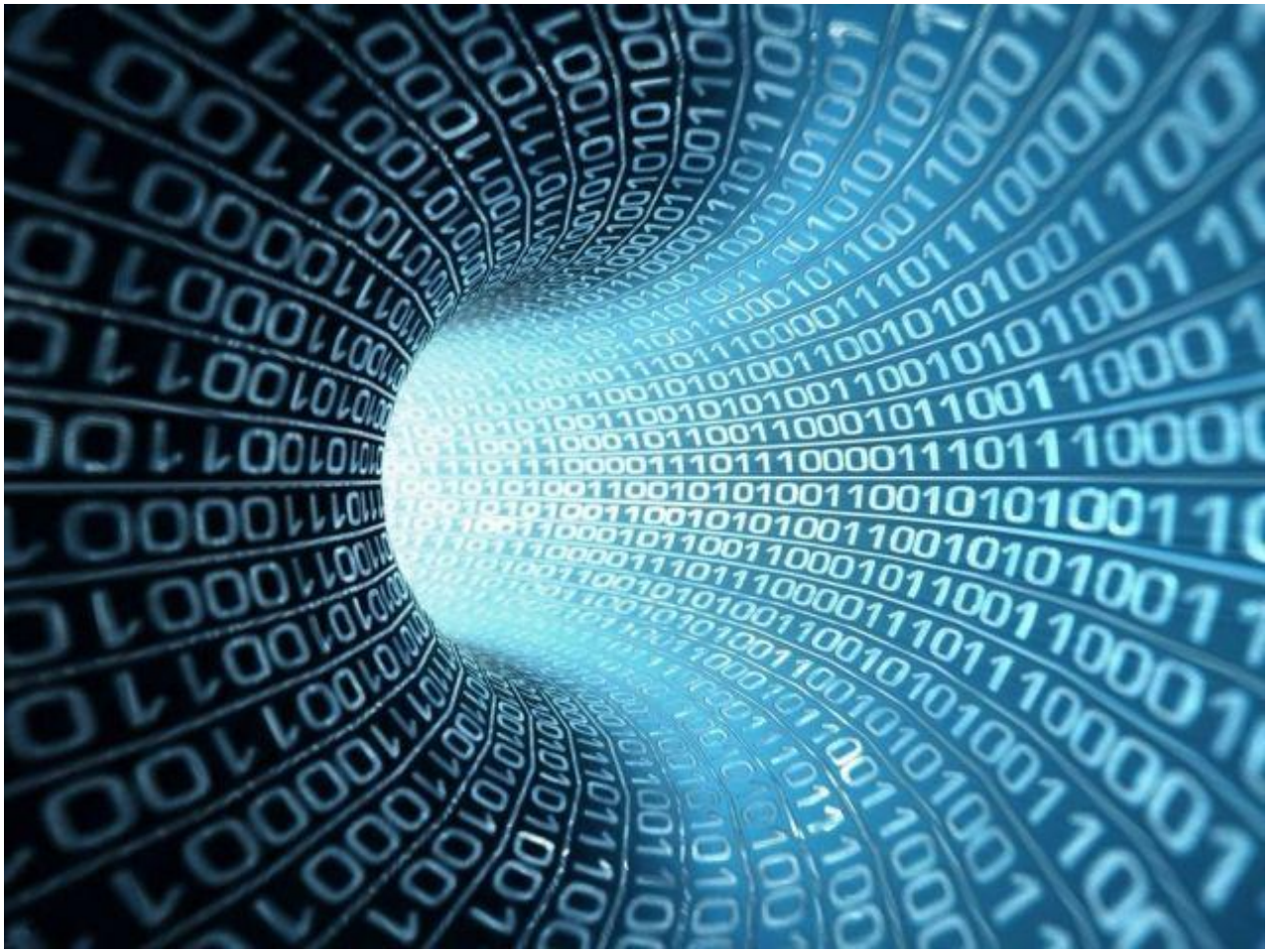
- 研究**大数据基础设施安全**能力的评估以及加强大数据框架下的安全技术，如数据标签法、Hadoop、NoSQL等，这些基础设施、基本技术，都将直接影响大数据下的信息安全。
- **推动信息安全的自主可控，提倡“可信计算”。**所谓的“可信计算”就是，软件不再做功能上的黑名单，而是换以白名单来进行控制。
- 围绕大数据突出的安全和隐私问题，**构建数据全生命周期的安全管理体系，**结合大数据处理体系的特点，尤其关注分布式环境下的并行计算隔离；分布式集群的数据访问控制；以及对敏感、重要数据的分级管控、加密处理和审计追踪等安全保障措施。
- **风险自适应的访问控制。**在大数据场景中，安全管理员可能缺乏足够的专业知识，无法准确地为用户指定其可以访问的数据。
针对这种场景讨论较多的一种访问控制方法。
- 在大数据环境下，**发展基于密码认证、攻防、风险控制、安全集成电路设计等信息安全技术。**

立法保障大数据安全

❑ 为了防止数据泄露，邬贺铨院士认为首先要从法律上约束制裁。“**大数据时代，开放数据和保护数据安全都需要通过立法来保证**，如果没有相应的法律，我们很难判断哪些数据应该共享，哪些数据不应泄露，谁可以用，谁不可以用，出了问题很难找出谁是幕后黑手，目前我们国家没有信息安全法，未来需要从法律上约束。”

❑ **搞网络安全不用灰头土脸，立了法就可以光明正大地干了，这是国际惯例。**

上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



谢 谢 ！