数据安全

第1章 引论

计算机与大数据学院 刘延华

福州大学
FUZHOU UNIVERSITY

数据无处不在

超级应用

交通系统 平安系统 城管系统 医疗系统 …

环境系统 旅游系统 智慧亚运 移动办事 …

区县中枢1

特色应用1
特色应用2
…
中枢平台

区县中枢2

特色应用1
特色应用2
…
中枢平台

区县中枢X

特色应用1
特色应用2
中枢平台

城市数据资源平台

城市数据大脑

城市算法服务平台

城市计算资源平台

行业系统

互联网数据 公共数据 企业数据 政务数据 …

一个统计

全球互联网用户每天：**Big Data**

- ✓ 发送电子邮件**2940**亿封
- ✓ 发送推文**5**亿条
- ✓ 在**Facebook**创建数据多达**4PB**
- ✓ 发送**WhatsApp**消息达到**650**亿条
- ✓ 发送微信消息达**500**亿条

预测明年全球数据总量将增长**10**倍，达到**44ZB**。

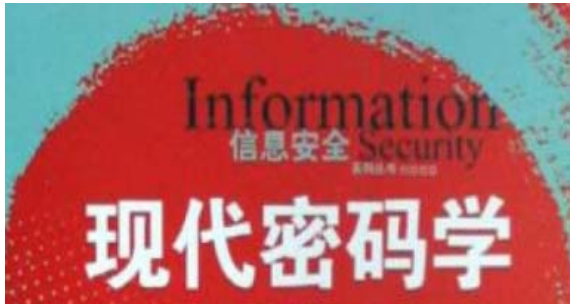
福州大学
FUZHOU UNIVERSITY

数字生活UP，形成高大数据墙！ 翻墙？

什么是安全？为什么不安全？

腾讯安全

奇安信
新一代网络安全领军者



第1章 引论

1.1 数据安全的概念

1.2 数据安全的态势

1.3 脆弱性，威胁，风险

1.4 数据安全标准

1.1 数据安全的概念

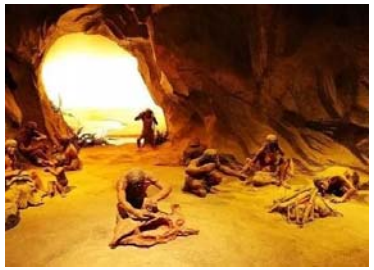
□术语：数据安全 data security

定义：以保护数据的保密性、可用性和完整性等为中心的安全。（GB/T 25069—2010）

□ 《中华人民共和国国家安全法》

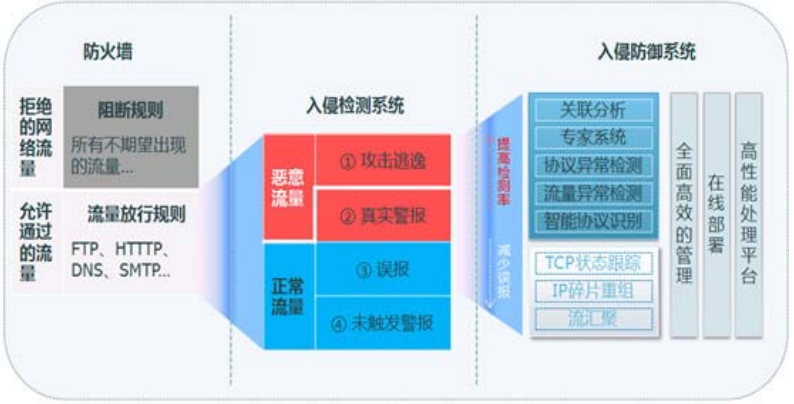
第二十五条中明确要求“实现网络和信息核心技术、关键基础设施和重要领域信息系统及数据的安全可控”。

1.1 数据安全的概念



□ 对数据安全的认识？

- ✓ 限制数据采集就能保护数据安全？ *回到远古*
- ✓ 精准服务等于隐私侵犯？ *精准推荐、精准广告*
- ✓ 网络安全+系统安全就能够数据安全问题？ *传统安全*
- ✓ 数据不流动，会更安全？
- ✓ ……



数据安全

福州大学
FUZHOU UNIVERSITY

1.1 数据安全的概念

- 从数据存在状态的角度，数据的安全可以认为包括**数据库管理下**的数据安全和**非数据库管理下**的数据安全两个部分。

- 在数据库管理模式下，数据的安全性依赖于**数据库管理系统**所采用的**安全策略**、**安全模型**和**安全机制**。一个成熟的商业化数据库管理系统能够为数据安全提供高强度的安全性。

- 数据在非数据库管理下，其安全性则由**使用用户或管理用户**来负责，如数据在**拷贝**、**传输**、**流转**、**交易**等环节中，往往没有数据库管理系统的安全约束。

对比来看，数据库管理下的数据安全性更高。

1.1 数据安全的概念

安全属性角度

□ 术语：可用性 availability

根据授权实体的要求可访问和可使用的特性。

[GB/T 29246—2017，定义2.9]

□ 术语：保密性 confidentiality

信息不能对未授权的个人、实体或过程可用或泄露的特性。

[GB/T 29246—2017，定义2.12]

□ 术语：数据完整性 data integrity

数据没有遭受以未授权方式所作的更改或破坏的特性。

[ISO/IEC 27040:2015，定义3.9]

1.1 数据安全的概念

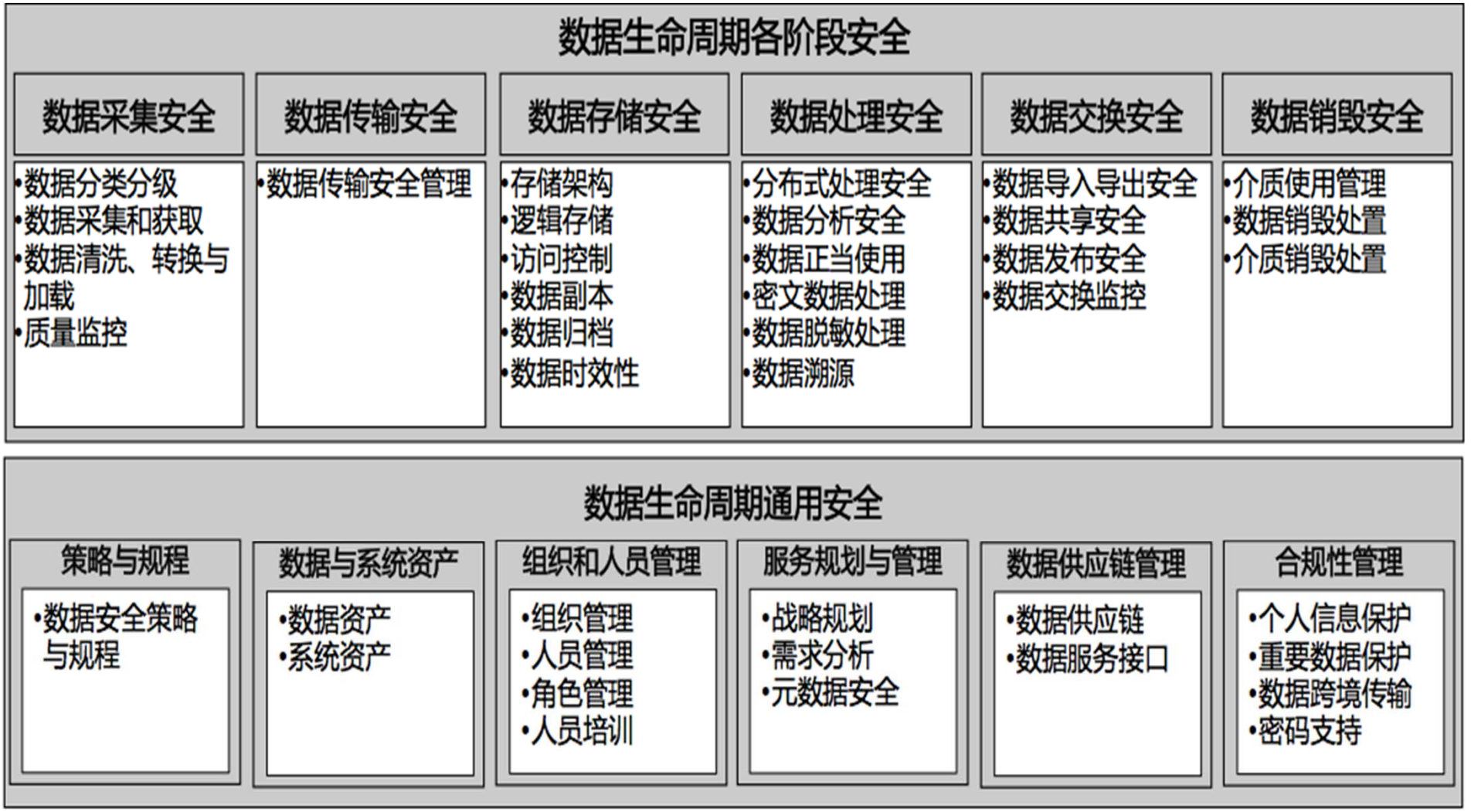
数据生命周期角度

□ 术语：数据生命周期 data lifecycle

数据从产生，经过数据采集、数据传输、数据存储、数据处理（包括计算、分析、可视化等）、数据交换，直至数据销毁等各种生存形态的演变过程。

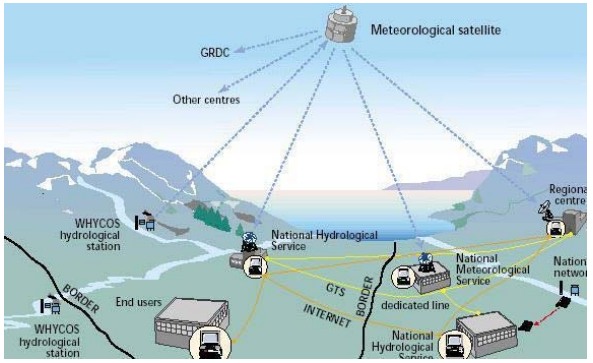
[GB/T 35274—2017，定义3.2]

从数据安全角度，应该对数据整个周期链进行有效安全监测与管理，这也是目前急需解决的重要任务。



国标**GB/T 37988**《信息安全技术 数据安全能力成熟度模型》

1.1 数据安全的概念



□ 数据采集安全

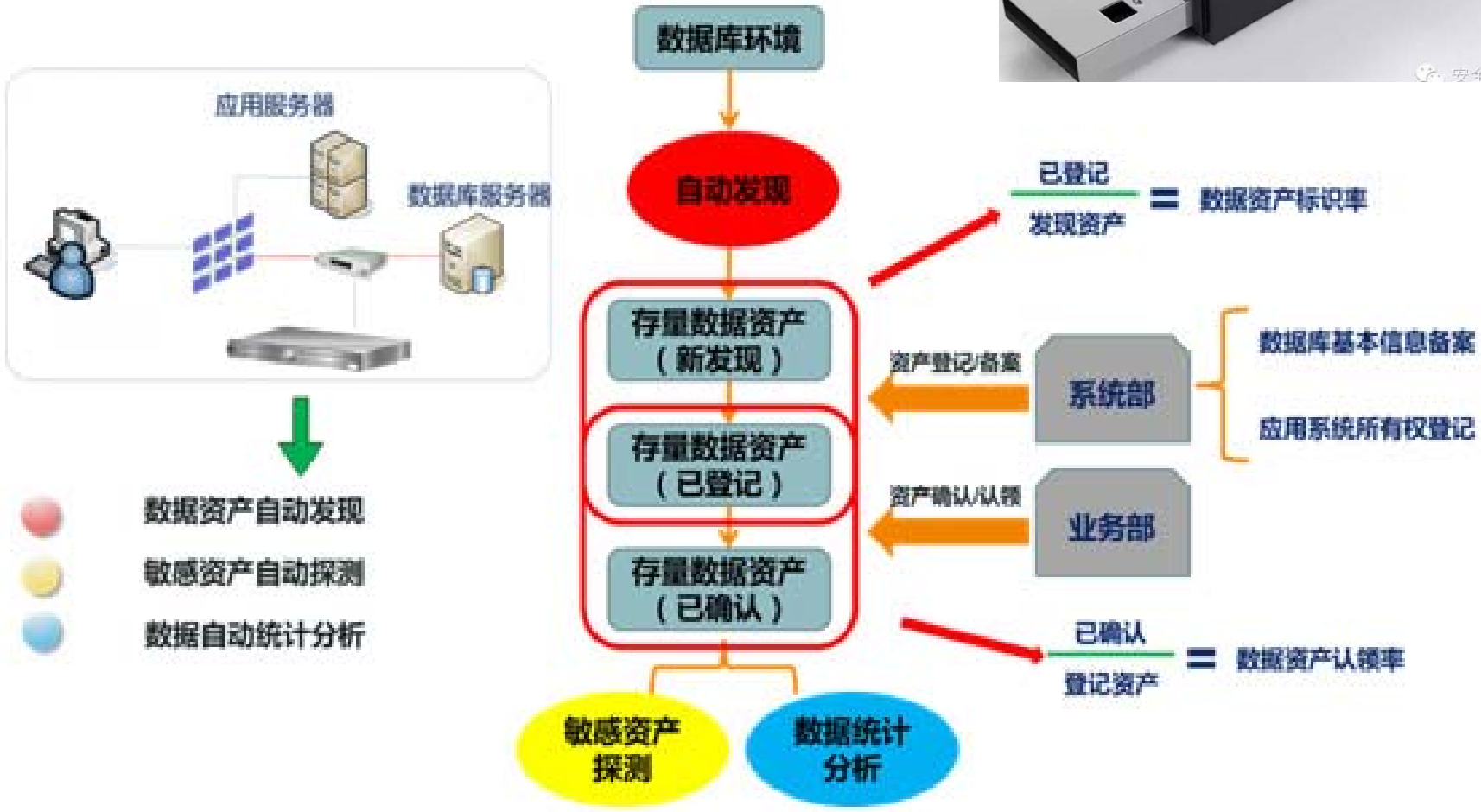
数据采集的可信性是一个重要关注点。其面临的安全威胁是数据被伪造或刻意制造，如虚假评论、数据粉饰等，可能诱导分析数据时得出错误结论，影响决策判断。

因此，如何对采集到的数据或属性进行识别、评估、去伪存真，提高识别非法数据源的技术能力，确保数据来源安全可信。

资产摸底

福州大学
FUZHOU UNIVERSITY

1.1 数据安全的概念


应用服务器
数据库服务器
数据资产自动发现
敏感资产自动探测
数据自动统计分析
数据库环境
自动发现
存量数据资产
（新发现）
存量数据资产
（已登记）
存量数据资产
（已确认）
敏感资产
探测
数据统计
分析
已登记
发现资产
=
数据资产标识率
资产登记/备案
系统部
数据库基本信息备案
应用系统所有权登记
资产确认/认领
业务部
已确认
登记资产
=
数据资产认领率

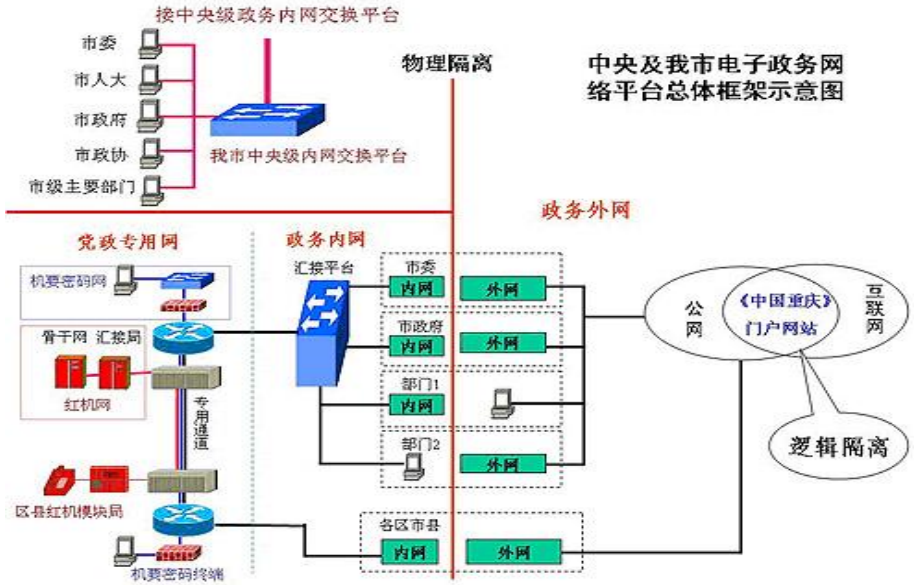
1.1 数据安全的概念

聚合数据

□ 数据传输安全

✓ 根据计算需求，在不同计算实体之间进行传输、分析和计算。确保数据传输中的安全性是一个挑战。

✓ 如此，即可能失去对数据的安全控制，数据安全边界并不存在。



政务内网、政务外网

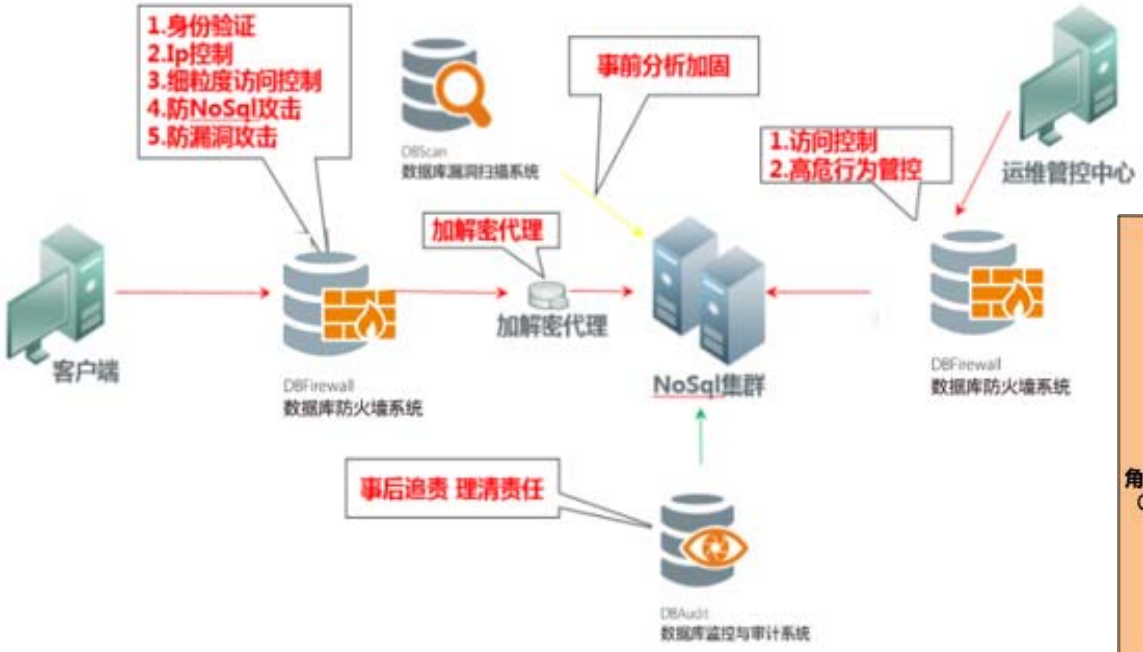
机密性和完整性

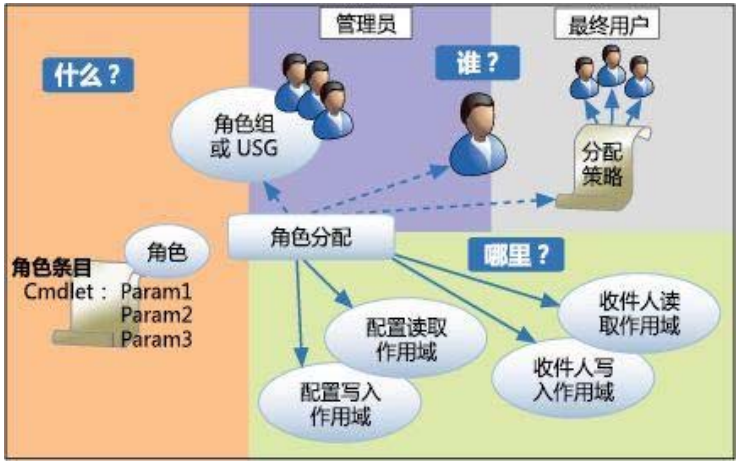
1.1 数据安全的概念

访问控制技术

□ 数据存储安全

目前，数据多是分布式存储在大数据平台中，采用云存储技术，以多副本、多节点、分布式的形式存储各类数据。也有数据集中存储。存在被非法入侵和数据被泄露的风险。





1.1 数据安全的概念

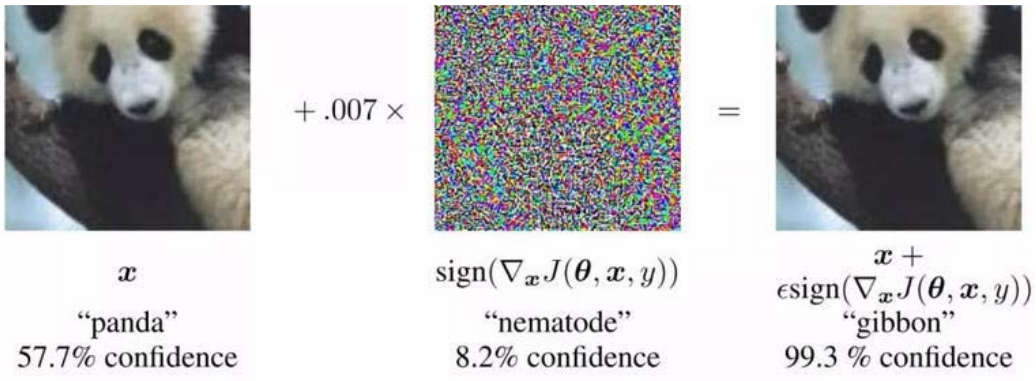
□ 数据处理安全

✓ 数据处理，即对数据进行操作、加工、分析等过程.

✓ 数据处理阶段是整个周期的核心阶段，数据处理安全与否直接关系到整体数据安全。

✓ 此阶段对数据接触最深入，所以安全风险也比较大。

✓ 包括数据脱敏、数据分析安全、数据正当使用、数据处理环境安全...



1.1 数据安全的概念

□ 数据交换（共享）安全

✓ 数据交换 **data interchange**

为满足不同平台或应用间数据资源的传送和处理需要，依据一定的原则，采取相应的技术，实现不同平台和应用间数据资源的流动过程。[GB/T 35274—2017，定义3.11]

✓ 数据共享 **data sharing**

让不同大数据用户能够访问大数据服务整合的各种数据资源，并通过大数据服务或数据交换技术对这些数据资源进行相关的计算、分析、可视化等处理。[GB/T 35274—2017，定义3.12]

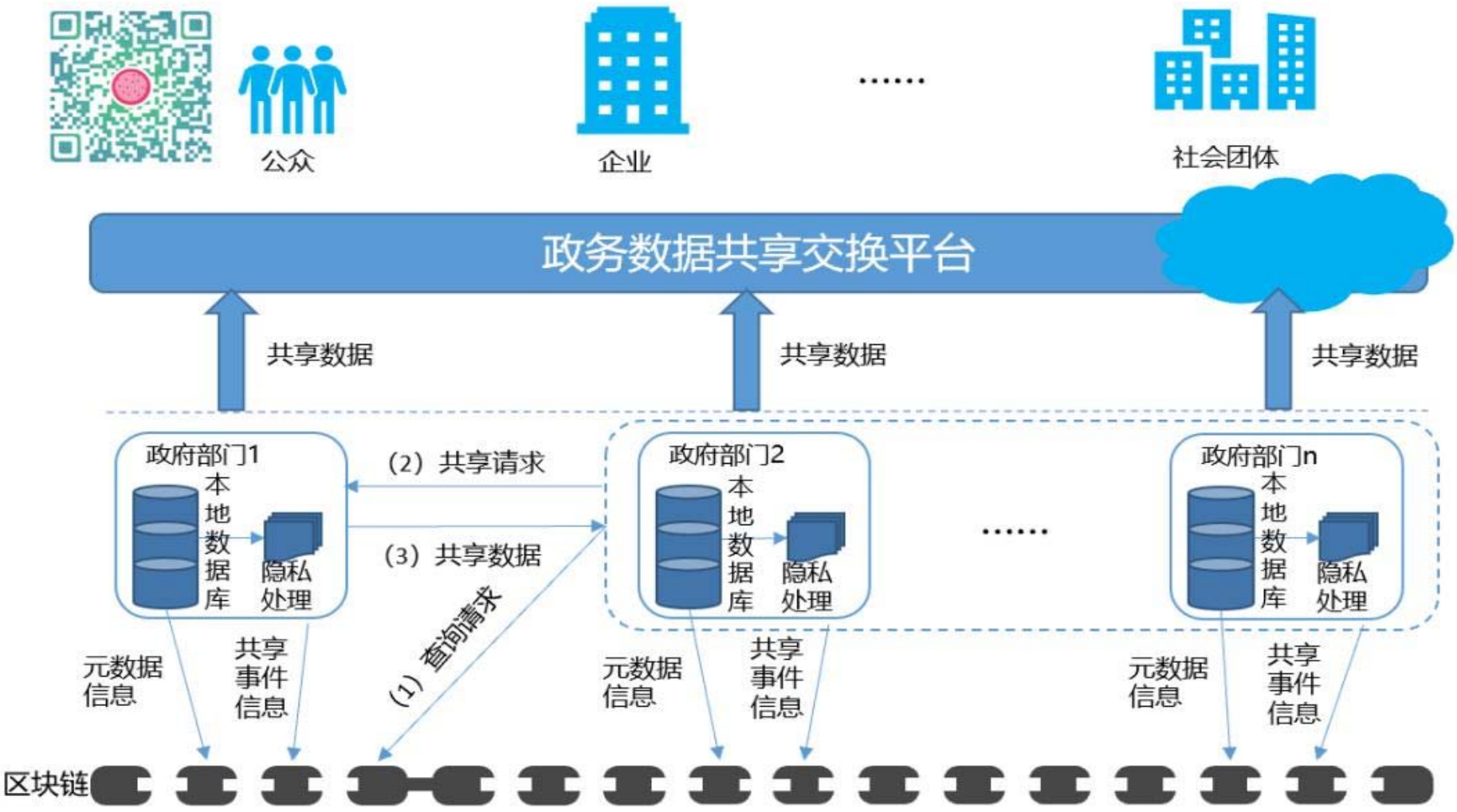
1.1 数据安全的概念

□ 数据交换（共享）安全

✓ 数据的互联、共享、整合是数据应用的需求。

✓ 由于数据资源跨个人、跨部门、跨管理域、跨省（甚至跨境）共享使用，可能导致**数据泄漏**。

机密性和完整性　　数据可信



数据安全

福州大学
FUZHOU UNIVERSITY

1.1 数据安全的概念

□ 数据交换（共享）安全

✓ 数据供应链 **data supply chain**

对大数据服务提供者提供的数据采集、数据预处理、数据聚合、数据交换、数据访问等相关数据活动进行计划、协调、操作、控制和优化所需的可用数据资源形成的链状结构。[GB/T 35274—2017，定义3.10]

注1：数据供应链目标是将大数据服务所需的各种数据和系统资产，通过计划、协调、操作、控制、优化等数据活动，确保大数据服务提供者能在正确的时间，按照正确的数据服务协议送给正确的大数据使用者。

1.1 数据安全的概念

◘ 数据销毁

当前，数据的安全销毁或安全删除是一个重要问题。

✓ 删除 **delete** [GB/T 35273—2017，定义3.9]

在实现日常业务功能所涉及的系统中去除个人信息的行为，使其保持不可被检索、访问的状态。

✓ 如果其存储在云端或云平台的数据删除不彻底；

✓ 如果废旧的手机存有敏感数据的碎片；

✓ 如果报废的硬盘曾经存储涉密数据；

✓ 如果交换、传输、暂存于他处的隐私数据并未销毁；

✓ … ***数据确实被删除?? 哪些工具？？原理？？***

1.2 数据安全的态势

□ 典型数据安全事件

Top 10. 未知（2.01亿条）

2020年1月，一个数据库，包含超过2亿条被暴露在公共互联网上的敏感个人记录，托管在Google Cloud服务器上，包含大量关于美国居民及其财产的**高敏感度**统计数据，例如姓名、地址、电子邮件地址、信用评级、收入、净值、房地产市场价值、投资偏好等。

Top 9. 微软（2.5亿条）

2020年1月，微软表示其用于存储客户支持分析结果的服务器发生数据泄露。共涉及**2.5亿条记录**（包括电子邮件地址、IP地址以及客户支持案例的详细描述），数据在未经密码保护情况下被意外公开。

https://github.com/woj-ciech/LeakLooker

1.2 数据安全的态势

□ 典型数据安全事件

Top 8. Wattpad **(2.68亿条)**

2020年6月，一个包含超过2.68亿条记录的数据库遭遇入侵，归属于加拿大写作博客平台Wattpad。恶意攻击者入侵了Wattpad的SQL数据库，其中包含用户账户凭证、电子邮件地址、IP地址以及其他敏感数据。

Top 7. Broadvoice (3.5亿条)

2020年10月，美国VoIP服务供应商Broadvoice由于配置错误，泄露超过3.5亿条客户记录，其中包括用户姓名、电话号码、通话记录以及留给医疗机构及金融服务公司的语音邮件，无需任何身份验证即可轻松访问。

1.2 数据安全的态势

□ 典型数据安全事件

Top 6. 雅诗兰黛（4.4亿条）

2020年1月，美国化妆品巨头**雅诗兰黛**一套**未受保护的数据库**将4.4亿条内部记录意外暴露在互联网上。信息包括电子邮件地址、内部文档、IP地址以及该公司内部教育平台的相关信息。

Top 5. 新浪微博（5.38亿条）

中国最大的社交媒体新浪微博于2020年3月遭遇**信息泄露，**超过**5.38亿**用户的个人信息被摆在暗网及其他在线网站上公开销售。黑客称，敏感数据（包括1.72亿用户的昵称、性别、居住地以及电话号码）是从官方平台抓取获得。

1.2 数据安全的态势

□ 典型数据安全事件

Top 4. Whisper (9亿条)

2020年3月，高人气秘密分享应用Whisper将9亿条用户记录暴露在网上。除了大量匿名文本以及与内容相关的元数据以外，泄露的信息还包括位置坐标及其他敏感数据。**所有信息都可在未经密码保护的公共数据库上直接查看。**

Top 3. Keepnet Labs (50亿条)

2020年3月，英国网络安全厂商Keepnet Labs收集历史违规数据并通知商业客户，**为了加快执行流程，该公司在迁移数据库时会临时将防火墙关闭10分钟。**导致黑客通过不受保护的端口，以无需密码的方式访问数据。由此泄露了收集到的50亿个电子邮件地址与密码。

1.2 数据安全的态势

□ 典型数据安全事件

Top 2. Advanced Info Service (83亿条)

2020年5月，泰国最大GSM手机运营商Advanced Info Service公司，被发现一套完全开放的ElasticSearch数据库，包含多达4 TB、总计83亿条DNS查询与Netflow数据等，可被用于映射用户行为。

Top 1. CAM4 (108.8亿条)

2020年3月，在视频直播网站CAM4上发现一台未受保护的ElasticSearch服务器，泄露7 TB数据，近110亿条记录，包括全名、电子邮件地址、聊天与电子邮件清单、密码哈希、IP地址以及付款记录等。

1.2 数据安全的态势

□ 典型数据安全事件

✓ 2018.08华住集团超5亿数据被拖库，疑似数据泄露，美爵、禧玥、漫心、诺富特、美居、CitiGo、桔子、全季、星程、宜必思、怡莱、海友等多家华住旗下酒店均在其列!

✓ 2020.08印度支付处理公司 Juspay 超过 1 亿用户的借记卡、信用卡信息在暗网上销售。Juspay 主要为亚马逊、Swiggy、MakeMyTrip 等公司处理支付业务。

✓ 美国网络安全厂商**FireEye (火眼)**于2020.12月透露，其内部网络被某个国家所突破，窃取了FireEye掌握的安全工具套件，“Red Team/红队工具”。 红队工具是一种网络武器库，能够复制全球最复杂的黑客攻击方法，这可能成为全球新一波攻击浪潮的起点。

1.2 数据安全的态势

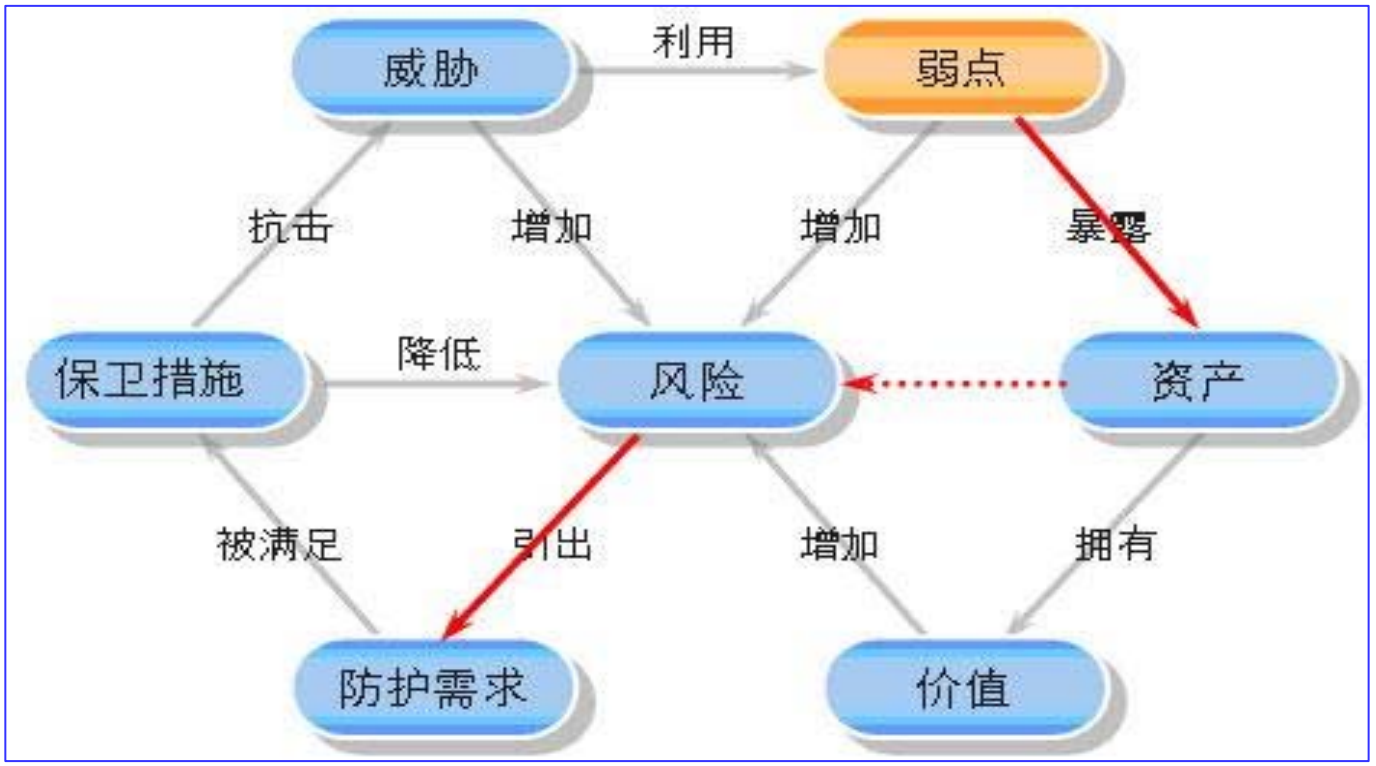
□ 典型数据安全事件

✓ Parler是Twitter的竞争产品，最近reddit社区传出，所有Parler用户数据（包括参加国会抗议示威活动人员）已经公开暴露，任何人都可查询。

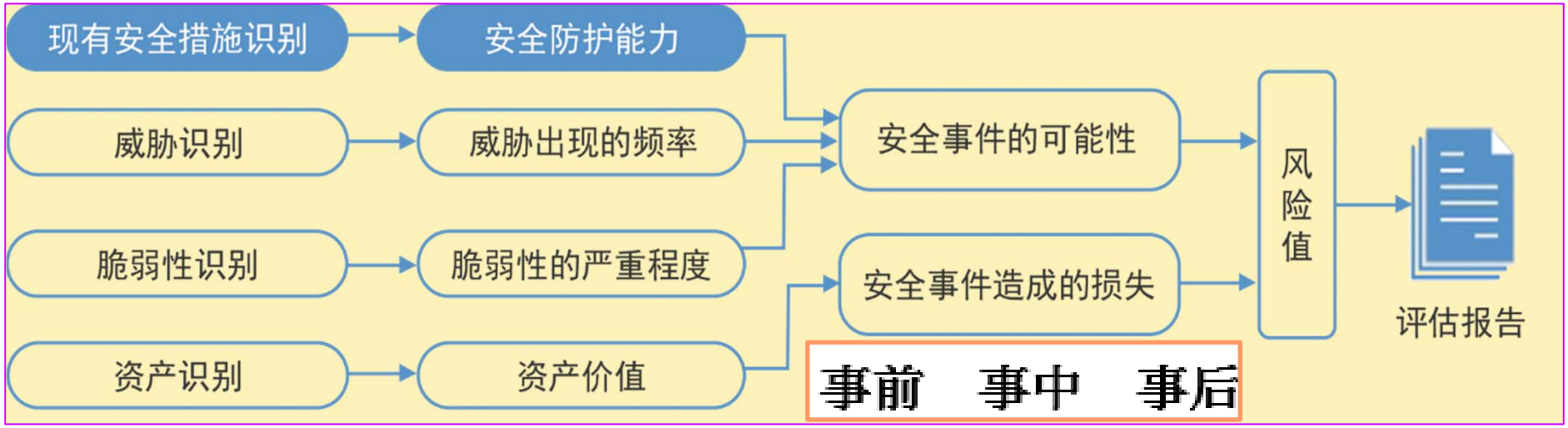
✓ 工信部通报63款侵害用户权益行为**APP**，关于侵害用户权益行为的APP通报（2020年第七批），此次检测发现，APP未经用户同意，私自收集设备MAC地址信息；将用户个人信息发送给第三方SDK的问题较多。

✓ 直播软件、优惠券发布、社交软件、新闻、小说、小游戏等.

1.3 脆弱性，威胁，风险


威胁
利用
弱点
抗击
增加
增加
暴露
保卫措施
降低
风险
资产
被满足
引出
增加
拥有
防护需求
价值


BUG


现有安全措施识别
安全防护能力
威胁识别
威胁出现的频率
脆弱性识别
脆弱性的严重程度
资产识别
资产价值
安全事件的可能性
安全事件造成的损失
风险值
评估报告
事前
事中
事后

1.3 脆弱性，威胁，风险

□ 术语：脆弱性　vulnerability

可能被一个或多个威胁利用的资产或控制的弱点。

[GB/T 29246—2017，定义2.89]

□ 数据的脆弱性

明文，无分级，无签名，无访问控制，无隐私处理，漏洞

…

据漏洞(CVE)数据统计，2020年报告了18,103个漏洞，大多数为高度严重级别，占比达57.1%。（属于系统&网络）

1.3 脆弱性，威胁，风险

□ 术语：威胁　threat

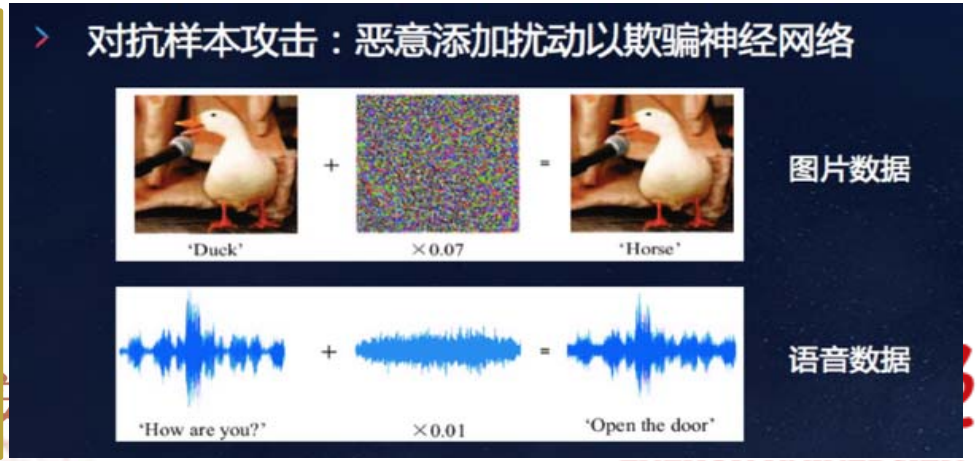
可能对系统或组织造成危害的不期望事件的潜在原由。

[GB/T 29246—2017，定义2.83]

□ 数据安全威胁：

误用，滥用@随意传播发布，窃取，篡改，拖库-刷(洗)库-撞库，DDoS攻击，暴力破解，伪造身份，伪造样本，社会工程，AI攻击、APT攻击…

AI攻击
@对抗样本攻击
@窃取数据AI黑盒交互
@窃取参数
@窃取超参数



1.3 脆弱性，威胁，风险

□ 授权人员非故意错误行为

由于用户不慎而造成意外删除或泄漏，如授权用户无意访问了敏感数据，还可能修改或删除了信息；用户对数据作了非授权备份。虽无恶意，但违反安全策略。

□ 推理攻击

推理攻击是指利用先进计算技术，从非敏感数据中推理出敏感数据，是一种十分微妙的安全威胁。

1.3 脆弱性，威胁，风险

- 社交工程的攻击

攻击者使用高级钓鱼技术，如钓鱼网站或邮件等，用户主动或无感地将机密数据提供给攻击者，造成数据泄露。

诸如“卖茶*”、“卖酒*”“爱心*”之类的骗局，我们要抱定一个决心：坚决不把钱和数据给对方！

1.3 脆弱性，威胁，风险

- **数据库管理配置的错误（安全管理能力弱**

配置的缺陷一旦被黑客利用，数据库很容易因外部攻击而造成数据泄露。 错误配置的数据库系统安全等级会明显降低，通常远低于数据库默认配置。

- **存在未打补丁的漏洞**

如果这些未打补丁的漏洞不属于公开漏洞，则入侵检测系统、防恶意软件等安全设施根本无法有效识别，一旦漏洞被利用，后果很严重。如：0day漏洞一旦被利用，就会数据库系统带来致命的威胁。

1.3 脆弱性，威胁，风险

□ 内部人员的攻击

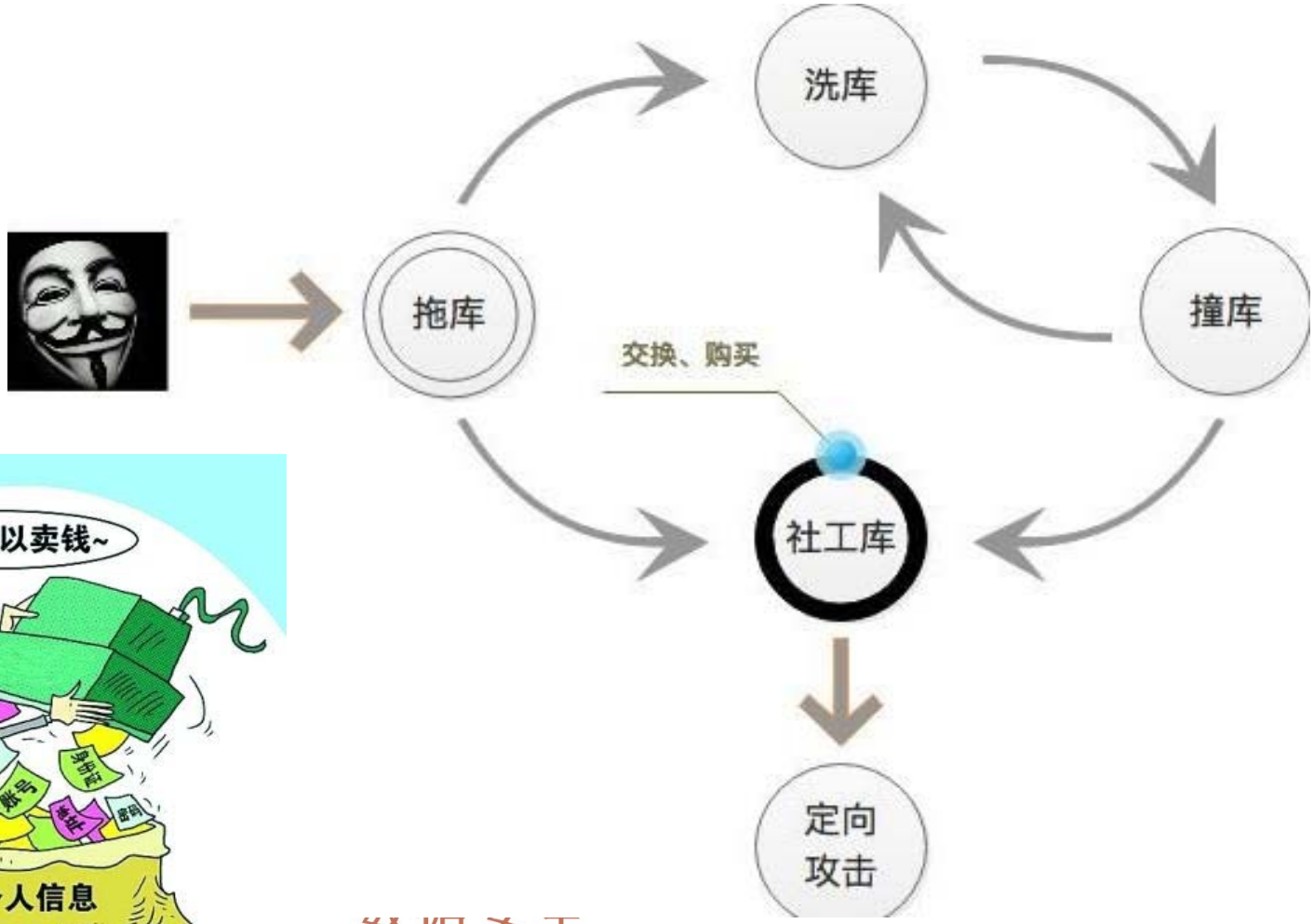
统计表明，约90%的数据库攻击源自企业内部。当受到利益驱使或存有报复倾向，在系统内即可攻击，绕过了外部防火墙系统，更加容易。

□ 高级持续性威胁（APT攻击）

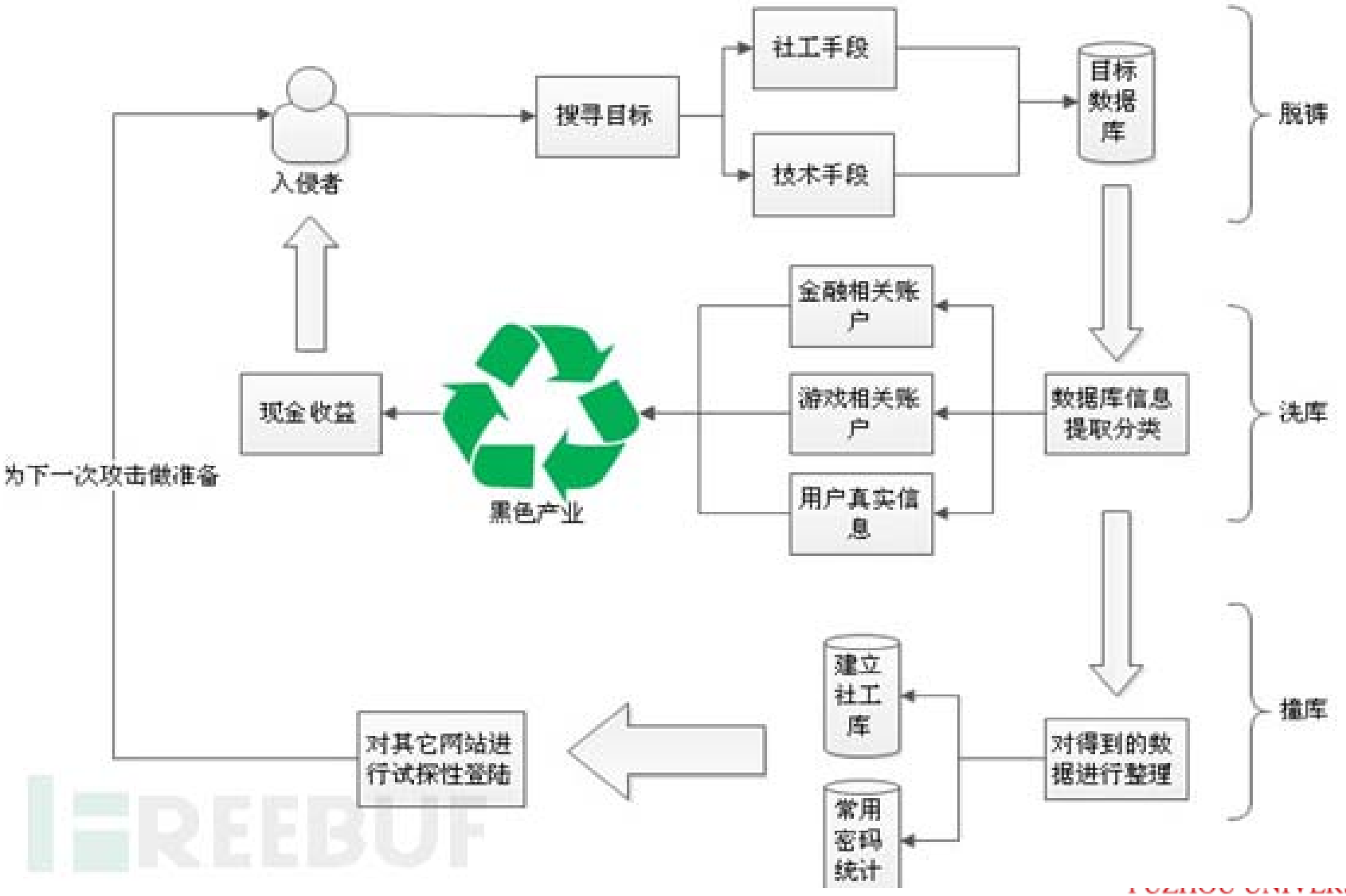
APT攻击通常是由专业公司，甚至政府机构发起的，具有极强隐蔽性，且目标性明确，多以重要的机密的数据为攻击对象。目前，这类攻击还没有明确的技术解决方案，给数据安全造成很大威胁。

1.3 脆弱性，威胁，风险

- 拖库-洗库-撞库攻击

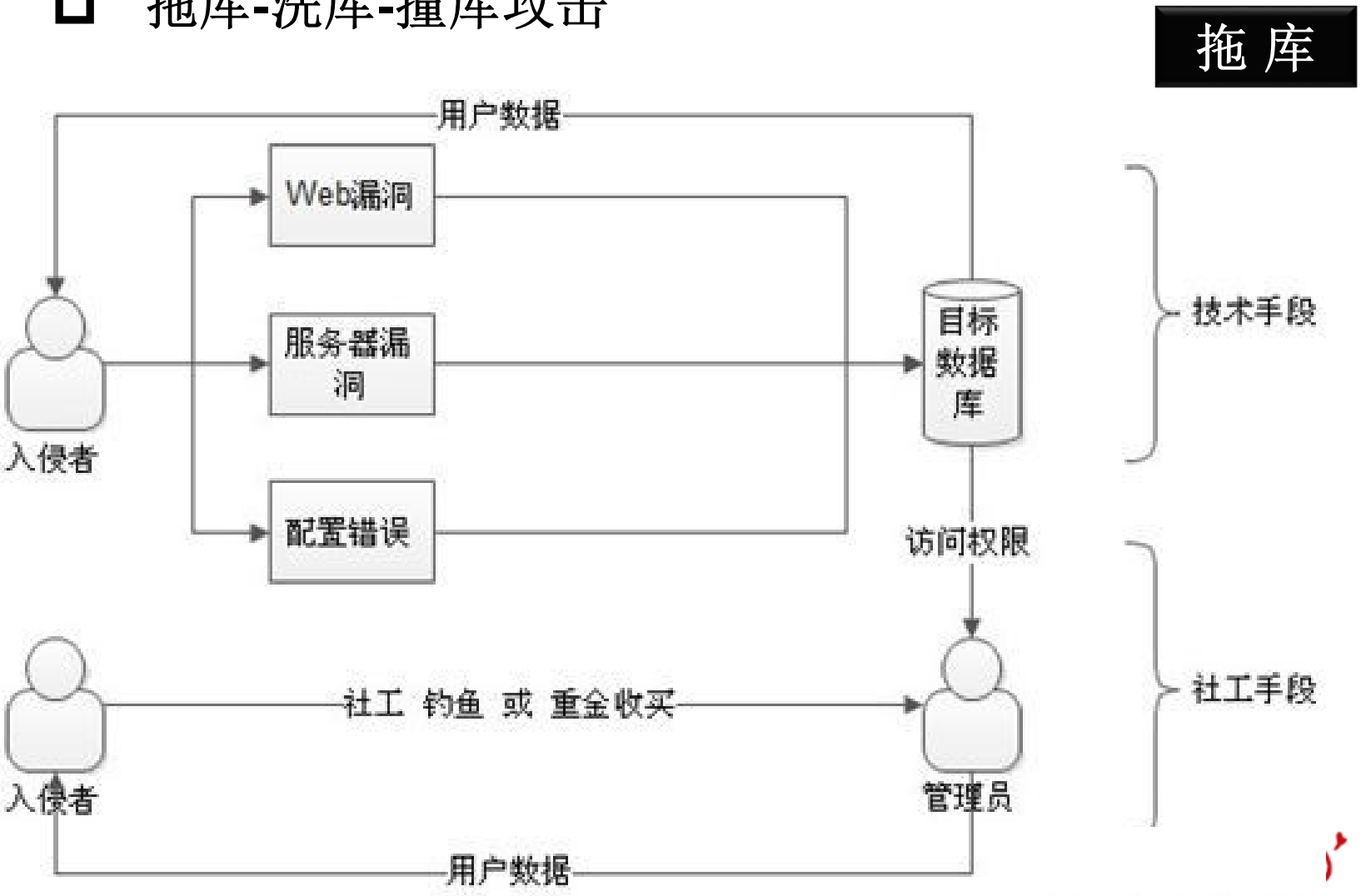


1.3 脆弱性，威胁，风险



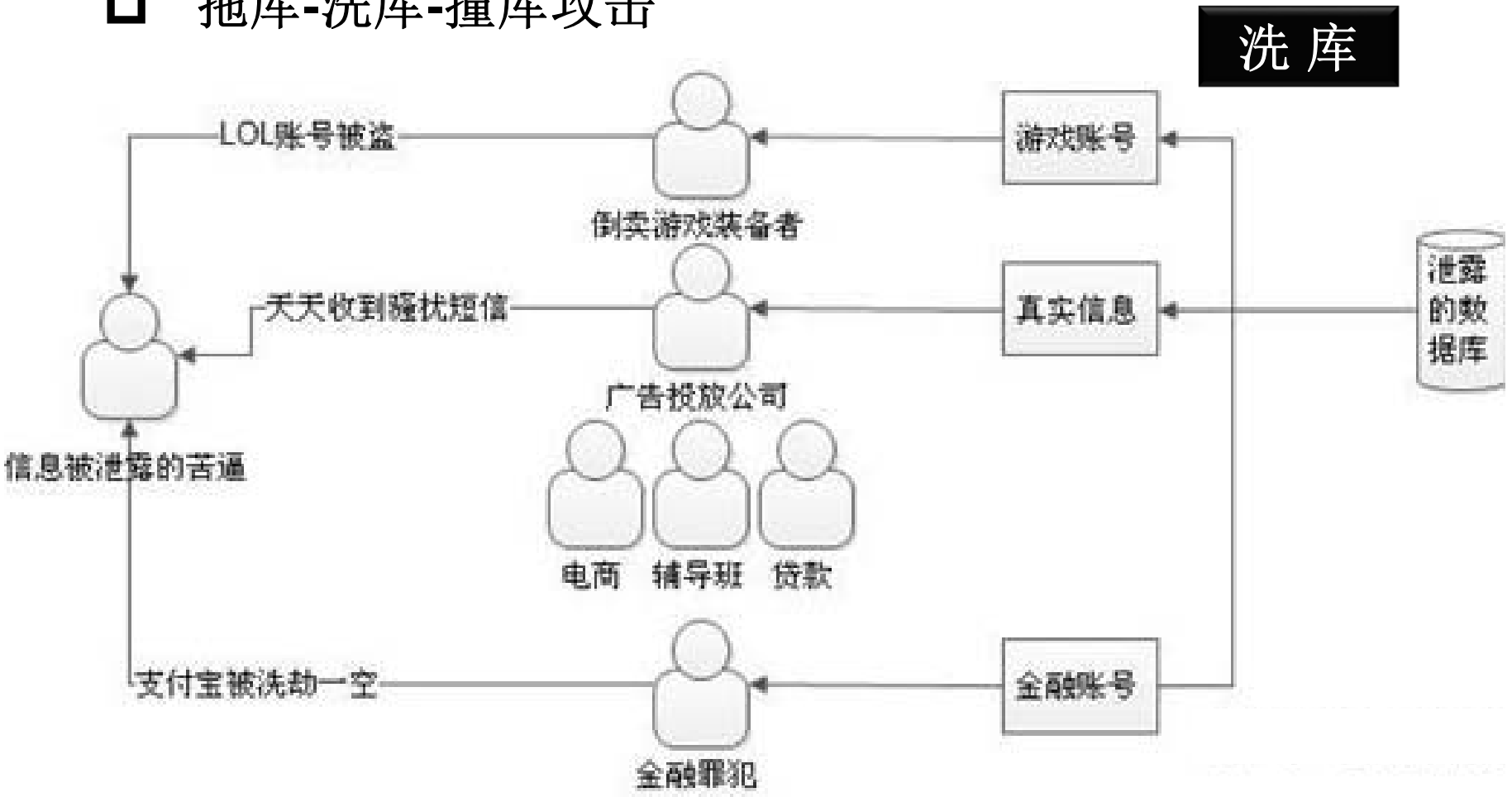
1.3 脆弱性，威胁，风险

- 拖库-洗库-撞库攻击



1.3 脆弱性，威胁，风险

- 拖库-洗库-撞库攻击



1.3 脆弱性，威胁，风险

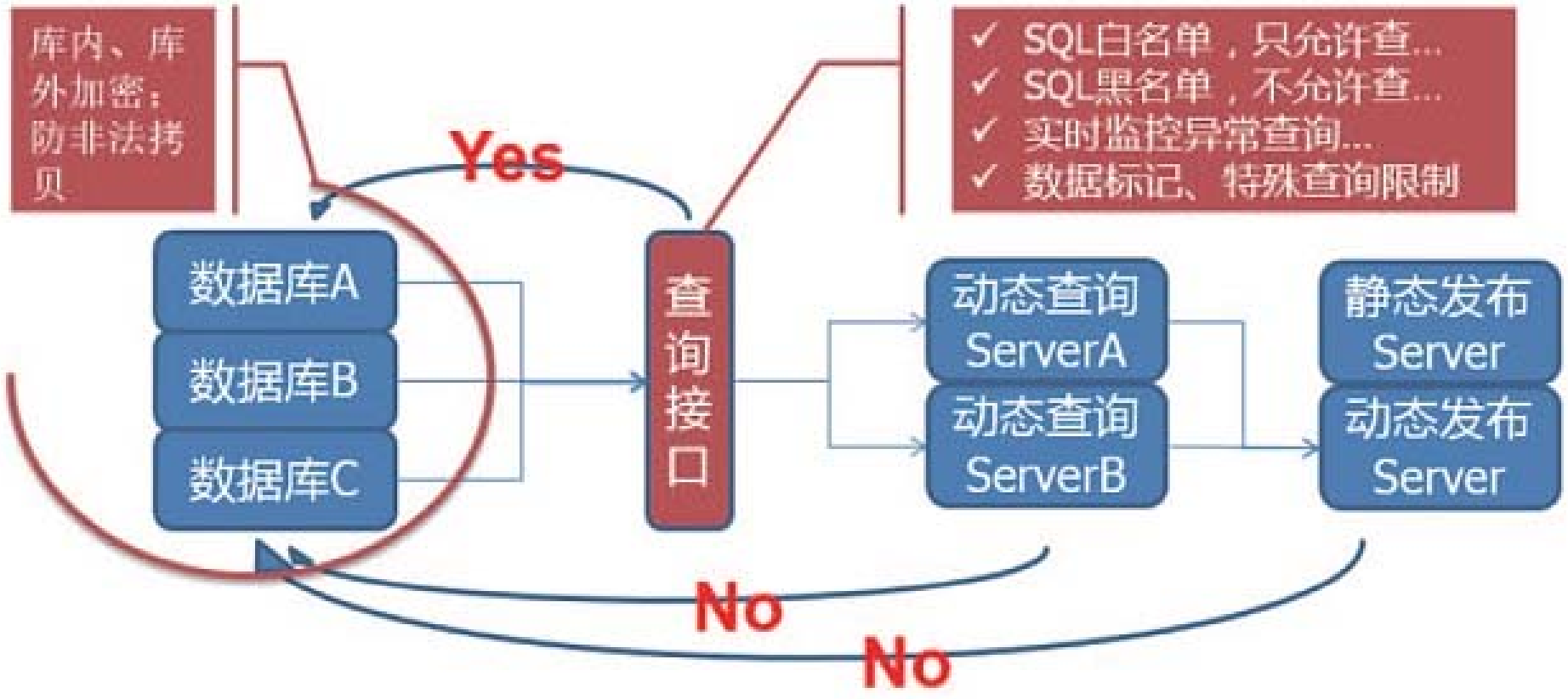
- 拖库-洗库-撞库攻击

撞 库

1.3 脆弱性，威胁，风险

- 拖库-洗库-撞库攻击

防范方法



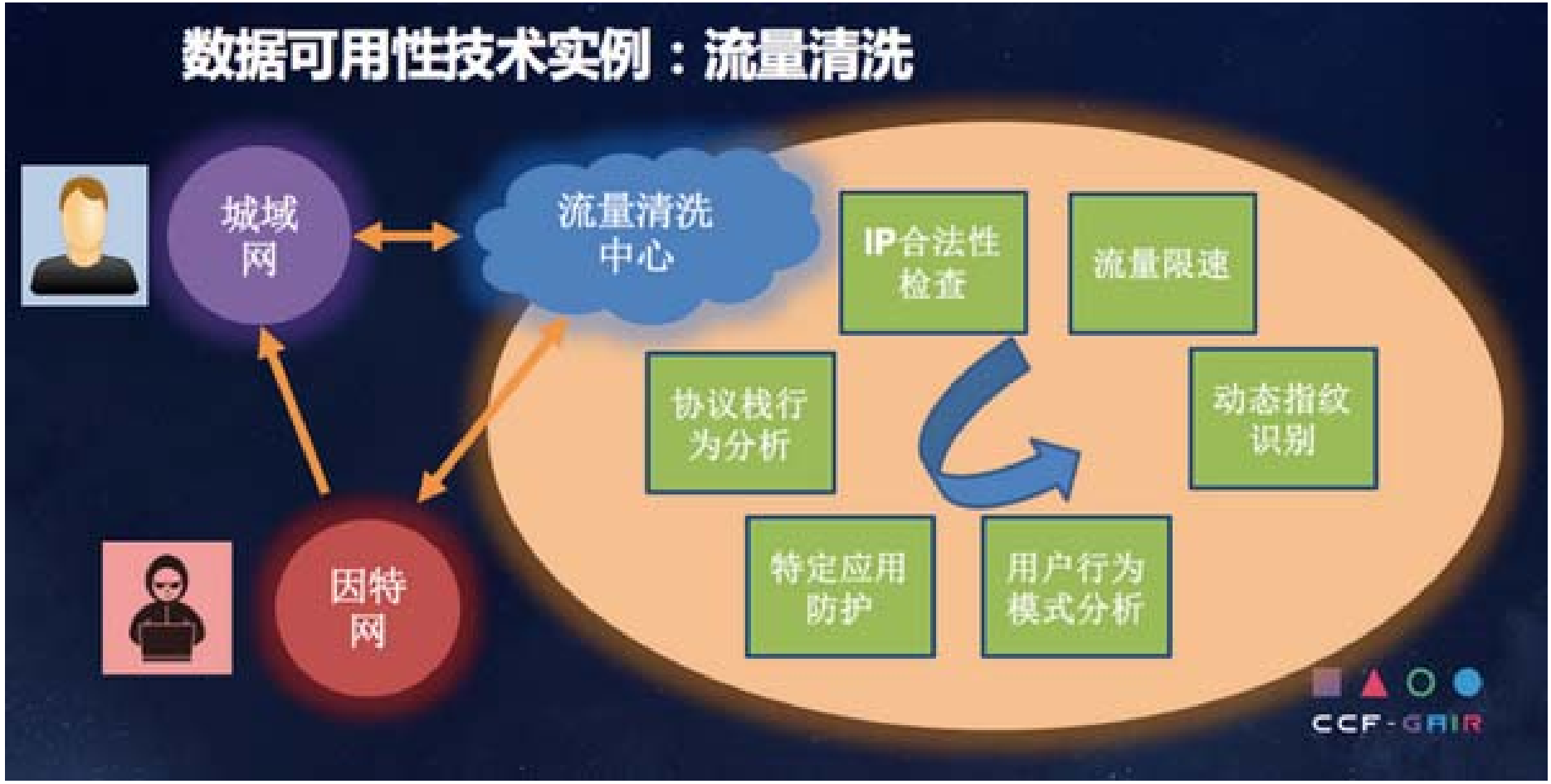
1.3 脆弱性，威胁，风险

- 数据传输、交换、流转、交易过程中，存在被人为窃取、恶意窃听、被篡改等安全威胁；
- 计算机、存储介质等硬件故障引起数据库内数据的丢失或破坏，例如，设备故障、磁盘损毁造成数据信息的破坏；介质安全
- 电源故障、自然灾害等其他因素也可能给数据安全带来风险；物理威胁

1.3 脆弱性，威胁，风险

- 数据可用性，面临的威胁主要是DDos攻击，防御技术手段主要是本地防护、云端防护和源端防护，多管齐下，把影响控制在最低范围内。
- DDoS防护技术是流量清洗，采用域名解析和其他方法，对网络流量进行分析，通过IP合法性检查、流量限速、动态指纹识别、特定应用防护等，去做DDos防护，提升数据可用性。

1.3 脆弱性，威胁，风险


数据可用性技术实例：流量清洗
城域网
流量清洗中心
IP合法性检查
流量限速
协议栈行为分析
动态指纹识别
特定应用防护
用户行为模式分析
因特网
CCF-GAIR

1.3 脆弱性，威胁，风险

□ 术语：风险　risk

对目标的不确定性影响。[GB/T 29246—2017，定义2.68]

注1: 影响是指与期望的偏离（正向的或反向的）。

注2: 不确定性是对事态及其结果或可能性的相关信息、理解或知识缺乏的状态（即使是部分的）。

注3: 风险常被表征为潜在的事态和后果，或它们的组合。

注4: 风险常被表示为事态的后果和其发生可能性的组合。

注5: 在信息安全管理体系的语境下，信息安全风险可被表示为对信息安全目标的不确定性影响。

注6: 信息安全风险与威胁利用信息资产或信息资产组的脆弱性对组织造成危害的潜力相关。

1.3 脆弱性，威胁，风险

□ 数据安全风险

信息泄露，篡改，不一致，不服务，损毁，数据非授权使用，个人隐私泄露…

□ 个人信息 personal information

以电子或其他方式记录的能够单独或与其他信息结合识别自然人身份的各种信息，包括与确定自然人相关的生物特征、位置、行为等信息，如姓名、出生日期、身份证号、个人账号信息、住址、电话号码、指纹、虹膜等。

1.3 脆弱性，威胁，风险

□ 个人敏感信息 personal sensitive information

指一旦泄露、披露或滥用可能危害人身和财产安全、损害个人名誉和身心健康、导致歧视性待遇等的个人信息。通常情况下，身份证号、银行卡号、健康记录、生物特征等属于个人敏感信息。

□ 匿名化 anonymization

对个人信息进行技术处理，使得个人信息主体无法被识别，且处理后的信息不能被复原。

数据脱敏–静态托名，动态脱敏

1.4 数据安全法律法规&标准

- 针对大数据安全合规方面的需求，大数据安全的标准化体系建设对规范和推动大数据产业的安全发展具有重要作用。当前，国内外标准化组织都认识到大数据的安全问题，并积极开展大数据安全的相关标准研究制定工作。

- 我国开展数据安全标准工作的组织主要是全国信息安全标准化技术委员会，在确保安全合规的情况下，如何规范指导各行业的大数据服务提供商，构建适合不同领域大数据的管理安全防护体系，以提供安全的大数据服务，仍然是急需解决的关键安全问题。

1.4 数据安全法律法规&标准

- 2016年11月我国颁布了《中华人民共和国网络安全法》

第二十一条 国家实行网络安全等级保护制度。网络运营者应当按照网络安全等级保护制度的要求，履行下列安全保护义务，保障网络免受干扰、破坏或者未经授权的访问，防止网络数据泄露或者被窃取、篡改。

第四十条 网络运营者应当对其收集的用户信息严格保密，并建立健全用户信息保护制度。

第四十一条 网络运营者收集、使用个人信息，应当遵循合法、正当、必要的原则，公开收集、使用规则，明示收集、使用信息的目的、方式和范围，并经被收集者同意。

1.4 数据安全法律法规&标准

- 2016年11月我国颁布了《中华人民共和国网络安全法》

第四十二条　网络运营者不得泄露、篡改、毁损其收集的个人信息；未经被收集者同意，不得向他人提供个人信息。但是，经过处理无法识别特定个人且不能复原的除外。

第四十四条　任何个人和组织不得窃取或者以其他非法方式获取个人信息，不得非法出售或者非法向他人提供个人信息。

1.4 数据安全法律法规&标准

- 2016年11月我国颁布了《**中华人民共和国网络安全法**》

第四十八条　任何个人和组织发送的电子信息、提供的应用软件，不得设置恶意程序，不得含有法律、行政法规禁止发布或者传输的信息。

第七十七条　存储、处理涉及国家秘密信息的网络的运行安全保护，除应当遵守本法外，还应当遵守保密法律、行政法规的规定。

1.4 数据安全法律法规&标准

- 2019年5月，国家互联网信息办公室发布《**数据安全管理办法**（征求意见稿）
- 2020年7月，《**中华人民共和国数据安全法**（草案）》在中国人大网公布
- 2020年9月中国银保监会印发《**监管数据安全管理办法 (试行)** 》

1.4 数据安全法律法规&标准

- 2020年9月，中国在“抓住数字机遇，共谋合作发展”国际研讨会上提出《**全球数据安全倡议**》.

引起国际社会广泛关注，尤其是某国。

- 2020年10月，全国大人常委会法工委发布了《**中华人民共和国个人信息保护法**（草案）》征求意见稿

- **数据安全法、个人信息保护法**为2021年全国人大审议重点.

1.4 数据安全法律法规&标准

截至2019年12月，已有四项数据安全国家标准正式发布：

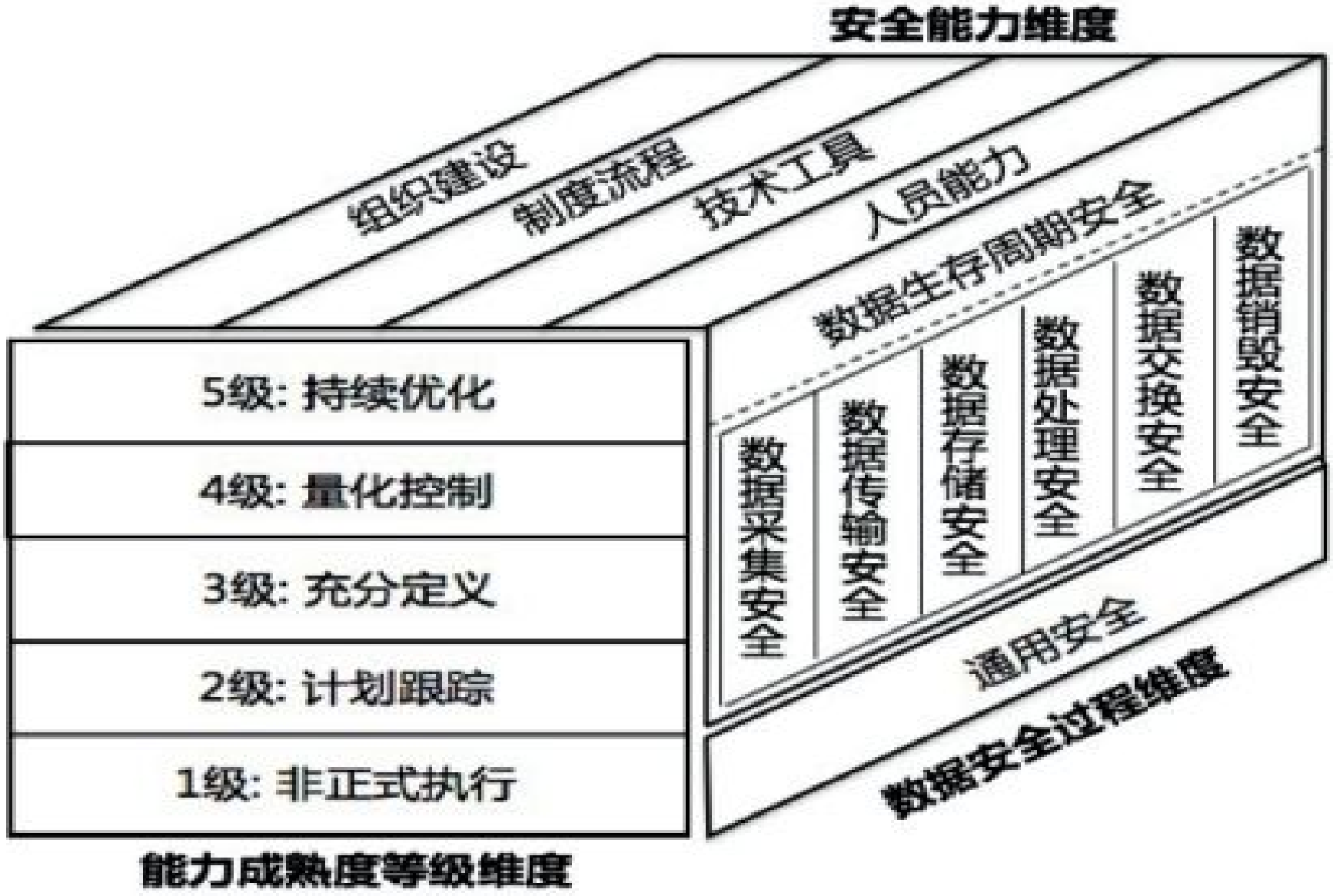
1.GB/T 35274-2017《信息安全技术 大数据服务安全能力要求》

2.GB/T 37932-2019《信息安全技术 数据交易服务安全要求》

3.GB/T 37973-2019 《信息安全技术 大数据安全管理指南》

4.GB/T 37988-2019 《信息安全技术 数据安全能力成熟度模型》

1.4 数据安全法律法规&标准



1.4 数据安全法律法规&标准

- 2012年，新加坡发布《个人数据保护法》
- 2018年，美国发布《数据保护法》
- 2018年，印度发布《个人数据保护法案》
- 2019年，美国制定《2019年国家安全和个人数据保护法》
- 2020年，英国发布《国家数据战略》

巴西《个人数据保护法》2015年
韩国《个人信息保护法》2011年
日本《个人信息保护法》 2015年

欧盟
《数据保护指令》1995
《通用数据保护条例》2015

澳大利亚 《隐私保护原则》
规定，数据主体获得明确告知后同意的，
才可以将个人数据传输至境外数据接收者。

附：2021年数据安全技术发展趋势

01 基于**零信任安全**的新一代远程办公体系建设成新热点.

02 **信息安全法律法规**密集出台推动数据安全解决方案变革.

03 基于**隐私计算**的大数据安全**共享技术**趋于落地.

04 基于大数据的**用户异常行为感知**技术(UEBA)趋于应用.

05 **智能语义分析**技术成为敏感数据**精准发现**的重要手段.

06 **数据流动**复杂化，驱动**数据安全风险评估**与预测技术发展.

07 **端点安全**防护由被动防护走向**主动防御**.

08 “互联网+”蓬勃发展，**敏感信息泄漏**防范日趋紧迫.

09 **数据安全销毁**作为数据生命周期的重要环节受到重视.

10 基于差分隐私的信息公开与情报分析**攻防对抗**技术成为新趋势.

谢谢大家，一起交流学习！

QQ：10068 0 2383