

可信数据库分析

肖珍 尹少宜 (mobile 组) 谢敏 (xml 组)

1. 引言

长期以来，邮件，财务记录，医疗图像，质量保证文档，订单记录一直是有巨大价值的资产，它们记录的事件是商业运作等关键性事务的决策基础。随着信息技术的发展，这些资料越来越多地以电子记录的方式被保存在数据库系统里，使得用户能快速的读写这些数据。但另一方面，为了谋取私利的攻击者也可能对这些数据进行修改而不留下痕迹，从而影响商业决策，损害国家和公众的利益。

1.1. 社会背景

2001 年 12 月，美国最大的能源公司-----安然公司，突然申请破产保护，此后，公司丑闻不断，规模也“屡创新高”，特别是 2002 年 6 月的世界通信会计丑闻事件，“彻底打击了（美国）投资者对（美国）资本市场的信心”（Congress report, 2002）。在安然事件中，公司 CEO，CFO 等通过发布虚假消息（虚假交易事件）和编造虚假的公司财务报告（虚报利润等）来误导公众，哄抬股价。另一方面，负责对安然公司财务进行审计的独立审计公司安达信的审计人员每年都会收到安然的巨额审计费，所以违背了审计人员的道德准则，对安然存在的问题视而不见，没有及时的向公众披露。在此案的审理过程中，美国检察官发现安然的公司的财务报告等运营数据完全是不可靠的，不真实的，从而导致了最后公司破产、股价暴跌、雇员失业，股民财产损失巨大，而公司的 CEO，CFO 等却通过抛售获得了巨大的个人收益。

在这一案件中，正是由于电子数据的管理不善，所以带来了国家和公众的巨大损失。随后，针对美国国内大规模上市公司的财务丑闻和审计标准屡屡拉响警报。在这些丑闻中，许多公司主管声称他们不应当对虚假财务报表负责，甚或他们根本不知道这些是虚假报表。为了转变这一信用危机局面，挽回公众对政府的信心，投资者对资本市场的信心。美国国会和政府加速通过了一系列法律法规，来规定企业的电子数据如何保留和储存的问题，从而准确地从上市公司的电子文档记录中掌握有关上市公司内部运营的信息。

其中包括：约束证券经纪商的美国证券交易委员会规范 SEC（Securities Exchange Commission），全美证券交易商协会行为规定（NASD 3110），约束医疗保健业的美国健康保险便利和责任法案（Health Insurance Portability and Accounting Act，简称 HIPAA），规定生命科学的联邦条例 21CFR 第 11 部分，规定美国国防部电子记录管理应用（RMA）软件的设计标准的 DOD 5015.2-STD 标准，约束上市公司财务管理等行为的萨班斯-奥克斯莱法案（Sarbanes-Oxley）法案，金融服务业的数据安全性规范 Gramm-Leach-Bliley 法案

(Gramm-Leach-Bliley Act) 等众多法案。这些法案对电子记录在完整性、保密性、可存取性、可靠性等各方面都有明确规定。

1.2. 法律背景

1.2.1. 企业财务：萨班斯法案

在安然事件等一系列华尔街财务丑闻中，许多公司主管声称，他们不应当对虚假财务报表负责，也根本不知道这些报表是虚假的。美国政府由此制定了影响极为深远的萨班斯法案，该法案的另一个名称是“公众公司会计改革与投资者保护法”。法案的第一句话就是“遵守证券法律以提高公司披露的准确性和可靠性，从而保护投资者及其他目的”。该法案管制对象仅限于在资本市场运作资金超过 7500 万，并且每季度必须向证券交易委员会提交报表的股份公司的美国上市公司和美国企业的海外分支机构及子公司。该法案规定了强化信息披露、监管责任、内部控制和外部审计等制度，要求公开上市公司需定时地公布准确详细的财务报告，强制设置审计委员会，并规定该委员会由独立董事组成，该独立董事没有担任公司管理层级职务，也不从公司领取薪金。其主要目的是确保审计人员的独立性，授予其审核公司财务记录的权限与自主性，并监督经理人的工作绩效。该法案对各公司信息保存有如下规定 (<http://www.kahnconsultinginc.com>):

- 公司的首席执行官和首席财务官必须亲自证明，他们提交的财务报表是真实的，否则将承担刑事责任。
- 对信息保护的能力。要求公司运用“细致入微”的存取控制和保护手段，防止非授权或因疏忽而更改、毁坏或破坏业务记录和财务信息。
- 对信息准确跟踪的能力。要求公司能够提供审计人员与保存关键记录和信息的系统之间所有交互行动的“审计轨迹”。信息和记录以及文档管理软件、硬件和安全存储环境形成了关键的“内部控制”，它确保各公司其财务和业务信息是准确和可靠的。
- 对信息长期保存的能力。要求公司确保用于保留要求记录的存档和存储系统和介质将支持长期可靠存取。对某些特定信息要求保存长达七年的时间。
- 从该材料提交的财政年度末开始计算，若不可提供 5 年内工作材料供审计或审查，将处以最高 5 年的监禁，并可被课以罚款，或单独课以罚款。
- 为了阻碍美国联邦调查，故意更改、销毁、损坏、隐匿、包庇、伪造任何记录、文件或者有形物体，导致记录的完整性受损或者在任何记录、文件或者有形物体中制造虚假条目导致记录的可靠性受损的人都将承担刑事责任，使其作为官方处理证据的价值受影响，可以被判以最高 20 年监禁，并可被课以数额不等罚款，或单独课以数额不等的罚款。

1.2.2. 证券交易：SEC 17a-4

如今的证券业的信息交流基本上是通过电子方式，包括电子邮件、即时信息传输以及各种电子表格（票据、文件、批准书等）。美国证券交易委员会要求券商将所有客户通信的电子资料和其他经纪记录都存储在不可删除、不可改写的介质上。除此之外，美国证券交易委员会还要求企业对各种频繁和广泛的信息需求做出迅速反应。美国证券交易委员会 SEC

(Securities & Exchange Commission) 的第 17a-4 条法案 (简称: SEC 17a-4) 对交易记录的保存有如下规定:

(<http://www.law.uc.edu/CCL/34ActRIs/rule17a-4.html>)

- 数据必须存储在“不可改写、不可删除”的介质上,也就是说只能以“不可覆盖,不可擦除”的方式保存记录;
- 自动验证存储介质记录过程的质量和准确性;
- 将原存储介质及其副本单元(如果合适)以及此电子介质上存储信息的保存期的日期和时间序列化;
- 有能力响应交易委员会或自律机构的要求随时将电子存储介质上保存的索引和记录方便地下载到任何可接受的介质上。

对电子记录所做出的以不可改写、不可擦除的格式保存的要求是要确保信息的完整性。验证信息的质量和准确性实际上也是对信息完整性的要求。在 SEC 17a-4 所规定的“能够及时下载保存在电子存储介质上的索引和记录”,是对电子记录的可存取性的具体规定。当然,保密性对所有经纪行的运营来说也是一个重要考虑事项。

1.2.3. 医疗健康: HIPAA 法案

1996 年,美国国会通过了健康保险便携性和责任法案(HIPAA),其中有关个人的医疗健康信息隐私权的条款于 2003 年 4 月 14 日生效。该法案打破了传统的由医疗提供者拥有患者个人医疗记录的观念,转化为以客户为主,由个人自己来决定自己的医疗信息将如何被使用的观念。政府的立场是,通过在线访问患者资料的方式可以大大提高医疗服务的质量,但同时也应当保护患者的隐私,防止对秘密数据的不正当使用。尽管该法案本身并没有对数据的存储方式做出要求,但要求医疗机构和其他相关机构(医院、保险公司和卫生维护组织)使用安全的系统和介质来对所有患者的记录进行电子化管理,保存 HIPAA 安全标准要求的相关文档,并接受对这些资料和相关过程的定期复查。

该法案对多种医疗健康产业都具有规范作用,包括交易规则、医疗服务机构的识别、从业人员的识别、医疗信息安全、医疗隐私、健康计划识别、第一伤病报告、病人识别等。该法案的主要目标如下:

- 保证劳动者在转换工作时,其健康保险可以随之转移;
- 保护病人的病例记录等个人隐私;
- 促进国家在医疗健康信息安全方面电子传输的统一标准。

HIPAA 条例定义:

- 受保护的医疗信息(Protected Health Information, 简称 PHI)包含以任何形式或者媒体传播的所有的医疗信息,不管是口头的还是有记录的。
- PHI 主要是由以下对象所创建或者接收到的:医院、健康计划部门、保健服务商、相关票据交换所、医疗信息系统提供商、医科大学、甚至只有一个内科医生的办公室,雇主,保险公司,学校等。
- PHI 是和以下信息相关的:某个个人过去、当前或者未来的身体或者心理健康状况;向患者提供的医疗服务;过去、当前或者未来对医疗服务的支付费用
- PHI 能够直接或者间接用于确认患者个人身份。

HIPAA 条例有以下规定:

- 对任何形式的 PHI 的存储、维护和传输都必须遵循 HIPAA 的安全条例规定,并且大部分组织必须在两年内达到要求。

- 对于违反 HIPAA 安全条例的行为，可以处以最高为 25 万美元的罚款和最长为 10 年的监禁。
- 保密性：对数据访问的保护和监控，保护数据免受非法访问，如病人的病例属于个人隐私，应予以保密
- 一致性：保护数据免受非法修改和删除。
- 可用性：系统和数据处于可访问和运行阶段的时间长度。

1.2.4. 生命科学和制药行业：FDA 法案

美国食品与药物管理局 FDA (Food and Drug Administration) 的 21 CFR (Code of Federal Regulations) Part 11 (Electronic Records and Signatures) 法案，意在简化药品从开发到投入市场的过程，并使之更有效。该条例的目标是，编制并管理有关药品开发、药品测试和批量制造的信息流，从而加快这一过程，并使之更加安全。无论是确保药品在获得审批之前经过彻底测试，还是保证药品经过充分的实验和调查，严格的记录都是必不可少的。由于大多数制药公司掌握的药品临床测试数据都与患者有关，它们必须同时按照 HIPAA 法案的要求，保证数据的机密。具体的对电子记录和签名有如下规定：<http://www.21cfrpart11.com>

- 确保电子记录的真实性、完整性。
- 验证系统以确保具有准确性、可靠性，一致的、所希望的性能，和识别无效或被篡改报告的能力。
- 保护记录以使它们在整个保留期内都准确而且可以随时检索。
- 将系统访问权限制到有权访问的人。
- 使用安全、由计算机生成而且加盖时间戳的审核跟踪，“其保留期至少应与主电子记录保留期一样长”
- 对开放系统的控制。“文档加密等附加措施”

通常,制药厂在将一种新药品推向市场的过程中,将会产生大量的文字资料。FDA 发布这一管理法规,提出确认系统以确保精确性,及保护记录以支持记录的“精确性和及时检索”标准。其基本要求就是提供系统验证和保留管理能力,以确保数据的完整性。

从以上这些美国法案对于电子记录存储的规定举例,我们可以看到,为了达到这些要求,需要有一个可信的数据库来存储和管理电子记录,使得里面存储的电子记录是可信的。那么可信的数据库包括哪些方面的技术,有什么要求呢?下面将详细的分析。

2. 可信的数据存储

2.1. 理论和案例分析

记录管理实际上就是记下那些对于企业和组织的业务非常重要的事情作为历史备案,便于审计及今后的所有调查、取证和分析等工作。可信数据库存储的电子记录需要满足以下要求,提供以下特性:

- 信息的完整性: 所有相关记录都保存。
- 信息的准确性: 数据记录描述准确, 精确。
- 信息的可靠性: 数据是真实可靠的, 没有虚假数据。
- 信息的安全性: 相应权限控制; 数据不能被非法使用;

- 信息的可获取性：能够在适当的时间以适当的格式访问任意的数据。
- 信息的不可更改性：历史数据记录不能被更改（被覆盖，被擦除）。
- 信息的时效性：所有信息都是在各项相关政策和规定所要求的时限上满足以上要求的，数据有自己的生命期。
- 日志：能够安全地记录所有数据操作：创建、改动和删除等日志信息。

以公司的交易记录为例：

- 完整性要求对某一项交易相关的数据都要有记录，以免出现数据的不一致性，比如收支不平衡等；
- 准确性要求相关数据一定要与事实相符合，交易的价格数量等不能有超过容忍的偏差，利润必须与事实吻合，不得虚报以哄抬股价，也不得漏报以偷税漏税；
- 由于数据反映了已经发生的事件的情况，是一种“证据”，所以必须是可靠的，不能有虚假的交易记录并不存在的交易事件；
- 安全性要求逻辑上由相应的访问权限控制等，比如只有交易员有权限提交和修改交易记录，物理上要求对存储介质有备份措施等；公司必须明确应当如何移动和存储数据，授权的人员应当如何以及在何时访问并修改数据，以及是否应当在一定的时期之后销毁数据。公司策略还必须确保未授权的人员无法对数据进行不正当的访问、更改或者删除数据。
- 可获取性要求任何对数据的合理的访问都应该能够快速响应，以满足业务的效率要求和相关部门的审计要求。比如审计师能够在可以容忍的响应时间内查询到需要审计的交易记录；对电子信件、数据、文件的察看有响应时间上的要求，是要避免被审计机构如公司利用时间拖延而伪造相关记录，时间拖延越长，造假的可疑性就越高。
- 由于电子记录非常容易修改，删除并且不留下痕迹，所以必须要求历史记录以不可覆盖，不可擦除的方式存储，对记录的修改只能以添加新记录的方式进行，以完全的反映该记录从创建开始的所有变化情况，只有当该记录过期之后，才可以删除或者备份。
- 时效性要求文档的保存都有规定的时间，而不同产业的电子数据也各有不同年限要求。比如萨班斯法案，要求会计等相关资料要留存 4 年以上；比如医疗记录必须保存 21 年，HIPAA 要求保存 30 年；甚至美国证券交易委员会的第 17a-4 条法案（SEC 17a-4）更要求资料要保存到该业者结束营运为止。

在美国推出的众多关于电子记录管理的法案中，上述要求或多或少的贯穿于全部管理法规要求，是电子记录管理的趋势所在。这些法案的颁布也促使公司采取相应的技术措施来保证“法规遵从性”，以便在突如其来的配合调查中能够从容因应，否则将会受到法律的处罚。

2002 年 11 月左右，美国政府对华尔街（Wall Street）5 家知名的金融公司予以罚款，包括高盛证券（GS）、摩根史坦利（MWD）、花旗集团（C）的投资银行部门：U.S. Bancorp.（USB）所属的 U.S. Bancorp Piper Jaffray、以及德意志银行（DB）的证券部门，总计罚款达 830 万美元，受罚原因就是：未依联邦主管部门的要求将电子信件进行留存。这就是违反了可信数据库的电子记录存储的完整性。美国对证券业者业务相关的电子文件保存规范要求 2 年内的数据要能立即被察看，而即便 2 年过后也依然要再保存 1 年，但这 1 年的保存就不再硬性规定立即调阅，可以用其它方式储存。诸如此类的要求，多是为了日后配合法令调查，包括金融业者的客户可能涉及洗钱，或金融机构本身可能违约交易等，届时持续留存的电子数据、文件就成为记录的证据。

为了提供可信的电子数据，可信数据库采用的存储介质要求考虑安全性、数据完整性、总拥有成本、性能、访问能力以及搜索功能。一个设计良好的系统也将具有灾难恢复功能，

除非发生重大的事故，否则记录将不能被更改，不会发生任何安全问题，重要数据也不会丢失。

为了满足上述的要求和特性，同时也要减轻企业的负担。可信数据库将采取什么样的技术呢？下面将分析一些已有的技术。

2.2. 技术实现

可信的数据存储要求：数据从创建到使用过程中一直保持可信。对于存储在一个数据库系统内的数据，如果是可信的，那么给定任何一组特定的输入，其输出一定是期望的结果。这要求数据库系统在任何时刻都具有正确的计算和推理能力，以及可靠的存储能力（存储的任何数据都可以被检索到，并且，检索到的任何数据都是被合法存储的）。也就是说，系统本身不存在任何错误的和欺诈的行为，同时也能够阻止任何恶意的入侵。（这些“错误的”、“欺诈的”、“恶意的”等的定义是针对具体的应用而言的）

为了实现这一要求，我们规定对于电子数据的存储满足以下要求：

- 为所有已经发生的事件创建正确的数据记录
- 没有虚假数据记录并不发生的事件
- 已存储的记录是规定的保留期内不可修改的（只能添加新的记录，不能对历史记录进行修改和覆盖），过期之后可以删除
- 有获取控制管理
- 在数据创建过程中，周期性的审计创建操作

为了满足上述要求，提供可信的数据，当前最普遍使用的存储技术是 **WORM** (write once read many)。如图 1 所示：它构建在普通的光盘、磁带、磁盘存储介质上，通过硬件设备的控制实现了数据只能一次写入，只能添加新的数据，不能修改历史数据的语义限制，从而为电子数据提供了最安全的保证。它提供了类似文件系统的接口，并被扩展以支持对文件和块的添加操作。我们做了一些合理的假定 1)任何用户不能从物理上来破坏 **WORM** 存储设备 2) 记录的创建是正确的，**WORM** 存储设备的运转也是正常的 3) 查询是正确的 4) 任何用户不能干涉查询的过程，例如修改查询接口等。

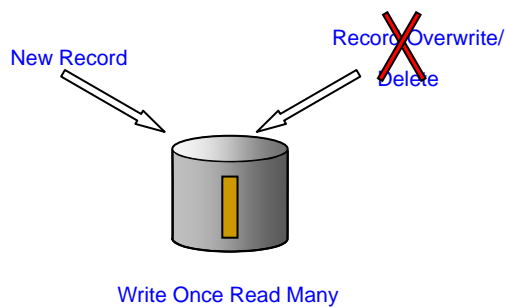
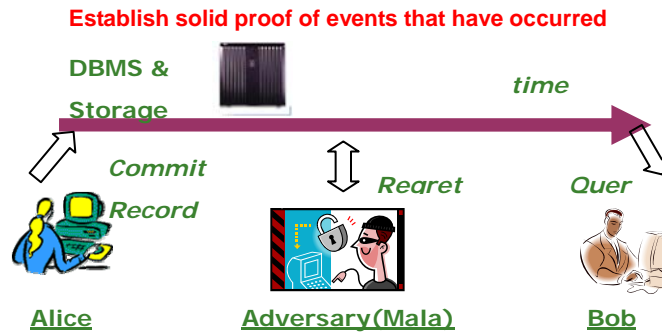


图 1: WORM 存储技术

在数据被创建到被使用的过程中，将面临怎样的威胁，我们将采用什么样的技术来保护数据不被攻击者破坏呢？我们将用一个示例来说明：

如图 1 所示：Alice 是一个合法的用户，她创建了一个记录 R，并提交到基于 **WORM** 存储的数据库里，在未来的某个时间，将有审计用户 Bob 提交查询并得到 R 作为结果。在 R 被创建到被查询的期间，有一个恶意的用户（Adversary）Mala，她不想让 Bob 得到 R 这个记录，所以会采取一些恶意的措施来掩盖 R 的存在。



Bob should get back Alice's data

图 2：威胁模型 1-----攻击者具有普通权限

如果 Mala 是一个普通的用户，那么她不能获得修改记录的权利，普通的获取控制机制和 WORM 存储特性就可以保证数据的安全。

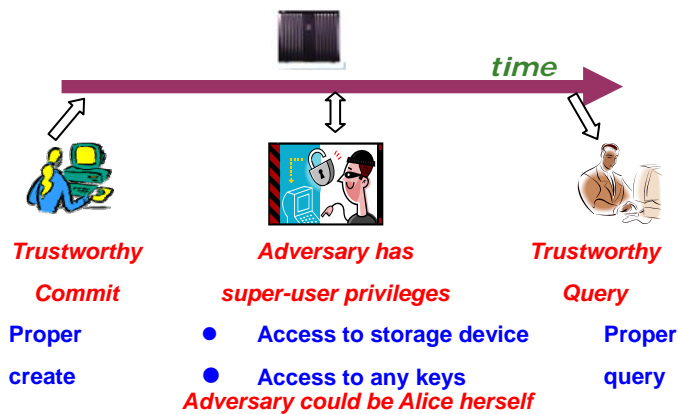


图 3：威胁模型 2-----攻击者具有超级权限

如果 Mala 是一个高级用户，情况就有所不同了，这在实际生活中也是常见的。公司的高层（CEO，CFO）或者高级技术人员等等往往具有超级用户的权限，能够进行所有普通合法用户的操作，例如往数据库里面写任何数据，读任何数据。如图 3 所示：在这样的情况下，普通数据库系统的获取控制机制就不起作用了，因为超级用户的权限是最大的。那么，简单的 WORM 技术还能保证数据的可信吗？不能，因为随着数据量的不断增大，数据库的查询主要是通过索引来执行的，所以虽然记录的提交和记录的查询都是可信的，但具有高级权限的用户（比如公司高层管理人员，技术专家）可以通过恶意篡改索引来从逻辑上隐藏某些数据，从而导致查询结果失真。如图 4 所示：

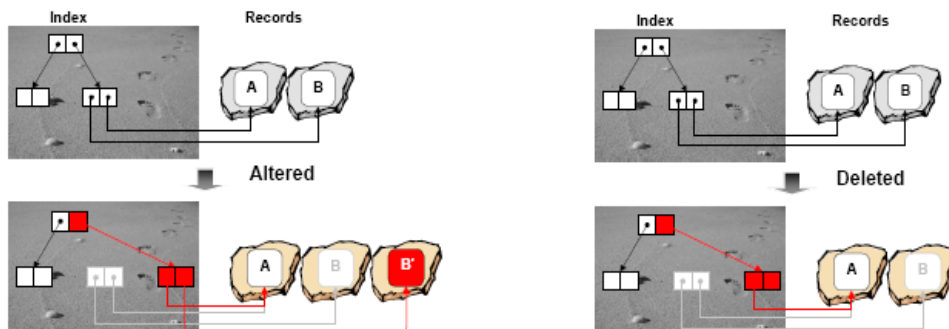


图 4：允许逻辑上修改数据的不可信索引

在沙地上记录的是不可信索引，表示索引很容易被修改，在石头上记录的是真实的记录，表示不能被修改。虽然记录 A, B 都存储在 WORM 存储设备里，但我们在查询的时候是通过索引来访问真正的数据。该索引是不可信的，它允许逻辑上的修改。因此通过修改索引指针，可以轻易的将 B 替换为 B' 或者将 B 隐藏，这样就达到了从逻辑上来隐藏记录 B 的作用。在这种情况下，我们面临的威胁模型变为如图 5 所示：

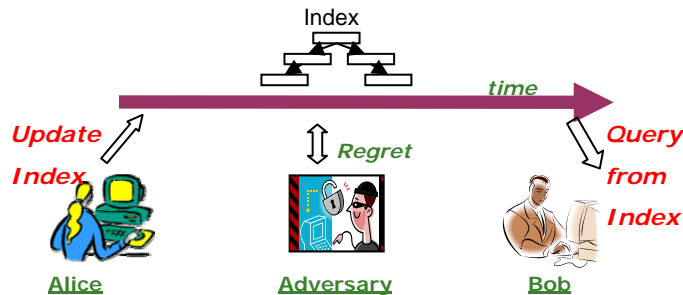


图 5: 威胁模型 3-----攻击者修改索引

记录的创建者提交记录的同时也更新了索引，记录的访问者通过索引来查询记录。攻击者由于是超级用户，所以可以从逻辑上来直接修改索引。该不可信的索引可能是存储在 WORM 里，也可能是存储在普通的存储介质里，但攻击者均可以通过添加新的路径来掩盖记录。

为了防止上述的攻击，我们提出了对可信索引的要求：

- 正确性：一旦记录创建，对该记录的索引项和路径都是不可更改的。更新索引的代码不会导致历史记录隐藏和修改。这意味着一旦记录被创建，则相应的索引更新也被提交到 WORM，除非 WORM 发生物理上的问题，否则该记录一定是通过创建时的索引来访问的。也就是说，记录的获取只依靠 WORM 存储设备的不可重写性。
- 持久耐用性：索引能支持记录的快速增长，对新记录即时更新，不需要周期性的删除和重建索引。如果没有即时更新，把记录放在缓存中，会面临该记录被修改的威胁。
- 索引必须支持高效的查询。
- 索引引起的空间增长必须是可接受的。
- 索引是不可分解的，已经过期的删除的数据不能通过索引被推测出来。

同时，对查询也要求是可信任的：即是正确的查询

3. 可信的数据隐私保护

3.1. 背景介绍

随着信息技术的发展，网络的连通性和磁盘存储空间的增大，整个社会正在经历一场数字化的革命，各种各样的电子数据不断产生，人们纷纷热衷于用数据来表达自己的意愿，来描述复杂的事物。对这场数字革命的到来欢欣鼓舞的人们，只沉醉于享受新事物的愉悦，毫无戒心的甚至争先恐后的将自己的个人信息公之于众，不管是填写各种注册信息还是使用各种诱人的服务，完全没有意识到自己已经毫无隐私可言。只有当清晨开机时那烦人的垃圾短信，以及邮箱里新增加的莫名其妙的垃圾邮件，才会让你好好反省是否应该保护自己的隐私！在使用各种服务时，一个不可避免的事实是只有提供更多的个人信息，才能获得更好的服务。

例如随着个人手持设备（PDA，Smart Phone 等）的普及，人们越来越多的使用基于位

置的服务（Location Based Service: LBS），包括紧急救援服务，基于位置的游戏，移动黄页服务等。虽然服务提供商不要求人们在请求服务的同时发送自己的唯一标志例如姓名，网络地址等，但要求用户发送自己的当前位置，只有个人位置信息越精确，获得的服务才越满意。在这种情况下，用户的位置就成为了个人隐私信息。服务商（攻击者）可以通过把用户位置和地图进行匹配以及某些经验观察来发现用户的真实身份，进而对用户的服务请求进行分析，发现用户的个人爱好等隐私。

另一方面，政府机构以及公共服务机构越来越多的发布包含个人信息的数据，比如医疗数据，选民数据等等，这些数据甚至可以作价出售。如果没有可信的隐私保护，那么攻击者将利用多个数据之间的联系来获得个人隐私信息。如图 6 所示，左图的医疗信息是从专门为政府雇员购买医疗保险的机构购买的，选民信息是从负责选举的机构处购买的。该医疗信息可以认为是匿名的，因为没有病人的姓名等唯一标志信息。但如右图所示，当攻击者把医疗信息和选民信息结合之后，通过出生日期，邮编，性别的匹配，就可以把选民姓名和疾病联系起来，从而获得了非常隐私的个人信息。

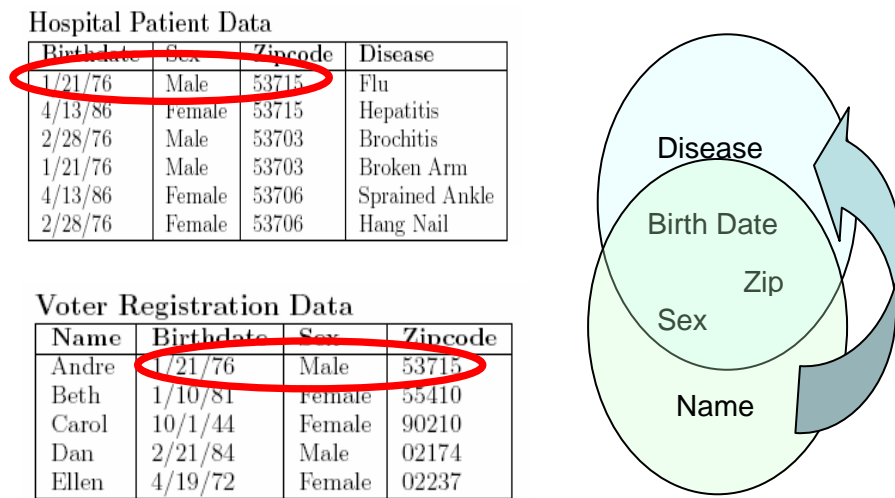


图 6：通过数据之间的匹配来识别隐私

3.2. 技术实现

3.2.1. 关系数据库环境

在不同的环境下，可信的数据隐私保护有不同的要求。公共发布数据环境下，包括公共医疗调查报告，选民数据发布等，要求首先必须隐藏能够惟一标志用户的个人信息，比如显式的名字。另外，还需要使某一个特定个人的信息不能从所有数据中被攻击者识别出来，比如生日，性别，邮编，电话号码等属性很容易被匹配到个人。当前普遍采用的一个方法是 k 匿名模型：一个关系满足 k 匿名，如果其中每一个元组所代表的个人信息都至少和关系中其他的 k-1 个元组不能区别。如图 7 所示：该关系中 Problem（疾病）是个人的隐私。在 Race, Birth, Gender, ZIP 属性上，每一个元组都至少包含了一个并发（相同的属性），所以攻击者不能识别出某一个特定个人的疾病信息。

Race	Birth	Gender	ZIP	Problem
Black	1965	m	0214*	short breath
Black	1965	m	0214*	chest pain
Black	1965	f	0213*	hypertension
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	obesity
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1964	m	0213*	obesity
White	1964	m	0213*	short breath
White	1967	m	0213*	chest pain
White	1967	m	0213*	chest pain

图 7: 满足 k 匿名的关系

3.2.2. LBS 环境

在 LBS 环境下，用户发送的位置越精确，获得的服务越好，但精确的位置更容易泄漏用户的隐私。为了保护用户的位置隐私，通常采用的方法是对用户的真实位置采用 cloaking 技术，使得用户的真实位置点 (location point) 被扩大为一个区域 (cloaking region)，服务提供商只能接收到该位置区域而不能识别用户的真实位置。使用广泛的是两种服务模型。

第一是采用匿名代理的三方模型，如图 8 所示：

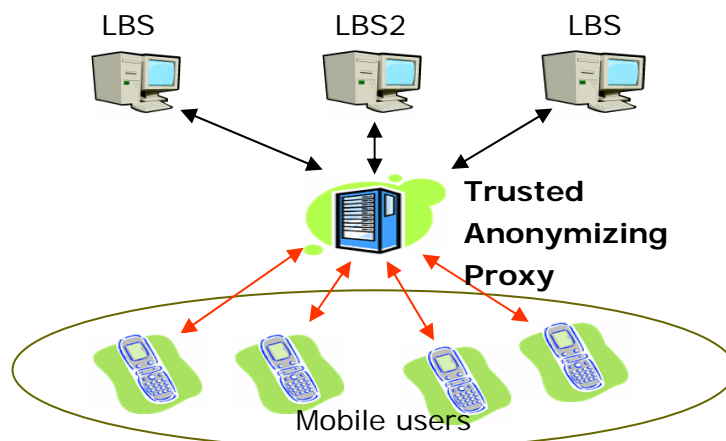


图 8: 可信隐私保护的 LBS 模型

该模型由三类对象组成：移动对象，可信的匿名代理，服务提供商。移动对象提出基于位置的服务请求，匿名代理对用户的位置进行匿名处理以保护用户的隐私，服务提供商提供各种不同的服务。在模型中信息的交互过程是：1) 移动对象向可信的第三方匿名代理提出服务请求，其中包含自己的服务内容和真实的位置信息 (location point)，还有一些关于隐私的要求。该请求通过安全的传输 (签名机制等) 到达匿名代理。2) 可信的匿名代理使用 cloaking 技术对用户的位置进行扩大，再把扩大后的位置区域和服务请求发送给服务提供商。3) 服务提供商响应服务请求，并把查询结果返回给匿名代理。4) 匿名代理对结果进行求精，选出最适合用户真实位置的结果返回给用户。

在该处理过程中，涉及到的关键技术是：如何进行位置的匿名处理；如何响应服务请求，

处理位置相关的查询；如何对结果进行求精。

匿名处理通常采用的方法也是 k 匿名模型。LBS 环境下的 k 匿名模型主要是针对位置信息的。如果某个用户的位置不能和其他 k-1 个用户的位置相区别，则该用户的位置满足 k 匿名。匿名代理在对用户的位置进行 cloaking 时，主要的问题就是怎么样找到一个合适的 cloaking region 使得它能够同时覆盖住 k 个用户的真实位置，并且还要满足用户的隐私要求（比如响应时间，匿名质量（cloaking region 的最小最大范围限制）等）。通常有两种找到 cloaking region 的方法。第一是静态的基于空间的划分方法。整个用户空间被预先划分为等大小的基本单元，用户的 cloaking region 为从用户所在的基本单元开始扩展而找到的最小的能够覆盖用户的并且满足隐私要求的区域。代表方法有基于 Quad-Tree 和基于 Grid。第二是动态的基于用户位置的划分方法。用户的 cloaking region 是从用户位置开始扩展，找到离自己最近的并且满足隐私要求的 k-1 个邻居之后所组成的区域。代表方法有基于 Graph 的。

处理位置相关的查询主要是处理 cloaking region 的查询，比如范围查询（range query），最近邻查询（knn query），采用方法主要是基于移动对象环境下的查询方法，代表方法有。。。

第二是用户-服务器结构的两方模型，如图 9 所示：

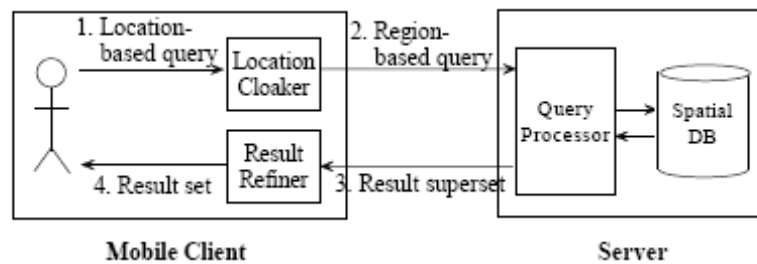


图 9：用户-服务器结构的 LBS 模型

该模型由两类对象组成：移动对象和服务器。移动对象能够使用 GPS 等设备来获取自己的位置并具有匿名处理及结果处理的能力，它和服务器直接进行通讯。在模型中信息的交互过程是：1) location cloaker 接受移动对象的服务请求，其中包含服务内容和真实的位置信息（location point），还有一些关于隐私的要求。2) location cloaker 使用 cloaking 技术对用户的位置进行扩大，再把扩大后的位置区域和服务请求发送给服务器。3) 服务器中的 query processor 响应服务请求，并把查询结果返回给用户的 result refiner。4) result refiner 对结果进行求精，并选出最适合用户真实位置的结果。

在该处理过程中的主要问题和上面的模型相似。

其他很多环境下，也需要考虑可信的隐私保护。比如 sensor 环境下，RFID 环境下等。

4. 可信的数据存储和可信的数据隐私保护的结合

可信的数据存储和可信的隐私保护是两种不同的契约规定，它们没有层次高低的区别，也没有相互依赖的关系，只是不同的应用要求，可以同时存在，也可以单独实现。只是不同的情况下对数据管理系统的实现提出不同的要求。类似安然案件这样的情况，公司的自身的运营情况记录等需要完全真实的提供给国家，审计者，投资者，调查者和所有公民个人以利于决策的，我们要求数据的存储是可信的，这里不要求隐私保护。

类似医院的病例记录，保险公司受理的保险记录等这些涉及到公众的私人信息的，同时也要完全真实的提供给国家的，我们要求数据的存储是可信的，但是对外发布的时候还要求可信的隐私保护（也就是发布公共信息的隐私问题）

类似用户请求服务时，用户的真实信息必须有一个第三方来存储，但这里的存储不一定要求可信，或者可信的要求更弱。由第三方再向外发布的时候要求可信的隐私保护。

5. 可信数据操作：外包数据库 —— 新的展望

对于简单的应用，也许仅仅有可靠的数据存储，有可靠的数据隐私保护，就能满足我们的需要。对于复杂的应用场景，我们往往还需要服务商能够提供针对这些数据的访问服务。一个提供数据库外包的服务商需要能够提供相关的查询，更新，访问控制等的操作。

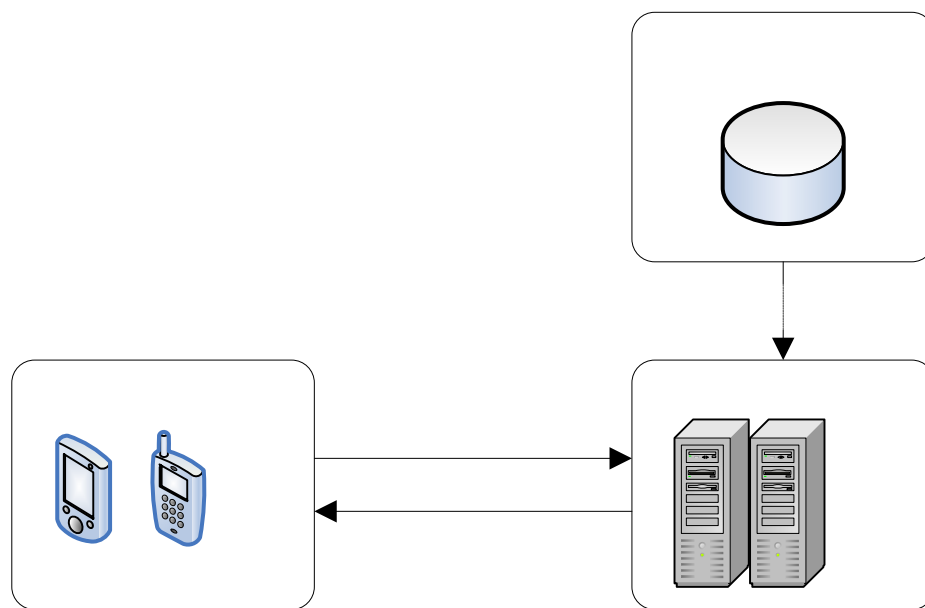
数据库外包会跟很多小的企业应用带来很多好处：可以避免维护数据库软件，硬件；可以节省人员 DBA 的开销等等。但是，数据库外包也会带来很多的挑战，例如：数据从远程访问，网络访问延时的影响；数据的安全性、隐私的考虑；数据操作的安全性考虑。

现今网络技术正在以惊人的速度发展，千兆网，万兆网的出现都是提供远程数据库服务成为了可能。对于数据的安全性，我们也可以通过加密数据来使得我们的数据不会被一个恶意的服务提供商获得。

但是还有一个非常重要的，非常有挑战性的问题在于我们如何保证在这些加密后数据上操作的可信性，换句话说，我们如何来保证我们的查询、更新操作真正得到了正确的、完全的执行？

我们说这个问题需要一分为二来看：首先，由于数据经过加密，在这些加密后的数据上的操作必然会受到限制，这是挑战之一；其次，假设我们在加密后的数据上能够执行查询，我们怎么去验证一个数据库服务提供商正确的执行了我们的查询、更新操作？我们将在下面的讨论中考虑上面的问题。

在讨论具体的问题之前，我们先来看看一个比较通用的数据库外包场景：



可以看到，在这个典型的数据外包场景中，一个数据所有者（Database Owner）将所有加密后安全的数据存放在一个数据库外包服务提供商（Service Provider），所有的用户（User Device）通过一个加密后的查询到服务提供商查询需要的数据。

在这样一个典型的数据外包场景中，一个值得注意的问题是我们的用户设备往往是一些存储和计算能力都有限的小型的手持设备，那么我们如何在这种受限的情况如何去验证一个不可信的服务提供商是否正确完整地执行我们的查询、更新操作，是否可信就是一个具有相当挑战性的任务。

下面我们将分两个部分来讨论，首先是如何在一个加密后的数据源上执行查询，然后讨论如何验证一个数据库外包提供商是否可信。

5.1. 在加密后的数据上执行查询

对于一些比较敏感的数据，为了避免一个恶意的服务提供商去利用数据去获取利益，我们往往需要将其加密，但是在加密后的数据上往往难于执行灵活的查询，因为，对于一个非等值查询来说，由于传统的加密算法不能保证加密后的数据和加密前的数据由相同的顺序，所以只能支持等值查询。

一部分研究工作中指出，我们可以通过把每一个准确值抽象成一个区间，然后通过在服务提供商存储这个抽象的区间而不存放原来的每个精确值来实现数据的加密，同时也可以保证我们能够执行所有的查询。但是这种方法由于丢失到数据的准确值返回的结果往往是一个结果的超集，需要用户对返回的结果进行过滤，这样会带来额外的代价。同时这种方法也面临着一个不可调和的问题就是，为了减小用户的代价，我们需要抽象的区间尽可能精确，但是抽象的区间尽可能精确又会造成安全性的降低，所以这种方法有其本身的缺陷。

另一部分最新的工作通过将原来的数据映射到一组新的数据上来达到加密的目的，这种方法维护数据项在加密后加密前数据中的顺序性来保证可查询性，这种方法没有之前方法存在的种种问题，而且查询返回的是一组准确的结果，所有较之之前的方法有较大的优势。

5.2. 查询的正确性，完全性检查

在数据库外包中，判断一个服务提供商是否可信，两个最为重要的条件就是：用户提交的查询返回的结果是否正确？用户提交的查询返回的结果是否完全？

正确性：

查询返回的结果是否正确是指，服务提供商返回的所有元组是否是原来数据库中的元组，服务提供商有没有修改我们的元组或者有没有返回一个恶意生成的元组。

如果一个服务提供商不能够满足查询的正确性，我们说这个服务提供商一定不是可信的。

因此，在之前的工作中，有很多的工作对这个问题进行了讨论，提供了各种加密，签名的方法来验证一个服务提供商是否满足查询正确性。

完全性：

一个服务提供商如果能够满足查询正确性，我们依然不能判定这个服务提供商是可信的，因为一个恶意的服务提供商还有可能删除部分数据，或者只返回原来结果的一部分。所以除了验证一个服务提供商的查询正确性之外，我们还需要知道一个服务提供商的查询完全性。

所谓的查询完全性是指，对于给定的一个查询，服务提供商能够给我们返回所有的查询结果，没有遗失任何一个正确的查询结果。

查询完全性的验证是一个非常重要的，非常具有挑战性的课题。之前的所有的工作往往都只能支持非常有限的查询类型，而且对数据由很强的假设，需要建立各种索引数据结构来完成验证的过程，所以非常具有局限性。

所以，能够提出一种开销小的，能够支持各种查询的完全性验证方法非常的有意义。

6. 研究点归纳

(1) 继续分析“可信的数据库”在不同行业的应用需求及不同的处理方式，并区别它和传统概念如安全性、保密性。

(2) 根据可信的本质特征及（法规）要求，提出相应的实现技术或解决方案。可信的数据存储从技术角度来看，既可以通过数据存储于数据库之后的内部实现机制（如存储方式、索引结构等）来从根本上对法律规范进行支持和遵从，又可以借助外部监测技术来监督和验证数据库系统的可信性。此外，还可以提出与技术相结合的一整套解决方案，如使用哪些（已成熟）的技术经过何种步骤来最终保证数据库中数据的可信性。

(3) 继续研究可信的数据隐私保护和可信的数据外包服务。