# 云存储安全

云存储安全这一部分主要从云存储安全的背景、云存储面临的安全问题以及保证云存储安全的关键技术三个方面进行云存储安全问题的分析。

云存储，顾名思义，就是将数据存储在云端的一种存储方式。这些数据包括个人档案、照片、文件和受版权保护的内容。云存储是伴随着云计算发展而来的，是云计算的资源存储，旨在通过互联网为用户提供更强的存储服务。云存储给我们提供了一种全新的存储模式，但凡事有利必有弊，跟互联网相关的事物也必然伴随着安全问题。通常情况下，云存储服务以数据的安全存储和管理为核心，相对于传统的网络应用，在云端进行存储和相关业务的处理会导致用户数据的所有权与管理权分离，由于云存储的特性使数据脱离了数据拥有者的物理掌控，这导致云存储服务的安全性、可靠性以及可用性都面临着巨大的挑战。 一方面，为了利用云平台的计算和存储资源，用户需要将数据存储到云服务器或者计算任务外包给云服务器，这将泄露用户的敏感数据和计算结果。另一方面，云服务器一旦出现问题，大量用户的数据与应用将会无法使用和运行。数据存储在数据中心中，通常不加密，或者没有加密到足够高的水平;更重要的是，大公司更容易受到人为错误的影响。其结果是由于技术错误或攻击者们的恶意目的导致信息丢失或泄漏高风险。 因此，随着云存储的不断普及，云存储安全也开始受到大众的关注。对于云用户来说，把大量的业务和数据放置于云端是一件冒险的事情 。例如，今年三月，全球20亿用户的Facebook卷入了“史上最大个人信息泄露风波”。3月17日，英国的“剑桥分析”数据公司被曝光非法收集 Facebook公司5000万用户的个人信息，受此影响，3月19日，Facebook公司股票一路狂跌，仅仅两天，市值蒸发了500亿美元……毫无疑问，对于公司和公众来说，数据泄露带来的影响和代价都是惨重的。 有数据表明：目前，大约有80%左右的企业不愿意将企业内的业务数据放在公有云上，主要是从数据安全性的角度考虑。 在这样的环境下，云端数据存储的安全问题自然而然地成为当下最受用户关心的问题之一。 云厂商如果要得到大家的信任，必须要让企业的数据是安全的，不会被窃听，不会被欺骗，而且云厂商能从一而终的提供服务。

云存储的安全问题存在于很多方面，比如，数据机密性：用户为了获得云存储服务，他们需要首先上传其本地数据存储到服务器中，而在需要的时候也要从服务器上下载数据。这一过程中数据需要经历公有云、私有云以及互联网上的各种通信线路的传输。数据存储在服务器上的时候，有可能会遭受非法窃取或篡改。这些情况都会使用户的重要敏感数据面临机密性的威胁；数据隔离问题：云计算最重要的一个核心就是云的虚拟化问题，对不同的云用户来说，应用的云存储系统是一个相同的物理系统，不再像传统网络一样有物理的隔离和防护边界，就存在虚拟系统被越界访问等无法保证信息隔离性的问题；应用安全：对于运行在云存储平台上的云应用，如果其本身未遵循安全规则或存在应用安全漏洞，就可能导致云存储数据被非法访问或破坏等问题；用户隐私：用户购买存储服务时，必然要向供应商提供一些个人信息，如信用卡账号，用户名等，这些信息都有可能被泄露。此外，用户的数字身份、证书以及存取操作记录等也需要隐私保护；版权风险：版权问题目前已大范围地出现在网盘服务中。一些个人或团体会将文件通过云存储的客户端上传至网盘中，然后通过分享的方式对其提供下载，大量的有版权的文件被通过这种特殊盗版方式进行传播，在这种情况下面临侵权问题的不仅仅是用户，还有云存储的提供商；运营停止：在目前这种互联网环境下，提供给公众的云存储服务每年投入的资金庞大，而且云存储盈利模式还并不清晰，这些都会导致服务商可能在一定时间后关停服务，那么用户数据留存问题就变成了最大的隐患。

基于上面所述的云存储所面临的安全问题，国内外学者在云存储数据安全进行了大量研究，主要集中在数据完整性审计、密文数据去重、可靠数据删除以及高效密文检索这4 个方面。

**1.完整性审计**

云数据存储解决了用户内存不足、硬件维护成本高等问题，但也带来了新的问题。事实上，用户并不完全清楚自己在云上数据的真实情况。为了保证数据在云上完整且正确地存储，用户需要定期对云服务器上的数据进行审计，因此需要用户能够高效地对云服务器存储数据的完整性和可用性进行审计。而事实上，用户并不会将云数据全部取回进行验证，因为海量的数据取回会消耗大量带宽以及终端的计算资源。因此，用户通过下载或使用关于原始文件的部分信息来执行云端数据的完整性验证。数据完整性审计是云存储的重要安全技术之一，用于用户（或审计者）验证其存储于云端的数据是否保持完整。完整性审计主要包含两种审计机制：可证明数据持有（Provable Data Possession，PDP）和可证明数据可恢复（Proof of Retrievability，PoR）。这两个方案都是基于审计方（包括数据拥有者）与云服务器通过挑战- 响应协议来验证数据的准确性或可恢复性。区别在于，PDP 在效率方面有巨大优势，在判断云中数据是否损坏上更加快速。而PoR 机制对所存储的数据进行了预先编码处理。因此，PoR 方案不仅能验证数据是否被损坏，并且能在数据受到一定损坏时恢复数据。随着学术研究的发展，PDP 和PoR 这两个方向的研究也在高效性、动态数据支持方面不断完善。

2003 年，远程数据存在验证的概念被提出后表明用户可以对文件进行次数的完整性验证。在此基础上，基于公钥密码体制的可证明数据拥有模型PDP 和可证明恢复模型PoR被提出，这两种模型分别用来检测存储在半可信服务器中的数据是否完整和检测数据是否可恢复。近年来，学者们对支持动态更新的PDP方案进行了大量研究，提出了很多特殊场景下（如多用户、带有重复数据、支持密文检索等）的数据完整性审计方案。

以PDP 为例，首先，用户对文件（明文或者密文）进行预处理和上传，包括数据的分块并计算数据块的校验标签。标签本质上是对数据块进行签名处理后获得的信息，他可以证明文件是否完整。现有两种方法对标签进行计算，一是基于对称密码学的方式，二是将公钥密码学用于计算标签。上传过程中用户将数据块以及对应的校验标签存储在云服务器上并删除本地备份，本地只保留与完整性审计相关的密钥信息。然后，用户再对云端数据进行完整性审计，审计者可以通过挑战—响应协议来完成数据完整性审计。其中，为了降低审计的通信开销，云服务器可以利用标签的同态性质，将与数据块相关的标签聚合成一个同态标签并反馈给审计者。审计者在验证阶段可以利用相关的密码参数验证云服务器上存储数据的完整性。如果在审计验证阶段输入的密码参数不包括数据拥有者的私密信息（如私钥），那么该审计者可以是除数据拥有者之外的任何第三方。基于该性质，我们可以称这种审计方案是公开可审计的。否则，完整性审计只能由数据拥有者完成，那么方案则是私有可审计的。

**2.密文数据去重**

在云存储服务带来诸多便利的同时，服务器中大量的冗余数据成为限制云存储发展的另一瓶颈。根据EMC 的调查报告显示，随着云端数据的快速增长，云存储中的冗余数据在备份应用中达到80% 以上，在文件系统中已经达到60% 以上，这些冗余数据消耗着大量的存储资源和管理资源。因此，如何通过删除重复数据从而保证数据存储的高效性成为了亟待解决的问题。为了解决上述问题，数据去重技术应运而生，通过数据去重技术可以消除数据冗余，相同文件只保留一个物理副本，从而有效降低用户端上传数据耗费的带宽及节省服务器端的存储空间。对于云计算平台而言，数据去重不仅可以减少存储时付出的硬件成本，还可以提升存储空间的利用率，提高访问效率。数据去重存储技术目前广泛应用于商业云存储及数据备份服务，例如Dropbox、Mozy 和Memopal 等商业存储中。然而现有的绝大多数去重方案仅适用于明文数据。为了保护敏感数据的隐私性，用户在上传敏感数据之前通常会使用加密算法对数据加密，由于不同的用户随机选取的密钥不同，因此，即使是相同的明文数据也会被加密成不同的密文，这使得云服务器无法判断两个密文数据是否是由相同的明文数据加密得到的。为了解决传统加密方案与去重方案不兼容的问题，收敛加密技术营运而生，通过将文件的哈希值作为其加密密钥的方案，收敛加密算法使得即使是不同的用户在不需要通信的前提下也能得到同样的密钥，这使得密文去重得以实现。

用户在上传文件之前，首先对文件F 进行哈希操作，得到加密密钥k，若云服务器中的文件与用户所拥有的文件相同，则得到的密钥k 也相同，然后基于密钥k 对文件F 进行加密，得到密文C，对密文C 进行哈希得到标签T。然后用户将标签T 上传到云服务器，云服务器根据得到的标签T和原来存储的数据标签进行对比，并将结果返回给用户。若用户得到云服务器返回结果是“重复”，则用户需要向云服务器进行数据拥有证明，证明自己拥有此数据；用户若收到结果为“没有重复”，则将密文及相应标签上传到云服务器进行存储。

通常，数据安全去重技术按照去重粒度可分为文件级去重和块级去重，顾名思义，在文件级去重中文件是去重执行的最小单元，服务器根据文件的标签进行去重并保留唯一文件副本。块级去重中，数据块是去重执行的最小单元，而服务器根据块标签进行去重并保留唯一的数据块副本。按照去重框架可分为服务器端去重和客户端去重，顾名思义，服务器端去重指的是在服务器端进行去重操作，用户不参与也不知道数据是否被执行去重；客户端去重是指去重在客户端进行，用户只上传不重复的数据给服务器。

**3.可靠数据删除**

由于云存储中用户的数据外包存储在云服务器上，这导致了用户数据的所有权与管理权分离，使得数据不在可控的范围内。对于用户来说，其外包的数据往往可能包含大量的敏感信息，这些敏感信息都可能会泄露用户的隐私。由于云服务器的不完全可信，传统的本地数据删除方法也无法适用于云场景中，云存储中的用户难以确信其数据是否被服务器安全删除。因此，可靠的数据删除服务就显得尤为重要。用户对可靠的数据删除的需求包括两个方面：一方面，用户希望在数据的生命期结束或它向云服务提供商请求删除数据后，这一数据就应该是永远不可访问的；另一方面，为了保证云存储的容灾能力，云服务提供商通常会保存用户的多个副本，但是在执行数据删除时，用户希望云服务提供商删除所有的数据副本。现有的数据删除技术主要包含两种：安全覆写方法和密码学删除方法。安全覆写方法是一种物理删除数据的方法，它对数据进行删除操作时首先需要对数据进行破坏，然后使用新的数据在原有数据的位置上进行覆写，从而使得用户数据无法恢复，进而实现可靠的数据删除；密码学删除方法指的是用户在上传敏感数据文件之前，通常会先对其数据文件进行加密操作，然后再将密文数据上传到云服务器上；用户所有数据文件的密钥按树形结构组织，然后将主密钥保存在物理安全的存储介质中，其他数据和加密的密钥树则被保存在一般存储介质。当数据需要被删除时，将密钥管理者所持有的密钥进行删除，也就是说即使云服务器保留了这一被加密的数据的密文，但是因为无法拥有相应的解密密钥，存储在云服务器上的密文数据无法解密，这样就能保证用户存储在云端数据的安全删除。此外，为了实现对每个数据文件的可操作性，每个数据需要单独选择不同的加密密钥，并在本地保留尽量少的主密钥，其它密钥通过主密钥加密方式存储在服务器中。

**4.高效密文检索**

许多的敏感信息如电话、个人材料、图像等存储在云端上。因此，为了保护客户的敏感信息，客户在上传数据到云端之前对数据进行加密处理。然而，加密技术破坏了数据原有的状态，使得在极其多的密文文件中搜索指定的文件变得十分困难。传统的明文检索方式不再适用于密文数据，高效的密文数据检索技术是保护云存储数据安全的一个关键技术。 在密文数据检索中，主要包含关键词搜索、多关键词检索、模糊关键词检索等多种检索方式。现有的加密数据检索方案分为4 种类型：

1. 利用索引访问类型。用户根据文件数据的关键词建立具有隐私保护性质的索引，其中索引与加密数据一一对应。检索时，只需根据索引查找检索请求，再由索引的信息返回相关文件。
2. 利用陷门信息访问类型。最常用的方法是利用文件关键词构造陷门信息，即用户的访问信息，从而可以直接使用文档的关键词检索加密文件，从而获得需要的资源。
3. 利用秘密共享方案访问类型。数据拥有者对数据计算多个隐私信息，并将隐私数据分散地存在不同的云中，只要服务器之间不相互勾结，则该方法能够对数据隐私进行保护。
4. 利用同态加密的访问类型。存在一些同态加密函数能够实现对密文数据的相关操作直接作用于明文数据上，因此可是现实密文状态下的信息检索。但同态加密在实际使用中计算开销和通信开销较大，不适合实际使用。

云环境下的安全密文检索过程通常为：首先，数据拥有者根据文件的明文信息提取关键词创建与文件相对应的索引表，然后对文件数据和索引进行加密，并将加密后的文件和密文索引发给云进行存储。当有客户需要检索指定文件时，客户根据需求向数据拥有者获取文件解密密钥以及生成索引陷门的密钥，并通过检索的文件创建检索请求，发给云服务器。云服务器收到客户的检索请求后执行检索操作，并将检索后的结果返回给客户。根据方案中使用的不同的密码体制，可搜索加密技术分为以下两个技术：对称可搜索加密技术和公钥可搜索加密技术。其中对称可搜索加密技术在使用过程中，数据拥有者和用户之间需要进行密钥协商，因此只有合法用户才能生成检索请求和解密密文。因为对称可搜索技术计算速度快的优势，已经逐步成为了学术研究的焦点。其中早期的研究工作主要集中在单关键词检索，之后大量的方案被提出，如相似性检索, 多关键词检索，模糊关键词等，使得可搜索加密技术应用更加广范。2004 年公钥密码被引入可搜索领域，提出了基于关键词的公钥可搜索加密方案。对于公钥可搜索加密技术中，用户利用共享者的公钥进行加密，双方不需要进行交互，因此应用场景更为广阔。

总的来说，云存储是伴随着云计算而生的，是大数据时代的产物。由于云存储是运行在云端的，就必然存在云存储安全问题，以上只是解决云存储安全问题的几种关键技术，关于云存储安全的研究还有很多。同时，云存储安全还面临着很多新的挑战，比如现在云存储的成本问题， 目前不管是国内还是国外，云存储的成本大概在每G每月2毛钱左右， 应该提高技术来减少云存储的成本。同时，在倡导低碳环保的今天，如何在提高云存储效率的同时还能保证安全，低碳、环保、绿色也是一个未来要考虑的问题。