一、选题背景

口 云技术具有优秀的可扩展性和可访问性，但是它的流行也导致了人们对数据隐私性和安全性更高的需求。为了保护数据的安全性，我们可以选择加密数据，但是数据加密也会带来另外的麻烦。举个例子，有一个托管在云服务器上的Web服务使用了大的量XML文档（dang）来存储数据。为了安全起见，文档是加密的。这时，如果Web服务的提供者想从这些文档中提取数据并计算的话是比较麻烦的。一个简单的解决方案是下载和解密所有文档，然后再提取数据，不过这在带宽和计算上都很昂贵。所以，一个可以直接对加密XML文档进行搜索和计算的方案就显得很重要了。

另外，目前未加密数据的搜索技术较为先进，而加密数据的搜索技术还有很大的进步空间。

1. 国内外相关研究情况

目前从加密的XML文件获取总体数据的问题并没有一个有效的解决方案，虽然存在这些文件的查询工作；

目前存在对加密的XML文档进行查询的方案，但是还没有一个很好的从加密的XML文件获取总体数据的解决方案；

.....

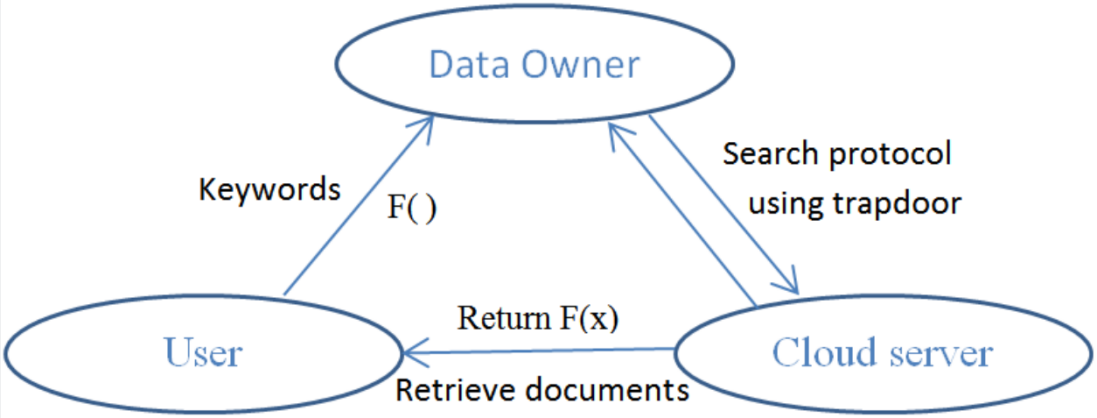
1. 研究内容

1、我研究的是处理大量XML格式的加密文档的问题，其中用户可以对XML中的某些元素进行搜索或计算。

2、用户可以进行的搜索包括多关键字搜索和短语搜索。

3、最终的解决方案要求可以用于部分或完全加密的XML文档，并允许在服务器端执行安全的计算，实现 求和、求平均值 等函数的计算查询。

这个方案的通信模型在公有云中涉及到三方：数据所有者、云服务器和用户。在私有云中，用户就是数据所有者。这个方案的搜索功能的参数有关键字和函数。首先，用户向数据所有者发送关键字kwi 和函数F(x)来发起查询请求。然后数据所有者加密请求并将其发送给云服务器。最终，云服务器向用户响应他所请求的文档以及函数的结果。



四、研究思路

1、这个解决方案利用索引表来快速地查询关键字所在的文档以及文档中的位置，用来实现搜索功能。

2、而为了在不信任的服务器上执行计算，我们需要用到同态加密。并且考虑将同态加密和对称加密一起使用，以减少计算和存储成本。

五、关键技术

1、倒排索引

它的实现过程是，首先，将所有文档都解析为关键字列表，然后，对关键字列表进行去重，最终，为每一个关键字生成倒排列表，它包含引用该关键字的所有文档。

1. 关键字位置索引

这也是一个倒排索引，它包含该关键字在一个文档中的所有位置，我们可以用它来实现短语搜索。

1. 同态加密

我们同态加密那些可能用于计算的数据，就可以对密文数据执行计算了。我们可以使用Paillier（胚丽而）密码系统进行同态加密，它可以实现加法和数乘的同态。

1. 对称加密

AES等对称加密算法比同态加密算法具有更高的存储和计算效率，所以我们可以使用对称加密算法加密那些不用于计算的数据，这种结合对称加密和同态加密的方案既能降低成本又能保证计算功能的实现。

1. 实践难点
2. 标签、属性和注释也可能包含查询关键字，这些情况下搜索功能很可能会返回错误的匹配结果。所以我们在使用短语搜索时，应该将某些符号作为关键字来检索，比如<、>和/，它们可以用来区分标签名和文本。
3. 口 当XML文档的规模不断扩大时，倒排列表占用的内存也会迅速增大。
4. 口 我需要足够多足够大的XML文档作为数据库，来检验查询和计算功能，另外还要对文档进行加密，所以有合适的 XML 数据库才能更好地实现这些功能。
5. 总结

1、这个解决方案除了实现最基础的多关键字搜索功能外，它还通过倒排索引实现了短语搜索功能，通过同态加密实现了计算功能。

2、它的优点是，结合同态加密和对称加密的方式大大降低了成本，提高了效率。

3、它的缺点是，尽管计算效率很高，但是很容易受到统计分析的攻击。特别是，当文档使用了一个已知明文模型中常用的标签，那么一些有意义的信息很可能会被泄露给云服务器。所以我接下来的工作，是完善这个解决方案以在保证计算效率的同时提高系统在统计分析方面的安全性（原则：扩散与混淆）。