**2023年网络技术挑战赛**

**作品设计文档**

**作品名称：基于全同态加密的医疗保险云数据隐私保护系统**

**所在赛道与赛项**： **A-ST**

**目录**

[1. 作品背景及意义 3](#_Toc26390)

[1.1. 保险隐私数据泄露问题 3](#_Toc26405)

[1.2. 云平台数据隐私保护问题 3](#_Toc3939)

[2. 现状分析 4](#_Toc7067)

[3. 作品特色功能 6](#_Toc11658)

[3.1. 基于全同态加密（CKKS）的安全存储与隐私计算 6](#_Toc17318)

[3.2. 基于SM4对称加密的非计算数据安全存储 7](#_Toc3723)

[3.3. 基于关键词的密文检索 7](#_Toc29534)

[4. 作品设计 8](#_Toc27851)

[4.1. 系统架构 8](#_Toc24451)

[4.2. 电子保单加密结构设计 9](#_Toc9809)

[4.3. 基于全同态加密和SM4对称加密的加密方案设计 9](#_Toc311)

[4.3.1全同态加密(CKKS) 9](#_Toc32260)

[4.3.2 SM4对称加密 12](#_Toc19662)

[4.4. 基于关键词的密文检索方案设计 13](#_Toc1999)

1. **作品背景及意义**
   1. **保险隐私数据泄露问题**

在数字化时代，除违法违规使用和收集用户信息外，因数据管理存在疏漏和导致用户信息泄露的事件同样值得关注与警惕。聚焦于保险业，对于手握大量用户个人身份信息、资金支付信息以及健康、习惯等数据的保险机构而言，保护用户信息尤为关键。

近年来，保险公司数据泄露事件屡见不鲜。2021年5月，法国安盛保险公司声明，其在亚洲的运营遭勒索软件攻击，造成安盛在泰国的下属机构Inter Partners Asia处理的部分数据泄露。2020年末，瑞典最大的保险公司Folksam证实，其在一次内部审计中，发现与数字合作伙伴共享了近百万客户的个人信息，这些信息在社交媒体上泄露。

2021年3月19日，银保监会消保局公布的罚单显示中信银行被处以450万元罚款。有消息称，该罚单疑似为2020年5月，脱口秀艺人池子举报中信银行违规私自对外提供其银行流水信息事件的处罚结果。

在2021年1月21日至2021年3月1日期间，美国第二大汽车保险公司Geico遭遇数据泄露，威胁行为人盗取投保人的驾照长达一个多月。威胁行为人滥用在线销售门户网站获取保单持有人的驾照号码并计划使用驾照号码以投保人的名义申请失业救济。

用户数据隐私对于保险机构非常重要，一方面源自于监管要求，另一方面，则是因为数据保密是客户与保险主体构建信任的关键，一旦发生数据泄露事件，无论是因为内控问题还是遭受攻击，都会导致用户对保险机构的信任缺失，造成难以挽回的影响。保险机构掌握着大量投保人个人身份信息、支付信息、健康数据、保单记录等，涉及消费者个人隐私及财产安全。监管也在数度强调信息安全重要性，根据《保险法》等法规文件，保险机构对在办理保险业务中获取的投保人、被保险人、受益人的业务和财产情况及个人隐私，负有保密义务。

* 1. **云平台数据隐私保护问题**

随着互联网的迅速普及，云计算和云平台等服务的推出，对网络信息安全提出了更高的要求。云平台是一种基于云计算技术的服务平台，它提供了一系列的计算、存储、网络、安全等基础设施服务和应用服务，可以帮助用户快速、灵活地部署和管理各种应用程序和服务。云平台通常包括公共云平台、私有云平台和混合云平台等多种形式，用户可以根据自己的需求选择不同的云平台来满足不同的业务需求。在云平台上，用户不需要关心底层的硬件和软件设施，只需要关注自己的应用程序和服务，通过云平台的服务接口和管理工具，可以快速实现应用程序和服务的部署和管理。在云平台中，各类信息在远程处理，或与远程协作处理，信息系统所处的环境不再是本地的、封闭的、个人的，而是远程的、开放的、共享的。

使用云平台可简化业务建立流程，实现高效率部署，通过将基础设施云化，随需扩展或收缩资源规模，可以提高IT资源利用效率，降低IT资源管理成本，节约IT总成本。在医疗保险行业，医疗数据共享可以方便医疗保险公司对被保险人的健康状况进行评估，从而做到为不同顾客提供个性化的保险方案和定价策略；减少保险产品风险，降低理赔成本；精准识别风险群体，指定更加有效的风险控制策略。因此，医疗数据共享是医疗保险行业发展的一大趋势，而使用云平台可以为医疗保险公司后续进行医疗数据共享建立良好基础。现有医疗保险公司也逐渐将部分业务转移到云平台。

不过，虽然云平台、云计算服务的出现加速了互联网时代的到来，但同时也给用户隐私保护方面带来了全新挑战。云平台的出现引发了一系列具有争议性的问题：在特定的服务中，不同客户之间的数据溢出；供应商系统遭到大量恶意软件攻击；“恶意人群操作”导致的资料外泄；共享信息的不安全性；以及黑客盗取供应商系统的数据。

虽然云平台在大量数据的存储管理方面有了极大的提高，但由于用户不放心将自己的隐私数据完全以明文方式存放在虚无飘渺的云端，云平台越来越多地考虑将用户数据加密后进行存储管理。常规加密方式无法对加密后的数据直接进行运算，这导致云平台的计算能力被极大程度的削弱。为此我们需要找到一种合适的加密算法，保证使用者能在云平台上直接对数据密文进行检索和统计分析。

1. **现状分析**

1978 年，3 位著名的密码学家 Rivest、Shamir 和 Adleman 第一次提出了“隐私同态”概念，之后立即引起了学术界广泛的关注。同态加密技术是一种特殊的加密技术，可以在加密状态下进行数据的加、减、乘等运算，而不需要先进行解密。简单来说，同态加密可以实现在不破坏数据隐私的前提下进行计算，对数据进行加密处理后，仍然可以进行各种计算操作，而不需要解密操作，最后再将计算结果解密得到明文结果。同态加密技术可以分为加法同态和乘法同态两种类型。加法同态允许在加密状态下进行加法运算，而乘法同态允许在加密状态下进行乘法运算。同时，还有一些高级的同态加密技术，如全同态加密，它是同态加密技术的高级形式，可以实现对任意运算进行加密计算，包括加法、乘法、比较、排序等。

同态加密的应用场景非常广泛，例如在云计算、大数据分析、人工智能等领域中，用户可以将数据进行加密后上传到云端进行计算，计算结果再通过加密的方式返回。这样可以有效保护数据隐私，同时也可以充分利用云端计算资源，加速计算速度。虽然目前同态加密还未被大范围实用，但是已经有学者在一些场景中提出了可行方案并取得了一定的研究成果，比如物联网、电子健康、电子投票方面等。

（1）物联网

近年来，物联网技术被广泛应用在智能交通、智能家居、环境监测、供应链管理等方面，带来了较好的社会效益和经济效益，但也面临着隐私泄露的问题。在由标签、传感器、基站构成的物联网结构中，节点负责收集数据信息并选择一条路径将信息安全地传输给基站。数据传输节点会比较目标节点与邻居节点间的相遇概率并选择最大相遇概率的节点作为中间节点来实现数据信息的快速转发。如果以明文方式发送相遇概率则会暴露节点间的交互状态隐私，为了在保证节点能够正确选择最佳路径发送数据信息的同时保护节点隐私，就必须对相遇概率进行加密。

Zouari 等人提出了一个将 ElGamal 同态加密和秘密共享技术相结合的隐私保护方案。该方案起到了一个安全中介的作用，通过将采集到的数据从感知层传输到应用层，同时保持了数据的完整性和机密性。该方案在不暴露任何一个节点信息的情况下，允许多个节点信息安全有效的组合。测试结果表明该方案达成了安全性和效率的最优平衡。李超良等人提出了一种基于同态加密的物联网隐私保护计算方案，对节点之间的相遇概率进行加密，以密文形式比较节点间相遇概率，使节点收集的数据信息能够以最佳路径发送给目标节点的同时保护了相关节点的隐私。

1. 电子健康

在众多医疗信息系统中，电子健康档案（Electronic Health Record，EHR） 系统作为记录居民各项健康信息数据的载体和各类健康相关活动的管理平台，己成为医疗信息化建设的一个重要部分。电子健康档案系统高效整合资源，记录了个人疾病发生、发展、治疗转归的过程，有利于改善居民健康状况，具有很高的医疗价值。

但是与此同时，系统所面临的信息安全风险大大提高，并逐渐成为制约其发展的瓶颈。通常这些健康数据涉及到个人隐私，敏感性极高，一旦泄露后果严重。因此，电子健康档案系统的数据安全研究并不仅限于数据的存储安全，还有数据的使用安全。针对上述问题，郭俊彦提出了一种基于属性的全同态加密算法，将同态加密应用到用户健康数据的存储过程中，并构建了一个基于同态加密的电子健康档案系统。通过改造 GSW 算法构造了基于矩阵重编码方案的对称加密体制，然后在对称加密体制的基础上构造了基于密钥策略的属性基全同态方案。

1. 电子投票

随着网络通信技术和信息安全技术的不断发展，新的投票方式——电子投票诞生，电子投票由于具有高效、快捷、成本低等特性，逐渐替代了传统的纸质投票和电话投票，成为目前各种政治及日常活动中常用的民意统计方式。电子投票系统的保密性和安全性是首要需求，要防止选票信息和投票人信息泄露。目前的电子投票系统都是在投票时对选票数据加密，在计票时将其解密并统计投票结果，这个过程存在选票被恶意篡改或窃取的可能，影响了投票的公平性和保密性。

同态加密可以很好地解决上述电子投票中的保密性问题。基于同态加密的电子投票方案通常是在投票阶段将投票人的选票信息用同态加密算法进行加密，计票时不需解密而是直接对密文进行计算，最后由权威中心对密文计算结果进行解密并公示，从而在保证投票匿名的同时也有效地防止了选票数据泄露和被篡改。

通过以上对同态加密应用场景的研究可以看出，同态加密目前主要被应用于需要进行隐私保护的数据存储、传输及计算的场景中，尤其在电子政务（电子投票）、医疗中都已有学者提出了较为实用的方案。在医疗保险系统中，个人健康信息及金融信息通常是非常敏感、不可泄露的，且部分数据需要进行计算，因此同态加密技术非常适用。

1. **作品特色功能**

本作品从当下云平台的数据安全性隐忧出发，计划设计一套基于全同态加密的医疗保险云数据隐私保护系统，旨在为各大保险公司提供数据保密存储与云端检索计算服务，实现对不同保险公司电子保单的存储、管理、检索、计算功能。基于全同态加密技术的数据检索计算方法可以直接对加密后的数据进行基本计算而不用将其解密成明文形式，既保证了用户数据的安全性，又保证了服务器能够对存储的数据密文直接进行操作，平衡了云存储系统中保证用户数据的安全性和服务于云计算之间的关系，而利用SM4加密算法结合使用也有效弥补了全同态加密算法加密速率不高的缺陷。同时，为保护信息检索时的数据安全，并提高检索速度，我们将保存在云平台上文件的关键词提取出来，实现基于关键词的密文检索方案。

本次作品设计实现的主要功能和特色是：基于全同态加密（CKKS）的安全存储与隐私计算，基于SM4对称加密的非计算数据安全存储，基于关键词的密文检索。

* 1. **基于全同态加密（CKKS）的安全存储与隐私计算**

为防止用户隐私泄漏和第三方恶意攻击云平台等异常现象，需要对用户的信息加以保护，因此常常选择在客户端将所有数据加密后再存储到云平台上，但由于加密后的数据失去了原有的意义，不能直接在云平台上进行计算和统计分析，这样在很大程度上限制了云平台的统计和计算能力。

因此，对于需要进行计算的数据，我们采用全同态加密算法——CKKS算法进行加密，这种算法可以在不暴露加密数据的情况下，支持对加密数据进行加法和乘法运算。CKKS算法是基于多项式环的全同态加密算法，它使用了多项式的离散傅里叶变换（Discrete Fourier Transform，DFT）和逆离散傅里叶变换（Inverse Discrete Fourier Transform，IDFT）技术，将多项式环的加法和乘法运算映射到离散傅里叶域上，从而实现对加密数据的加法和乘法运算。与其他全同态加密算法相比，CKKS算法具有高效性和可扩展性的优点，它可以在多项式环的上下文中进行计算，而不需要将加密数据转换为比特数组，从而降低了计算和通信的开销。此外，CKKS算法还支持基于概率的纠错技术，可以在一定程度上减少计算错误对加密结果的影响，适用于处理大规模数据和高维数据的加密计算任务。

在云医疗保险平台中，用户上传的数据将会被分为可计算数据和非计算数据分别加密存储。其中需要计算的数据采用全同态加密算法CKKS进行加密后保存在云服务器上，并且在需要的时候直接通过密文计算返回结果。

* 1. **基于SM4对称加密的非计算数据安全存储**

全同态加密算法能够实现明文与密文的加乘同态关系，这使得它适用于保密计算，但其加解密速度相对较慢。对于不需要进行计算的数据，我们采用加解密速度更快的SM4加密算法进行加解密，以提高检索效率。

SM4加密算法是国家密码管理局发布的一种对称加密算法，算法采用分组加密的方式，将明文分成128位一组，经过多轮迭代，得到密文输出。SM4算法采用了多轮迭代的结构，并加入了S盒、置换、线性变换等多种运算，增强了算法的混沌性和抗攻击能力，算法安全性高。相比全同态加密算法，SM4计算速度更快，能高效满足对大规模数据的加密需求。

在云医疗保险平台中，用户上传的数据将会被分为可计算数据和非计算数据分别加密存储。其中不需要计算的数据采用对称加密算法SM4进行加密后保存在云服务器上。

* 1. **基于关键词的密文检索**

信息检索是从有一组的信息集合中根据需要找出相关信息的过程和技术，是

人们有效利用和开发各种信息资源的有效途径。通常信息组织、检索及其它管理

过程在存储容量大、运算速度快的高性能服务器上进行。如果用户将数据以明文

直接上传到服务器上对其进行检索，存在许多隐私信息泄露的安全威胁。数据在

不安全的通信信道上传输被黑客截获，数据在不可信的服务器上存储可能被特权

用户利用。同时，在检索过程中若攻击者获取检索内容，将暴露用户个人习惯、兴趣等隐私。

因此，为保护信息检索时的数据安全，并提高检索速度，我们将保存在云平台上文件的关键词提取出来，并通过全同态加密算法和SM4对称加密算法对其加密，实现基于关键词的密文检索方案，并且可以通过对指定关键词的操作实现云平台的计算能力。

在云医疗保险平台中，当用户需要检索相关文档时，首先将相关关键词加密，并把加密后的关键词上传至云服务器中进行检索。云服务器通过计算检索请求的加密关键词索引与云中存储的关键词索引的相似度，将匹配结果返回给用户。最后用户根据自己需求下载对应的密文文件，并解密获得明文文件。只有数据所有者能使用密钥解密密文，其他人都无法获得明文。

1. **作品设计**
   1. **系统架构**

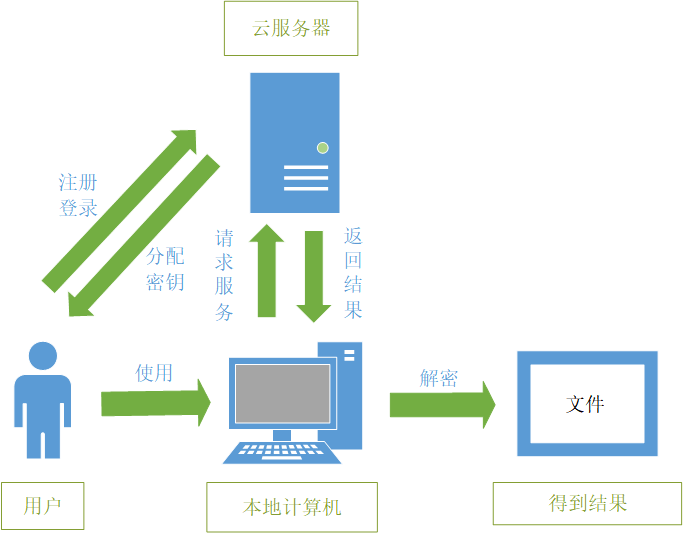


图1 系统架构图

（1）用户

用户初次使用本系统时，需要先进行注册，然后由系统分配全同态加密密钥和SM4加密密钥，并且云平台为用户分配云空间。之后登录时，系统就使用该用户专属的密钥进行加解密操作。

（2）本地计算机

用户使用本地计算机登录web网页，完成与云服务器的交互。用户可以使用文件加密上传，加密文件下载，加密数据云计算等功能。

（3）云服务器

集存储、管理、计算为一体的云端服务器。既可以安全地为每个用户存储加密后的隐私数据，也可以完成对数据的备份与管理，在收到用户请求时可迅速完成对密文的检索和计算功能，由于云服务器总机的运算能力远超过个人电脑，因此在云服务器上直接完成计算可节省大量时间。

* 1. **电子保单加密结构设计**

图2 电子保单加密结构

1. 对于投保险种数量、投保项目数量、保险费、保险时间等可能需要计算的数据，我们选用全同态加密算法对其加密后存储，如：计算分析所有投保人的平均投保时间、计算总保险费用等。
2. 对于保险单号码、投保人、投保人证件号码、联系电话等不需要进行计算，但需要统计筛选的数据，我们选用SM4对称加密算法进行加密，以提高加密速度。
   1. **基于全同态加密和SM4对称加密的加密方案设计**

### 4.3.1全同态加密(CKKS)

Cheon团队在2017年提出了基于CKKS的全同态加密方案，主要算法流程如图3所示，可以对浮点数进行近似计算，在多个全同态加密算法中效率较高。

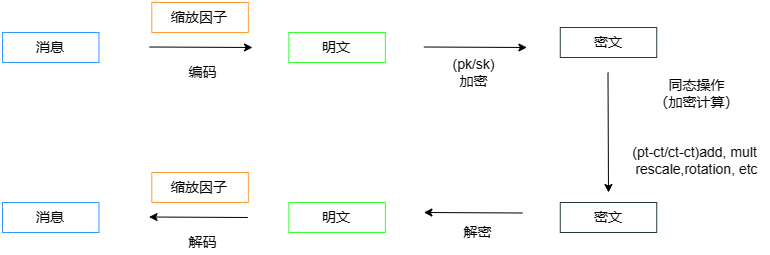


图3 CKKS算法流程

设CKKS方案的安全系数为，明文空间,映射为：

即，具体关系为，其中。方案详细描述如下：

1. **CKKS.KeyGen**

选择一个2的方幂的整数，一个整数，一个大整数，和一个实数;

采样，，，设私钥，公钥，其中;

采样，，设同态乘密钥，其中.

1. **CKKS.Encode**

对于明文向量，首先对其等比放大;

接着通过映射的逆将转化为多项式

1. **CKKS.Decode**

对于明文多项式，使用映射还原为向量并等比缩小

1. **CKKS.**

采样，，输出密文

1. **CKKS.**

对于，解密算法为，即

由于CKKS方案在加密时引入了噪声，所以其解密函数生成的明文与原始明文是不同的，但误差的数量级是远远小于明文的，所以该误差是完全可以忽略的。

1. **CKKS.**

对于，，密文的同态加法为相应位直接相加

由于密文的同态乘法会导致密文规模扩大，这将影响到解密算法的执行，从而使得密钥规模随乘法深度的增加而增大。故Cheon等人在方案设计时引入同态乘密钥实现对乘法密文的缩放，使其规模保持在中。

1. **CKKS.**

对于，，我们记

则的同态乘法表示为

但乘法操作会导致噪声增大，从而出现无法正确解密的情况，这时就需要在每次乘法之后进行一个“重缩放”的操作。

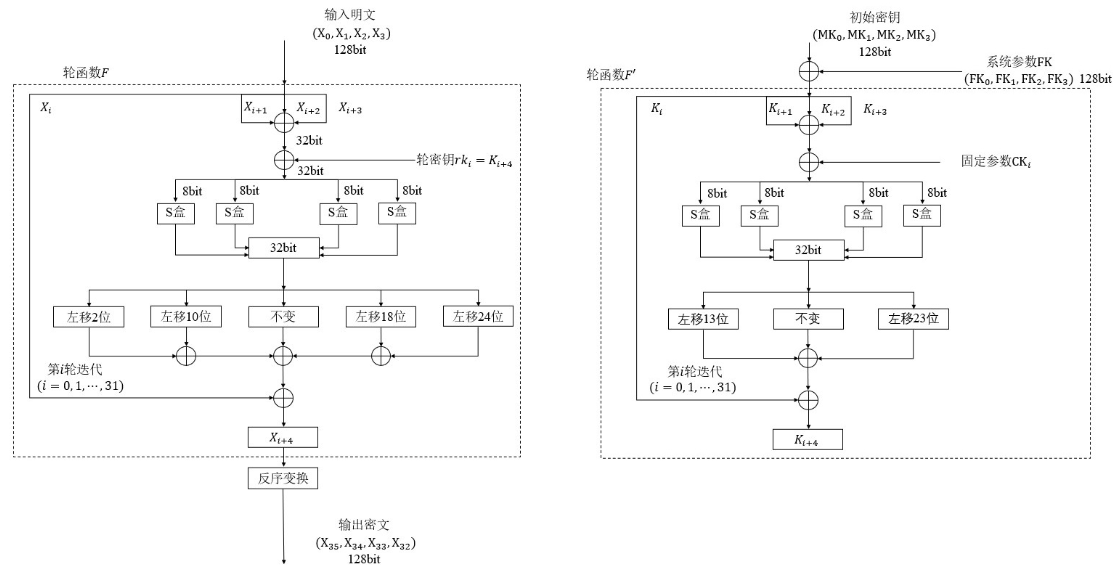
1. **CKKS.**

对于密文，我们对其进行“重缩放”

经“重缩放”后，其密文模数降低为

### 4.3.2 SM4对称加密

SM4算法是一种对称加密算法，也是中国商用密码算法。SM4算法采用分组密码方式，密钥长度为128位，分组长度为128位。SM4算法采用Feistel结构，将明文分成左右两部分进行加密，加密过程中采用了轮函数、S盒、置换等操作。SM4算法的密钥扩展采用了基于非线性的置换和置换线性变换的方法，具有较高的安全性。SM4算法是目前国际上公认的加密算法，已被ISO/IEC、IEC、GB/T等标准机构采用为标准算法。SM4算法具有高效、安全、可控等特点，适用于各种信息安全领域，例如数据加密、数据传输、电子认证等。在中国，SM4算法已被广泛应用于各种信息安全领域，例如物联网、金融、电子政务等。

图4 SM4 的加解密（左）和密钥拓展（右）流程图

SM4算法的加密流程如下：

（1）密钥扩展：将128位密钥扩展为32个字（4字节/字），得到32个子密钥。

（2）初始置换：将明文分成左右两部分，对每一部分进行初始置换。

（3）轮函数：对左右两部分进行16轮加密，每轮加密包括以下步骤：

a. 右部与当前子密钥进行异或运算。

b. 将异或结果输入S盒，得到新的右部。

c. 对新的右部进行置换。

d. 将置换结果与左部进行异或运算，得到新的左部。

（4）逆初始置换：将左右两部分进行逆初始置换，得到密文。

SM4算法的解密流程与加密流程类似，只是子密钥的使用顺序相反。具体来说，解密流程如下：

（1）密钥扩展：将128位密钥扩展为32个字，得到32个子密钥。

（2）初始置换：将密文分成左右两部分，对每一部分进行初始置换。

（3）轮函数：对左右两部分进行16轮解密，每轮解密包括以下步骤：

a. 左部与当前子密钥进行异或运算。

b. 将异或结果输入S盒，得到新的左部。

c. 对新的左部进行置换。

d. 将置换结果与右部进行异或运算，得到新的右部。

（4）逆初始置换：将左右两部分进行逆初始置换，得到明文。

* 1. **基于关键词的密文检索方案设计**

在云保险服务平台中，如果要对加密后的保单进行搜索，需要使用关键词进行筛选，以检索出原始保单。全同态加密算法能够实现明文与密文的加乘同态关系，这使得它适用于保密计算，但其加解密速度相对较慢。与之相比，SM4加密算法具有更快的加解密速度，因此可以提高检索效率。因此，我们将结合使用这两种算法。对于用于保密计算的信息，我们将其选为关键词，并使用全同态加密算法进行加密；对于其他信息，我们将使用SM4加密算法进行加密。这样一来，我们可以将加密后的关键词提取出来作为检索摘要，并在数据库中单独进行存储。用户在本地对数据加密上传到服务器，根据需要提出检索请求，服务器在对密文数据不解密的情况下进行检索，并返回用户检索结果。

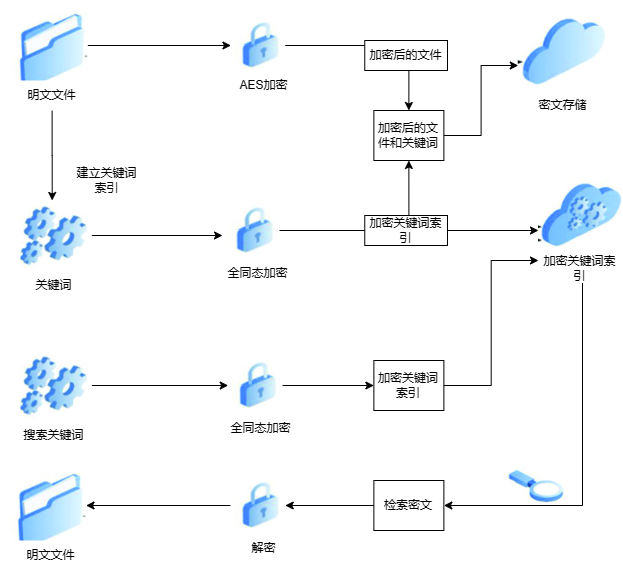


图5 基于关键词的密文检索方案设计图

**参考文献**

[1]刘家森,王绪安,王涵,赵凯洋,闫纪宁.云服务器中基于同态加密的关键词检索方案[J].科学技术与工程,2021,21(08):3180-3185.  
[2]李帅. 基于同态加密技术的云安全存储模型研究[D].中国矿业大学,2015.  
[3]王婧琳. 基于同态加密的金融数据安全共享方案研究及实现[D].哈尔滨工业大学,2020.DOI:10.27061/d.cnki.ghgdu.2020.003535.

[4] Cheon, J.H., Kim, A., Kim, M., Song, Y. (2017). Homomorphic Encryption for Arithmetic of Approximate Numbers. In: Takagi, T., Peyrin, T. (eds) Advances in Cryptology – ASIACRYPT 2017. ASIACRYPT 2017. Lecture Notes in Computer Science(), vol 10624. Springer, Cham. https://doi.org/10.1007/978-3-319-70694-8\_15