# 基于云计算的气象数据信息安全技术研究

马 欣

(61741 部队,北京 100097)

**摘 要**：提出一种有效提高加密云数据搜索效率的方法，通过制定搜索相关文件的标准并返回匹配文件，可以部分实现云计算的数据安全托管服务。相关实验通过与可搜索对称加密技术进行对比和分析，说明该方法具有更好的高效性和鲁棒性。

**关键词**：气象业务系统；云计算；加密数据保护；可搜索对称加密技术；搜索效率

**中图分类号**：TP301 **文献标识码**：A **文章编号**：1674-7720(2012)23-0064-03

## Meteorological data information security technology research based on cloud-computing

Ma Xin

(61741 People's Liberation Army, Beijing 100097, China)

**Abstract**: This paper presents an effective method to improve the searching efficiency encrypted cloud data. Data security manage services can be partly fulfilled through developing the cloud computing criteria for searching relevant documents and returning the matching files. The related experiments show the new method has better efficiency and robustness compared with searchable symmetric encryption technology.

**Key words**: meteorological operational systems; cloud computing; encryption-based data protection; searchable symmetric encryption technology; searching efficiency

随着气象业务的数字化和网络化程度不断发展，并行计算、分布式计算、网络存储等新技术方法已经得到广泛利用。但是主流的气象业务仍然基于传统的数据库管理系统，随着数据的增长需要购置大量的存储空间和处理设备作为支撑。相应的各级气象部门都需要对其进行维护管理，为此投入大量的人力物力。

云数据库是一种由大量计算节点构成的并行处理方式，这种方法突破了传统方式的瓶颈，解决了硬件空间有限和软件费用高昂的问题。云计算必将成为未来气象业务数据管理的最佳解决方案。但是由于气象业务数据的特殊性，导致在使用云计算带来便利的同时，也面临着数据泄露等风险，基于云计算的查询内容及数据的保护将成为利用云计算开展气象业务研究的关键问题之一[1]。

基于云计算的数据保护方法通常采用分发前进行加密，在用户端进行查询时发送匹配的数据文件的方法，如图1所示。数据拥有者通过对原始文件进行索引和加密后分发到云计算的服务器上；用户端则通过终端对服务器上的索引进行检索，服务器对其返回匹配的文件加密，如图2所示[2]。这样的方式在数据文件过大时可能会产生分发困难；此外，更多的用户只希望检索某些特定数据文件，一般通过关键字进行选择性搜索，但是这种纯文本搜索方法无法对云服务器上的加密云数据进行有效搜索[3-4]。

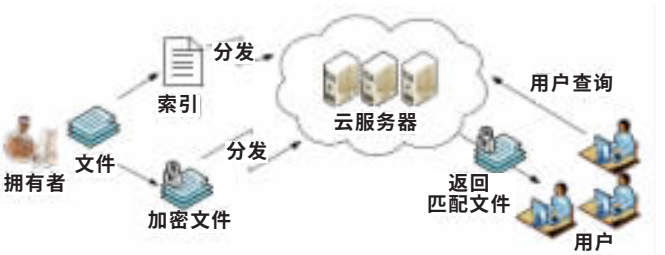



图1 基于云服务器的加密文件查询方法

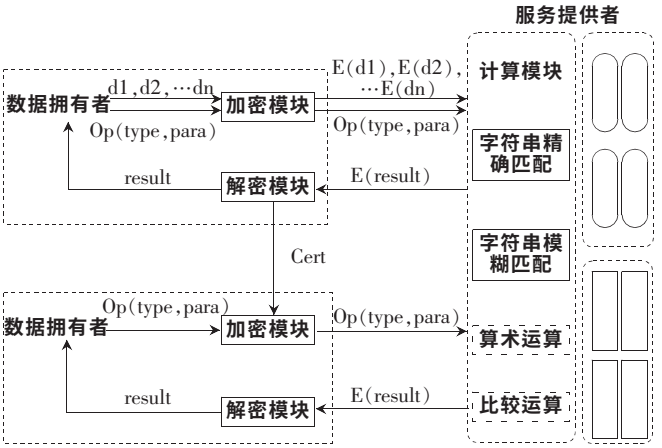



图 2 支持隐私保护的云计算模型

本文提出一种有效提高加密云数据搜索效率的方法，制定了一种搜索相关文件的标准，这种标准可以在按照关键字进行排名搜索的同时保护云数据不被泄露，部分实现云计算的数据安全托管服务。通过实验及其结果与可搜索对称加密技术进行对比和分析，可以认为本文提出的新方法具有更好的高效性和鲁棒性。

## 1 云计算的概念及其在气象业务中的应用

云计算是在处理器技术、分布式存储技术、并行处理和宽带互联网技术高速发展的环境下产生的以基础架构共享来实现的一种新型技术[5-6]。其技术核心为超大规模的分布式环境下的数据存储和网络服务，通过分布式的大规模集群和服务器虚拟化软件搭建来实现。每个“云”包括几十万台、甚至上百万台电脑，“云”中的资源可以无限扩展，并且可以随时获取。基于气象业务的云计算使客户通过远程存储云数据实现气象数据的资源共享，并通过配置获得高质量的气象业务应用和服务。通过这种新的计算模式可以在如下方面对气象业务获得改进和提高[7-9]：

(1)数据存储和采集

通常针对不同气象业务和数据需要提供多台不同的专用服务器，但在运行过程中每台服务器并非同时进行数据存储和处理，造成许多资源的闲置和浪费；通过建立气象业务系统的私有云可以实现云数据和资源在各级气象部门内部及与总部的共享访问。

(2)远程数据查询和访问

通过云计算的方式可以实现对存放在工作场所的气象数据的远程访问，用户只需通过认证就可以对气象数据进行不同权限的处理。

(3)气象数据预报

通过云计算可以整合分布在全国气象部门网络中的数据资源和计算资源等，提升整体系统的计算能力和使用效率，从而获得更加精确的气象预报数据。

## 2 一种改进的可搜索对称加密方案

云计算应用于气象数据业务时的关键问题在于：用户在此条件下对数据不进行直接操作和控制，传统的数据加密算法不能被直接采用。而针对存储数据进行数据安全检查时，也无法掌控全局数据。此外还特别需要针对敏感气象数据进行研究，在保持更新频率的情况下保证数据的安全性和准确性[10]。

传统的基于云计算的数据保护方法主要通过分发前对数据加密，这种加密方案支持对加密数据通过关键字进行搜索，其问题在于，这种技术只支持布尔搜索，没有捕获任何数据文件的相关性。

### 2.1 模型定义

云数据托管服务涉及三个不同的主体：拥有者 O、用户 U、云服务器 C。数据拥有者具有 $n$ 个数据文件 $F=(F_1,F_2,\cdots,F_n)$，拥有者希望对这些文件进行加密分发的同时仍可对其进行有效的数据查询。为此首先需要根据 $m$ 个不同的关键字 $W=(w_1,w_2,\cdots,w_m)$ 建立并存储一个安全的可查询索引 $I$。当需要对关键字 $w$ 进行查询时，授权用户向云服务器 C 提交一个搜索请求，云服务器接收到请求后对其进行响应并返回相应的文件内容。

### 2.2 可搜索对称加密方案

可搜索对称加密 SSE(Searchable Symmetric Encryption)技术允许数据的拥有者以加密方式分发数据，同时保持对加密数据的搜索能力。可搜索对称加密技术的基本策略包括初始化阶段和搜索阶段。

初始化阶段：

定义安全参数 $k,l,l',p$，由数据拥有者对参数进行调用，并初始化产生随机关键字 $x,y\underline{R}\{0,1\}^k$ 和 $z\underline{R}\{0,1\}^l$，输出 $k=\{x,y,z,l^p,l^{p'},l^p\}$，此时对服务器数据建立安全逆序索引，方法如下：

对数据扫描并从中提取不同的关键字 $W=(w_1,w_2,\cdots,w_m)$ 对其建立函数 $F(w_i)$。计算文件 $F_{ij}$ 的得分 $\text{Score}=\frac{1}{|F_d|}\cdot(1+\ln f_{d,t})$，其中 $f_{d,t}$ 代表文件 $F_d$ 中关键字 $t$ 的出现频率。

定义算法语义 $\varepsilon:\{0,1\}^l\cdot\{0,1\}^r\rightarrow\{0,1\}^r$，计算 $\varepsilon_z(\text{Score}_{ij})$ 并与文件 $F_{i,j}$ 的 ID 一起保存。

对于每个索引 $I(w_i)$，假设 $\pi:\{0,1\}^k\cdot\{0,1\}^*\rightarrow\{0,1\}^p$，根据不同的关键字函数 $f(w_i)$ 对包含关键字 $w_i$ 的文件个数 $N_i=F(w_i)$ 进行加密，并用 $\pi_x(w_i)$ 取代 $w_i$，最后输出索引 $I$。

搜索阶段：

对于感兴趣的关键字 $w$ 定义 $T=(\pi_x(w),f(w))$，通过 $\pi_x(w)$ 定位与索引匹配的列表并用 $f(w)$ 解密，并将与 $F(w)$ 相关的文件及其关联的已加密相关得分一起发送；通过关键字对相关得分解密，从而得到排名搜索结果。

上述方法能够很好地满足可搜索对称加密算法对安全保证的要求，但由于排名在用户端完成，需要付出很大的计算代价并进行后处理，发送全部搜索结果也会占用很大带宽，并且这种方法使得服务器可以保留搜索

到的文件和关键字之间的信息。

### 2.3 改进的可搜索对称加密方案

为了使得服务器在没有先验知识的情况下快速获得搜索排名，提出一种改进的可搜索对称加密方案，其基本策略如下：

初始化阶段：

定义安全参数 $k,l,l',p,|D|,|R|$，由数据拥有者对参数进行调用，并初始化产生随机关键字 $x,y,z\underline{R}\{0,1\}^k$，输出 $k=\{x,y,z,l^l,l^{l'},l^p,|D|,|R|\}$，此时对服务器数据建立安全逆序索引，方法如下：

对数据扫描并从中提取不同的关键字 $W=(w_1,w_2,\cdots,w_m)$ 对其建立函数 $F(w_i)$。计算文件 $F_{ij}$ 的得分 $\text{Score}=\frac{1}{|F_d|}\cdot(1+\ln f_{d,t})$，其中 $f_{d,t}$ 代表文件 $F_d$ 中关键字 $t$ 的出现频率。

定义算法语义 $\varepsilon_{new}:\{0,1\}^l\times\{0,1\}^{\log|D|}\rightarrow\{0,1\}^{\log|R|}$，其实现方法为：

$D\leftarrow\{\min(D),\cdots,\min(D)-1+RHGD\}$，其中 $RHGD$ 为随机 $HGD$ 采样函数，且 $R\leftarrow\{\min(R),\cdots,\min(R)-1+|R|/2\}$；得到 $coin\underset{\leftarrow}{R}RCG(k,D,R,m,ID(F_{i,j}))$，则有 $c\underset{\leftarrow}{coin}R$，最后输出 $C$ 作为对分数的加密结果。

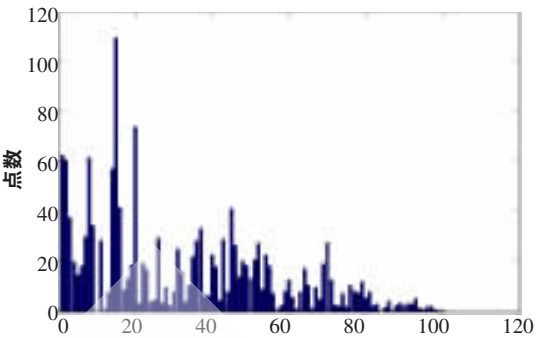
搜索阶段：

首先对于感兴趣的关键字 $w$ 定义 $T=(\pi_x(w),f(w))$，云服务器通过 $\pi_x(w)$ 定位与索引匹配的列表，云服务器获得文件 $F_{i,j}$ 的 ID 及其关联的已加密相关得分 $\text{Score}_{ij}(\varepsilon_{new}(\text{Score}_{ij}))$ 发送；对已加密相关得分解密，从而得到排名搜索结果。

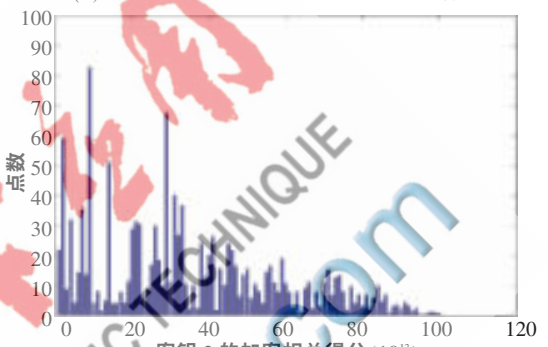
## 3 新方案性能评估

本节对新方案的安全保障性能进行分析和评估，在此过程中需要保证云服务器没有对数据或搜索关键字进行关联学习。实验部署在自行研发的气象数据内部实验平台上，该平台为基于内网的面向单位全体人员的云计算环境，部署在 20 台服务器上。从理论上而言根据上一节提到的新方案可以获得更加随机分布的加密值，可以降低被截获后解密的概率。图 3(a)、3(b)为新方案对同一关键字“Temperature”采用不同的随机密钥得到的加密相关得分对比，由图中可以看出，新方案使得不同的随机密钥对应的加密相关得分的分布产生明显变化，进一步降低加密数据被截获后解密的可能性。

根据本文第 2 节的内容可知，新方案的时间效率与相关分数域范围参数 $M$ 和加密得分范围参数 $R$ 有关，图 4 给出新方案下 50 次试验的时间效率测量平均结果，由图 4 可以看出，随着相关分数域范围参数 $M$ 的增加，所付出的时间也有所增长；同时随着加密得分范围参数 $R$ 的增长，所付出的时间代价减少；包括随机密钥产生在内的时间代价范围在 0.05 s~0.45 s 之间，说明本方案具有效率较高的特性。




(a)对同一关键字采用密钥 1 的加密相关得分




(b)对同一关键字采用密钥 2 的加密相关得分

图 3 对同一关键字采用不同密钥的加密相关得分

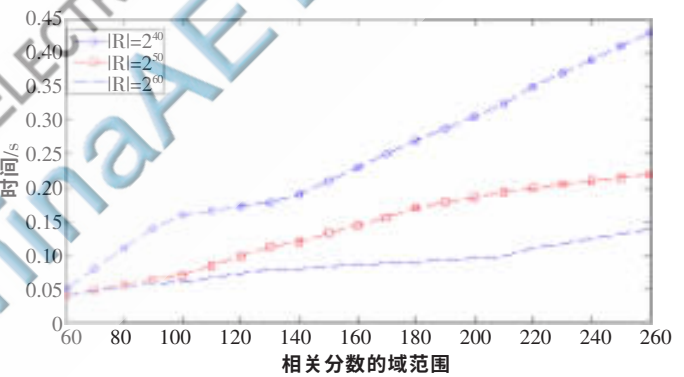



图 4 时间代价与相关分数域范围参数和加密得分范围参数的关系图

本文提出了一种改进加密云数据搜索效率的方法，通过制定搜索相关文件的标准并返回匹配文件，可以部分实现云计算的数据安全托管服务。相关实验通过与可搜索对称加密技术进行对比和分析，说明新方法具有更好的高效性和鲁棒性。下一步将改进该方案中运算速度，使之支持更大规模服务器集群的云计算环境，同时进一步研究加密数据排序问题。

参考文献

[1] 张洁.云计算的发展前景以及安全问题[J].信息与电脑，2012(1)：25-26.

[2] 黄汝维，桂小林，余思，等.云环境中支持隐私保护的可计算加密方法 [J]. 计算机学报，2011，34 (12)：2391-2402.

[3] Chang Y C, MITZENMACHER M. Privacy preserving

keyword searches on remote encrypted data [C]. in Proc. of ACNS'05, 2005.

[4] CURTMOLA R, GARAY J A, KAMARA S, et al. Searchable symmetric encryption: improved definitions and efficient constructions[C]. in Proc. of ACM CCS'06, 2006.

[5] GOLDREICH O, OSTROVSKY R. Software protection and simulation on oblivious rams[J]. Journal of the ACM, 1996, 43(3):431 - 473.

[6] 刘鹏.云计算[M].北京:电子工业出版社,2010.

[7] KAMARA S, LAUTER K. Cryptographic cloud storage[C]. Proceedings of Financial Cryptography: Workshop on Real-Life Cryptographic Protocols and Standardization 2010, January 2010.

[8] REN K, LOU W, KIM K, et al. A novel privacy preserving authentication and access control scheme for pervasive computing environment [J]. IEEE Transactions on Vehicular Technology, 2006,55(4):1373-1384.

[9] BONEH D, WATERS B. Conjunctive, subset, and range queries on encrypted data. Proceedings of TCC 2007[C]. Lecture Notes in Computer Science 4392, 2007.

[10] SHI E, BETHENCOURT J, CHAN T-H. H, et al. Multi-dimensional range query over encrypted data[C]. Proceedings of IEEE Symposium on Security and Privacy, 2007.

(收稿日期:2012-07-24)

作者简介:

马欣,男,1976年生,博士,主要研究方向:信息安全和云计算。