Bluce：一个高效的支持排序的关键词可搜索加密系统

福建师范大学　　张楠　陈兰香

I. 基本介绍

- 实现一个高效的支持排序的关键词可搜索加密系统
- 引用了一种改进的概率模型BM25L，改进了Lucene的评分模型
- Bluce系统精度更高，相关性更好，排序更加合理

II. BM25L模型介绍

$$S_c(s_i, d_i) = \sum_t^N IDF(s_i)\left(\frac{(g_1+1)\cdot(e_{td}+\delta)}{g_1+(e_{td}+\delta)}\right)$$

其中，

$$IDF(s_i) = \log\frac{N - n(s_i) + 0.5}{n(s_i) + 0.5}$$

$$e_{td} = \frac{f(s_i, d_i)}{1 - b + b\cdot\left(\frac{dl}{avgdl}\right)}$$

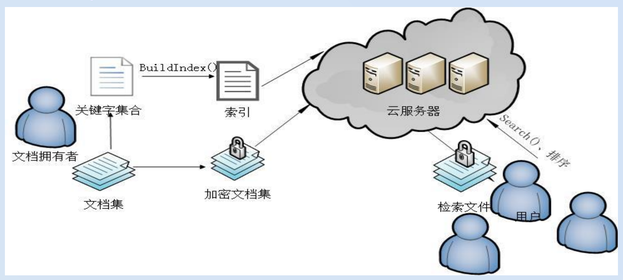
- N是集合中文档的总数
- $f(s_i, d_i)$是检索词s_i在文档d_i中的频率
- dl是文档d_i的长度
- $avgdl$是所有文档的平均长度
- g_1和b是自由参数
- $IDF(s_i)$是检索词s_i的IDF（文档频率倒数)权重
- δ是一个正数，本文取0.5

III. 方案描述

可以将方案形式化地表示为一个五元算法（*Keygen*，*BulidIndex*，*Trapdoor*，*Search*，*Decrypt*）简单描述如下：

- $K \leftarrow Keygen(1^k)$: 由用户执行的生成密钥的概率性算法，以安全参数k为输入，输出密钥$K=(k_1, k_2, k_3)$。
- $I \leftarrow Buildlndex\ (K, D)$: 由用户执行的创建安全索引算法，输入密钥$K$以及文件集合D，输出一个安全的密文索引$\Delta$。
- $Tw \leftarrow Trapdoor(K, w)$: 由客户端执行的陷门生成算法，输入密钥$K$和关键字$w$，输出陷门$Tw$。
- $IDw' \leftarrow Search(\Delta, Tw)$: 由服务器执行的密文检索算法，输入安全索引$\Delta$和陷门$Tw$，输出包含关键字w的密态标识符集合$IDw'$。
- $Dw \leftarrow Decrypt(K, IDw')$: 由用户执行的解密算法，输入对称密钥$K$和密文标识符集合$IDw'$，输出包含检索关键字$w$的明文文件集$Dw$。

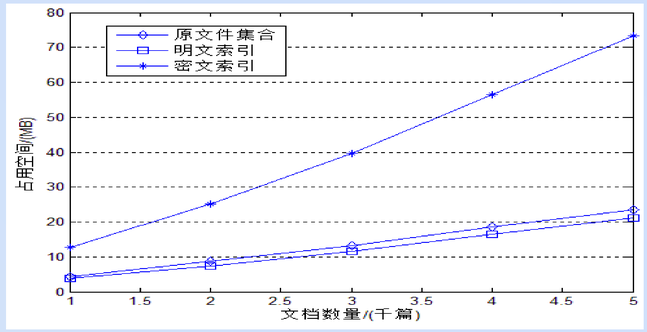
IV. 系统框架



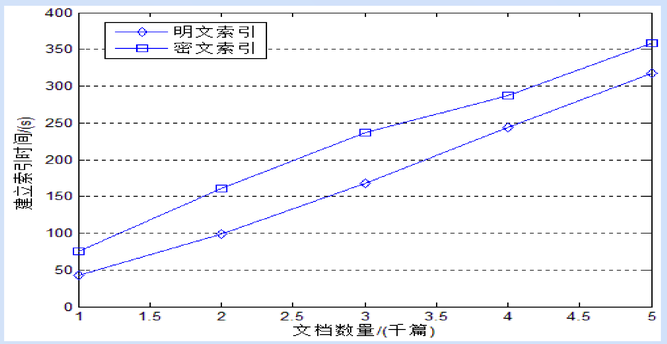
- 三大部分：文档拥有者、服务器、用户
- 两大核心处理部分：密文索引构建、密文检索

V. 性能分析

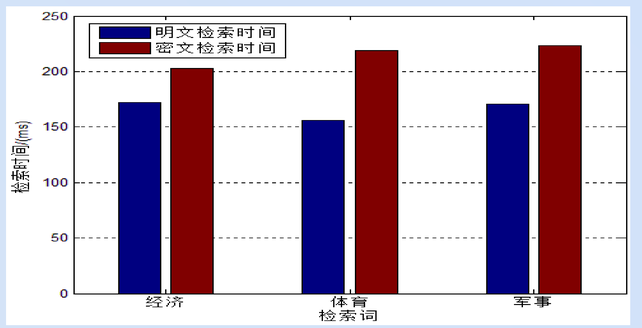
- 占用空间



- 索引构建时间



- 检索时间



- 查询性能

方法	军事		体育		经济	
	MAP	P@15	MAP	P@15	MAP	P@15
Lucene	0.3713	0.6672	0.2012	0.3422	0.4718	0.6931
Bluce-BM25	0.5536	0.7394	0.2312	0.4360	0.5618	0.7654
Bluce-BM25L	0.5822	0.7959	0.2346	0.4420	0.5841	0.7984

VI. 结论

BM25L模型具有更好的检索性能，可以在保证密文检索安全性的基础上提高密文检索的效率和查准率。

VII. 不足

密文索引占用的空间过大，需要在以后的工作中解决这个问题。