基于云计算平台的物联网加密数据比较方案

孟倩[1]，马建峰[2,3]，陈克非[4]，苗银宾[3]，杨腾飞[3]

（1. 西安电子科技大学通信工程学院，陕西 西安 710071；
2. 西安电子科技大学计算机学院，陕西 西安 710071；
3. 西安电子科技大学网络与信息安全学院，陕西 西安 710071；
4. 杭州师范大学理学院，浙江 杭州 310036）

摘 要：已有的短比较加密（SCE, short comparable encryption）方案能在确保物联网数据安全的前提下通过比较密文数据大小而推出明文数据大小。但 SCE 方案在密文比较以及生成标签的过程中会引入大量的计算和存储开销。为此，提出一种基于滑动窗口技术统一开窗的高效短比较加密（SCESW, short comparable encryption based on sliding window）方案。严格的安全分析表明，SCESW 方案在标准模型下满足弱不可区分性且保障了数据的完整性和机密性。同时，实验性能分析表明，SCESW 方案的存储开销是 SCE 方案的 $\frac{1}{t}(t>1)$ 且效率高于 SCE 方案。

关键词：短比较加密；滑动窗口；标准模型；弱不可区分性；完整性；机密性
中图分类号：TP309
文献标识码：A
doi: 10.11959/j.issn.1000-436x.2018065

Data comparable encryption scheme based on cloud computing in Internet of things

MENG Qian[1], MA Jianfeng[2,3], CHEN Kefei[4], MIAO Yinbin[3], YANG Tengfei[3]

1. School of Telecommunication Engineering, Xidian University, Xi'an 710071, China
2. School of Computer Science and Technology, Xidian University, Xi'an 710071, China
3. School of Cyber Engineering, Xidian University, Xi'an 710071, China
4. School of Science, Hangzhou Normal University, Hangzhou 310036, China

Abstract: The previously proposed short comparable encryption (SCE) scheme can infer the plaintext relationship by comparing the ciphertexts relationship as well as ensuring data security in Internet of things. Unfortunately, it will incur high storage and computational burden during the process of comparing ciphertexts and generating tokens. To this end, an efficient short comparable encryption scheme called SCESW was proposed, which was utilizing the sliding window method with the same size window. Formal security analysis shows that the scheme can guarantee weak indistinguishability in standard model as well as data security and integrity. The experimental results demonstrate that the storage of the SCESW scheme is $\frac{1}{t}$ $(t>1)$ times shorter than that of the SCE scheme and the efficiency of the SCESW scheme is superior to that of the SCE scheme.

Key words: short comparable encryption, sliding window, standard model, weak indistinguishability, integrity, confidentiality

收稿日期：2017-10-12；**修回日期**：2018-03-20
基金项目：国家高技术研究发展计划（“863”计划）基金资助项目（No.2015AA016007）；国家自然科学基金资助项目（No. 61702404）
Foundation Items: The National High Technology Research and Development Program of China (863 Program) (No.2015AA016007), The National Natural Science Foundation of China (No. 61702404)

1 引言

近年来，人们正逐步进入一个万物互联的物联网时代[1]，更多的智能终端需要接入互联网。云租户从物联网获取的海量数据通过云计算平台进行智能处理。敏感数据（如患者医疗记录）通常被外包到云服务器进而减少本地数据存储和管理开销。由于云服务器并不完全可信，因此，云租户在享受便捷的数据服务同时也会面临敏感数据泄露的风险。为了保证数据的安全性，通常将数据加密后再传到云服务器。由于在密文条件下对数据的操作比较困难，所以如何确保云租户对密文数据进行比较查询操作是研究的重要方向。

在传统的数据库加密中，国内外科研工作者针对密文数据查询操作做了许多工作。2004 年，Agrawal 等[2]针对数字密文查询提出一种保序加密（OPE）方案，但此方案导致巨大的存储开销和计算开销。为此，2009 年，Agrawal 等[3]又采用分布式存储的思想提出具有隐私保护的数据外包方法。此方案在一定程度上解决了密文的计算开销和存储开销的问题，但是方案的安全性不高。文献[4~7]分别提出安全性增强的 OPE 方案。2010 年，Kadhem 等[4]利用对明文消息进行分割然后添加随机数的思想提出一种多值部分保序加密方案。2012 年，Liu 等[5]通过添加噪声随机化每个索引的思想提出一种检索密文数据库的保序加密方案。2013 年，Popa 等[8]首次利用可变密文思想提出一种更安全的保序加密方案。虽然各行各业的科研工作者已经提出许多 OPE 方案，但是这些方案仍不能解决 OPE 的核心问题，即安全性问题，如图 1 所示。由于在日常生活中使用比较查询操作[9~12]非常普遍，例如，一个医疗机构的所有患者的年龄使用 OPE 进行加密，那么获得患者加密后的信息后就可以推断患者年龄的大小关系。因此，提出一种高安全的可比较加密方案是有必要的。

2013 年，Furukawa[13]采用前缀保护加密（PPE, prefix preserving encryption）思想提出一种基于请求的比较加密方案。该方案在一定程度上增强了 OPE 的安全性，但仍会带来巨大的计算开销和存储开销。为了进一步减少密文存储空间，Furukawa[14]采用 PPE 思想提出一种短比较加密（SCE）方案。对比基于请求的比较加密方案，SCE 方案在存储密文时将密文的二进制转化为三进制存储，从而使密

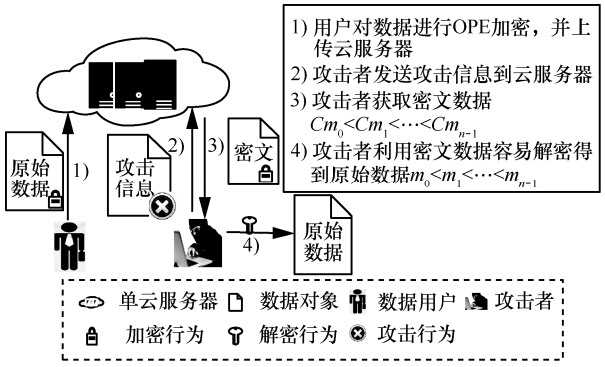

图 1 保序加密的核心问题

文存储长度减少，提高了数据库的工作效率。为了将比较加密方案应用到更多的实际场景中，Zou 等[15]将图像的像素点使用比较加密方案加密后再进行一系列的检索工作，从而提出一种安全有效的图像检索方案。为了将比较加密方案应用到拍卖系统中，Zhu 等[16]将比较加密方案与拍卖系统相结合提出了基于修改的比较加密方案拍卖系统。针对多个竞标者参与的情况，Guo 等[17]将比较加密与多线形对结合进而提出了安全的密封投标方案。上述这些实际场景中探讨的都是比较加密（CE）方案，为了进一步减少计算开销和存储开销，对 SCE 方案进行探讨从而提出 SCESW 方案。为了减少计算开销和存储开销，对滑动窗口技术进行了研究[18,19]。滑动窗口技术可以运用到幂指数运算中[19]，通过滑动窗口技术可以减少幂指数的运算量。Chen 等[18]利用滑动窗口技术提出一种高效的基于请求的比较加密方案。该方案利用开窗的思想将数据的二进制形式进行分组，这在一定程度上减少计算开销和存储开销。试想将滑动窗口技术应用到 SCE 方案中，进而提出一种高效的短比较加密方案，该方案在一定程度上大大减少计算开销和存储开销，从而使方案在实际应用中的效率得到提高。

SCE 方案在数据库的加密中是一种很好的选择方案，但是提出一种理想的 SCE 方案同时兼顾效率在实际应用中变得不可行。相比于理想特性下的加密方案，文献[13,14]中的方案均是考虑弱特性下的比较加密方案，并且这 2 种方案都具有很强的安全性。与理想特性下的加密方案相比，Furukawa 提出的方案中标签 *token* 会泄露部分信息。但是文献[14]中对其安全性进行评估，其结果表明在弱特性下，标签 *token* 泄露的信息很少，同时能保证很强的安全性。因此，提出弱特性下的加

密体制是有意义的。

结合滑动窗口技术和 SCE 方案，本文提出一种 SCESW 方案。由于原有的比较加密方案的存储开销和计算开销比较大，新的方案可以在保证安全的前提下使存储开销和计算开销降低，从而使效率提高。本文采用的滑动窗口法对数不再区分零窗口和非零窗口，而是对其进行统一开窗，使每个窗口大小相等。在存储密文时，SCESW 方案将密文的二进制转化为 $2^{t+1}-1$ 进制存储，从而使密文存储长度变小。这样的思想可以使 SCESW 方案的计算开销和存储开销减少。同时严格的安全分析表明，SCESW 方案在标准模型下能实现弱不可区分性。

2　预备知识

本节介绍滑动窗口法和 SCE 方案以及安全性分析中用到的知识。

2.1　滑动窗口技术

滑动窗口法[19]一般运用到幂指数运算中。例如，求幂运算 x^e 时，通常将 e 表示成二进制形式即 $e=(b_0,b_1,\cdots,b_{n-1});b_i\in\{0,1\}$。一般情况下将整数 e 分割为固定长度的块，然后进行非零块次数的乘法运算。如果使用的分块长度不同，就可以减少非零块从而减少乘法运算总数。这样的分割思想方法称为滑动窗口法[19]。

在实际应用中会对滑动窗口法做一些优化，因为在数的二进制表示中，零位和非零位都是有意义的，因此，不再区分零窗口和非零窗口，而是对数的二进制形式进行统一开窗，使每个窗口大小相等。这种通过减少计算量提高效率的技术受到了各行各业的广泛关注。

2.2　安全模型

本节首先介绍弱可区分游戏，然后再引入 SCE 方案弱特性下的安全模型。此部分是为证明 SCESW 方案满足在标准模型下的弱不可区分性做准备。

挑战者 C 和敌手 A 之间进行弱可区分游戏[14]。首先，挑战者 C 接收到安全参数 $k\in N$ 和范围参数 $n\in N$，然后执行参数生成算法 Gen，即 $Gen(k,n)=(param,mkey)$，并将生成的公共参数 $param$ 返回给敌手 A。敌手 A 向挑战者 C 发起询问，在这个游戏中挑战者 C 对询问的应答如下。

1) 挑战者 C 接收到任何一个询问数字 $0\leqslant num\leqslant 2^n$，然后执行标签生成算法 Der 并返回生成标签 $token=Der(param,mkey,num)$。

2) 挑战者 C 接收到任何一个询问数字 $0\leqslant num\leqslant 2^n$，然后执行加密算法 Enc 并返回密文 $ciph=Enc(param,mkey,num)$。

3) 挑战者 C 接收到一组需要进行询问的数字 $0\leqslant num_0^*<num_1^*<2^n$，挑战者随机地选取 $b\in\{0,1\}$ 然后产生密文 $ciph^*=Enc(param,mkey,num_b^*)$。

在这个游戏中，敌手 A 不允许做以下询问。

$$\exists\ l\ (0<l<n)$$

$$\text{s.t. } ((\alpha_l,\cdots,\alpha_{n-1})=(\beta_l,\cdots,\beta_{n-1}))$$

$$=(\gamma_l,\cdots,\gamma_{n-1})\wedge(\beta_{l-1}<\gamma_{l-1})$$

$$(num=\sum_{i=0}^{n-1}\alpha_i 2^i;\ num_0^*=\sum_{i=0}^{n-1}\beta_i 2^i;\ num_1^*=\sum_{i=0}^{n-1}\gamma_i 2^i,$$

$\alpha_i,\beta_i,\gamma_i\in\{0,1\}$)

游戏结束时，敌手 A 将 $b'\in\{0,1\}$ 发送给 C。游戏的结果为

$$Exp_{C,A}^k=\begin{cases}1, & b=b'\\ 0, & b\neq b'\end{cases}$$

定义 1[14]　在任意的多项式时间内，敌手 A 进行询问后，$Adv_{C,A}^k:=|Pr(Exp_{C,A}^k=0)-Pr(Exp_{C,A}^k=1)|$ 在弱可区分游戏中对于 k 是可忽略的，则称 SCE 方案是弱不可区分的。

3　SCESW 方案

在实际应用中会对滑动窗口法做一些优化，不再区分零窗口和非零窗口，而是对数的二进制形式进行统一开窗，使每个窗口大小相等。数的每个窗口包含 t ($t>1$) bit 的信息量，然后再进行标签生成、加密和密文比较等操作。与 SCE 方案相比，SCESW 方案的计算存储开销较低且效率较高。

3.1　SCESW 方案系统模型

图 2 给出了基于云计算平台的物联网加密数据方案，即 SCESW 方案的系统模型。该系统模型包括 3 个实体，即数据拥有者、云租户和云服务器。数据拥有者负责将数据上传到云服务器之前需要对共享数据进行加密；半可信的云服务器负责数据的存储和检索操作；云租户负责提交查询请求从而得到数据的大小关系。

3.2　SCESW 方案定义

SCESW 方案包含 5 个算法：参数生成 Gen、数据分块 Par、标签生成 Der、密文生成 Enc 和密文比较 Cmp。

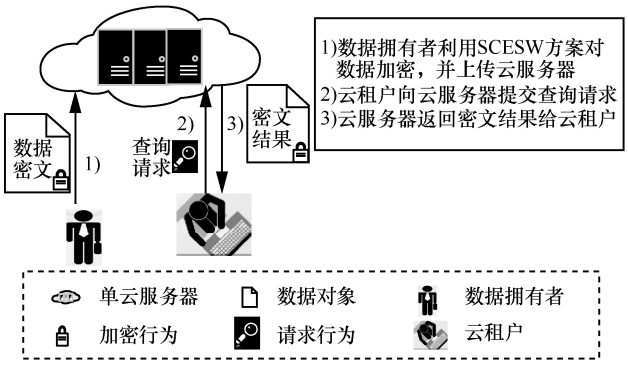

图 2 SCESW 方案的系统模型

1) *Gen* 算法

给定安全参数 $k \in N$ 和范围参数 $n \in N$。算法输出公共参数 *param* 和主密钥 *mkey*，即

$$Gen(k,n)=(param,mkey)$$

2) *Par* 算法

$num=(b_0,b_1,\cdots,b_{n-1});\ b_i \in \{0,1\}$ 是给定数字的二进制表示形式，数开窗后输出值是 $num=(B_0,\cdots,B_{m-1})$；$\frac{n}{m}=t$。

3) *Der* 算法

给定公共参数 *param*、主密钥 *mkey* 和数字 *num*。算法输出标签 *token*，即

$$token=Der(param,mkey,num)$$

4) *Enc* 算法

给定公共参数 *param*、主密钥 *mkey* 和数字 *num*。算法输出密文 *ciph*，即

$$ciph=Enc(param,mkey,num)$$

5) *Cmp* 算法

给定公共参数 *param*，密文 *ciph*、$ciph^*$ 和其中一个密文对应的标签 *token*。算法输出结果为

$$Cmp=\begin{cases}-1, num>num^*\\ 0, num=num^*\\ 1, num<num^*\end{cases}$$

3.3 SCESW 方案具体描述

本节将窗口大小设置为 $t\ (t>1)$ 来描述 SCESW 方案。假定数字的二进制长度为 *n* bit，每个窗口包含 *t* bit 信息，其中，*n* 是 *t* 的倍数。事实上，*n* 可以是任意长度，如果 *n* 不能被 *t* 整除，可以进行补零操作直到它是 *t* 的倍数为止。在具体介绍 SCESW 方案之前，表 1 中给出了 SCESW 方案用到的符号定义。

表 1 符号的定义

符号	定义
H_1	$Hash_1$ 函数
H_2	$Hash_2$ 函数
H_3	$Hash_3$ 函数
H	散列函数
mkey	主密钥
I	*k* 位随机数
n	数字长度
h	散列函数运算个数
m	开窗个数
param	输出参数

SCESW 方案从以下 5 种算法进行描述：参数生成 *Gen*、数据分块 *Par*、标签生成 *Der*、密文生成 *Enc* 和密文比较 *Cmp*。

1) *Gen* 算法

给定安全参数 $k \in N$ 和范围参数 $n \in N$。随机地选择 H_1、H_2、H_3 满足条件 $\{0,1\}^k \times \{0,1\}^* \to \{0,1\}^k$。算法输出公共参数 *param* 和主密钥 *mkey*。其中，$param=(n,H_1,H_2,H_3)$。

2) *Par* 算法

给定数字的二进制表示形式 $num=(b_0,b_1,\cdots,b_{n-1});\ b_i \in \{0,1\}$，算法输出结果是开窗值为 *t* 的分组数据。数开窗后表示为

$$num=(B_0,\cdots,B_{m-1})=\sum_{i=0}^{m-1}B_i(2^t)^i;\ \frac{n}{m}=t \tag{1}$$

其中，$B_0=(b_0,\cdots,b_{t-1}),\cdots,B_{m-1}=(b_{n-t},\cdots,b_{n-1})$。

3) *Der* 算法

给定公共参数 *param*、主密钥 *mkey* 和开窗后的数 *num*。按照式(2)生成标签 d_i，*Der* 算法输出的标签为 $token=(d_1,d_2,\cdots,d_m)$。其中，有

$$d_i=H_1(mkey,B_m,B_{m-1},\cdots,B_i),\ i=1,2,\cdots,m \tag{2}$$

4) *Enc* 算法

给定公共参数 *param*、主密钥 *mkey* 和开窗后的数 *num*。*Enc* 算法中随机地产生 $I \in \{0,1\}^k$ 和标签 $token=(d_1,\cdots,d_m)$，按照式(3)生成 f_i，输出密文 $ciph=(I,(f_0,f_1,\cdots,f_{m-1}))$。为了使密文长度变短，$(f_0,f_1,\cdots,f_{m-1})$ 转化为整数 $F=\sum_{i=0}^{m-1}f_i(2^{(t+1)}-1)^i$ 存储。其中，有

$$f_i=H_1(d_{i+1},I)+H_2(mkey,d_{i+1})+B_i \mod(2^{(t+1)}-1) \tag{3}$$

5) *Cmp* 算法

给定公共参数 *param*，密文 $ciph=(I,(f_0,\cdots,f_{m-1}))$、$ciph^*=(I',(f_0',f_1',\cdots,f_{m-1}'))$ 和其中一个密文对应的标签 *token*。由式(4)比较得到第一个不同的窗口

$$c_j=f_j-f_j'-H_3(d_{j+1},I)+H_3(d_{j+1},I')\ \bmod(2^{(t+1)}-1) \tag{4}$$

其中，$j=m-1,\cdots,1$。

Cmp 算法输出结果为

$$Cmp=\begin{cases}-1(num>num^*),1\leqslant c_j\leqslant 2^t-1\\ 0(num=num^*),\quad c_j\equiv 0\\ 1(num<num^*),2^t\leqslant c_j\leqslant 2^{(t+1)}-2\end{cases}$$

参数生成算法主要用来生成后面步骤中用到的公共参数 *param* 和主密钥 *mkey*。标签生成算法主要用来生成和数字 *num* 相关的标签 *token*，数字 num^* 生成的 $token^*$ 与此过程类似。密文生成算法主要用来生成和数字 *num* 相关的密文 *ciph*，数字 num^* 生成的密文 $ciph^*$ 与此过程类似。密文比较算法主要利用之前生成的密文数据和与其中一个数字相关联的标签进行比较操作，最终判断出一对密文 *ciph* 和 $ciph^*$ 的第一个不同窗口上的差值关系。

4　安全性分析

首先对挑战者 C_A 以及挑战者 C_B 作一下说明。

挑战者 C_A 与文献[14]定义 1 中的挑战者 *C* 除下面几点外其他都相同。1) 游戏开始前，C_A 分配密钥 *mkey*。2) C_A 对 $H_1(mkey,\cdot)$、$H_2(mkey,\cdot)$ 分别组建一个表进行模拟，即对于输入 C_A 模拟输出 $output=H_1(mkey,input)$、$output=H_2(mkey,input)$。若 $(input,output)$ 在对应的表中则令 *output* 为 $output^*$；否则，C_A 随机选取 $output\in\{0,1\}^k$，然后将 $(input,output)$ 添加到表中。

挑战者 C_B 与挑战者 C_A 除下述之外其他都相同。假设数字 num_0^*、num_1^* 生成的对应标签分别是 $(\overline{d_1},\overline{d_2},\cdots,\overline{d_m})=token_0^*$、$(d_1^*,d_2^*,\cdots,d_m^*)=token_1^*$，其中 $num_0^*=\sum_{i=0}^{n-1}\beta_i 2^i=\sum_{i=0}^{m-1}B_i(2^t)^i$、$num_1^*=\sum_{i=0}^{n-1}\gamma_i 2^i=\sum_{i=0}^{m-1}E_i(2^t)^i$，$\frac{n}{m}=t$。此时，$\forall\, l\leqslant i\leqslant m$，有 $\overline{d_i}=d_i^*$，其中，*l* 满足以下关系，即对数字 num_0^*、num_1^* 有 $(B_l,\cdots,B_{m-1})=(E_l,\cdots,E_{m-1})\wedge(B_{l-1}<E_{l-1})$ 成立。C_B 对模拟散列函数 $H_3(\overline{d_i},\cdot)$、$H_3(d_i^*,\cdot)$，$\forall\, 0\leqslant i\leqslant l$ 分别组建一个表，模拟与前面一致。

定理 1　若 H_1、H_2、H_3 函数是伪随机函数，则 SCESW 方案满足弱不可区分性。

证明　假设存在多项式时间，敌手 *A* 进行询问后，$Adv_{C,A}^k:=|Pr(Exp_{C,A}^k=0)-Pr(Exp_{C,A}^k=1)|$ 在弱可区分游戏中对于 *k* 是不可忽略的。因此，可以将其表示为 $|Pr(Exp_{C,A}^k=0)-Pr(Exp_{C,A}^k=1)|\geqslant\varepsilon$，并认为 H_1、H_2、H_3 可以从随机函数中进行区分。这样就与本文的前提假设 H_1、H_2、H_3 是伪随机函数相矛盾。下面考虑挑战者 *C*、C_A、C_B 参与的一系列游戏从而去证明该定理。这里首先给出引理 1~引理 3 的定义以及证明。

引理 1　若 H_1、H_2 函数是伪随机函数，则对于任意的多项式时间，敌手 *A* 进行询问后，$|Adv_{C_A,A}^k-Adv_{C,A}^k|$ 对于 *k* 是可忽略的。

证明　事实上 *mkey* 只是用作 *H* 函数的输入并且敌手 *A* 很难得到 *mkey*。故引理中的伪随机函数满足不可区分性。基于这一事实该引理得到证明。

引理 2　若 H_3 函数是伪随机函数，则对于任意的多项式时间，敌手 *A* 进行询问后，$|Adv_{C_A,A}^k-Adv_{C_B,A}^k|$ 对于 *k* 是可忽略的。

证明　假设敌手 *A* 对以下一个查询数字进行查询 $num=(A_0,A_1,\cdots,A_{m-1})$。如果在定义的弱区分游戏中 $\forall\, l\leqslant i\leqslant m$，有 $B_i=A_i$ 成立，则这样的查询是不允许的。因此，这种情况不予考虑。

考虑以下情况，如果对于 $\forall\, l+1\leqslant i\leqslant m$，有 $B_i=A_i$ 成立，那么这意味着敌手很难获得标签 $(\overline{d_1},\overline{d_2},\cdots,\overline{d_m})$、$(d_1^*,d_2^*,\cdots,d_m^*)$ 的任何信息。因为 $(\overline{d_1},\overline{d_2},\cdots,\overline{d_m})$、$(d_1^*,d_2^*,\cdots,d_m^*)$ 是随机选取的并且不会被泄露，所以 $\forall\, 0\leqslant i\leqslant l$，$H_3(\overline{d_i},\cdot)$、$H_3(d_i^*,\cdot)$ 的输出满足伪随机函数的不可区分性。

引理 3　对于任意的多项式时间，敌手 *A* 进行询问后，$Adv_{C_B,A}^k=0$。

证明　事实上，密文 $ciph^*$ 不依赖于 B_i。数字 num_0^*、num_1^* 对应的密文之间的不同仅仅依赖于 $(f_i)_i,\forall\, i=0,\cdots,l-1$。由于 $H_3(\overline{d_i},\cdot)$ 是随机选取的，故 $\forall\, i=0,\cdots,l-1$，每个 f_i 都不依赖于 B_i。基于这一事实该引理得到证明。

从引理 1 和引理 2 以及前面提到的矛盾假设可知，只要 H_1、H_2、H_3 是伪随机函数，则

$|Adv^k_{C,A}-Adv^k_{C_B,A}|<\varepsilon$。由引理 3 可知 $Adv^k_{C_B,A}=0$，故 $|Adv^k_{C,A}|<\varepsilon$。因此，定理 1 得到证明。

若 H_1、H_2、H_3 函数是伪随机函数，SCESW 方案满足弱不可区分性。除此之外，SCESW 方案还满足完整性。可以这样分析，假设一对需要比较的密文 $ciph$ 和 $ciph^*$，由式(1)知分别是数字 $num=\sum_{i=0}^{n-1}b_i2^i=\sum_{i=0}^{m-1}B_i(2^t)^i$，$\frac{n}{m}=t$ 和数字 $num=\sum_{i=0}^{n-1}\beta_i2^i=\sum_{i=0}^{m-1}B_i'(2^t)^i$，$\frac{n}{m}=t$ 生成的密文。其中，t 表示开窗的窗口大小，m 表示窗口总数。

由式(2)可知，num 和 num^* 生成的标签 $token$ 分别表示为 $token=(d_1,d_2,\cdots,d_m)$、$token^*=(d_1',d_2',\cdots,d_m')$。此外，$num$ 和 num^* 对应生成的密文表示为 $ciph=(I,(f_0,f_1,\cdots,f_{m-1}))$、$ciph^*=(I',(f_0',f_1',\cdots,f_{m-1}'))$。

为了使密文长度变短，$(f_0,f_1,\cdots,f_{m-1})$、$(f_0',f_1',\cdots,f_{m-1}')$ 分别转化为 $F=\sum_{i=0}^{m-1}f_i(2^{(t+1)}-1)^i$、$F^*=\sum_{i=0}^{m-1}f_i'(2^{(t+1)}-1)^i$ 存储。

在这些关系中可以看出，标签 $token$ 中的分量 d_i 和 d_i' 仅与 B_i、B_{i+1}、B_i'、B_{i+1}' 和 $mkey$ 有关。假设 l 为 num 和 num^* 的第一个不同的窗口，那么对于 $i=l+1,\cdots,m-1$，如果 $B_{i+1}=B_{i+1}'$，则有 $d_{i+1}=d_{i+1}'$。

若 $num=num^*$，$\forall i=0,1,\cdots,m-1$，Cmp 输出 0。若 $num\neq num^*$，对于第一个不同的窗口 j，有

$$\begin{aligned}c_j&=f_j-f_j'-H_3(d_{j+1},I)+H_3(d_{j+1},I')\ \bmod(2^{(t+1)}-1)\\&=(f_j-H_3(d_{j+1},I))-(f_j'-H_3(d_{j+1},I'))\ \bmod(2^{(t+1)}-1)\\&=((H_3(d_{j+1},I)+H_2(mkey,d_{j+1})+B_j)-((H_3(d_{j+1}',I)+H_2(mkey,d_{j+1}')+B_j'))\ \bmod(2^{(t+1)}-1)\\&=B_j-B_j'\bmod(2^{(t+1)}-1)\end{aligned}$$

c_j 的值有 $2^{(t+1)}-2$ 种情况，因为 $B_j-B_j'=0$ 时即为 $num=num^*$ 的情况。如果 c_j 满足 $1\leqslant c_j\leqslant 2^t-1$ 时，则 $num>num^*$。如果 c_j 满足 $2^t\leqslant c_j\leqslant 2^{(t+1)}-2$，则 $num<num^*$。因此可以说明 SCESW 方案满足完整性。

以下说明 SCESW 方案的机密性，即攻击者不能得到除第一个不同窗口的差值之外更多的信息块。可以这样分析，由 SCESW 方案的算法可知，数据的产生与 H 函数和主密钥 $mkey$ 有关。如果主密钥 $mkey$ 不泄露给非法的用户，则非法用户仅通过标签信息和密文信息不能得到原始明文的敏感信息。在密文比较算法 Cmp 时，假设密文为 $ciph=(I,(f_0,f_1,\cdots,f_{m-1}))$、$ciph^*=(I',(f_0',f_1',\cdots,f_{m-1}'))$。当比较得到第一个不同的窗口 j 时，标签 $token$ 中的分量 d_j 依赖于 d_{j+1} 和 B_{j+1}。当 $B_{j+1}=B_{j+1}'$ 时才能进一步比较 $c_j=B_j-B_j'\bmod(2^{(t+1)}-1)$。如果 $B_{j+1}\neq B_{j+1}'$ 时，则 $d_{j+1}\neq d_{j+1}'$，此时，就不满足进行比较的条件，故 $c_j\neq B_j-B_j'\bmod(2^{(t+1)}-1)$，这样比较的结果是错误的。故当第一个不同窗口出现之后 Cmp 将不再继续工作，直接输出第一个不同窗口的差值关系。因此攻击者不能得到除第一个不同窗口的差值之外更多的信息块。

本节可以保证 SCESW 方案在标准模型下满足弱不可区分性、完整性和机密性，定理 1 是用来证明 SCESW 方案在标准模型下的弱不可区分性，其他分析表明 SCESW 方案的完整性和机密性。

5 性能分析

本文采用 C++来实现 SCE 方案和 SCESW 方案，实验结果表明，SCESW 方案在存储、计算开销和效率上都要优于 SCE 方案。

5.1 理论分析

表 2 是 SCESW 方案与 SCE 方案的理论性能对比。由于每一步的生成计算开销主要来自于 H 函数的运算，因此，需要计算开销 H 函数的个数。表 2 中 L 表示数字的第 $L+1$ 位。2 个数字 $num_0^*=(\beta_0,\beta_1,\cdots,\beta_{n-1})$ 和 $num_1^*=(\gamma_0,\gamma_1,\cdots,\gamma_{n-1})$，$L$ 满足 $(\beta_L,\cdots,\beta_{n-1})=(\gamma_L,\cdots,\gamma_{n-1})$ 且 $\beta_{L-1}<\gamma_{L-1}$。另外 $\frac{n}{m}=t$。

表 2 SCESW 方案与 SCE 方案的对比

对比项	SCESW 方案	SCE 方案
标签长度/bit	mk	$(n+1)k$
加密阶段开销/ms	$3mh$	$3nh$
标签生成开销/ms	mh	$(n+1)h$
比较阶段开销/ms	$2(m-L+1)h$	$2(n-L+2)h$

下面给出 2 种方案对比的详细介绍。在参数生成算法 Gen 中，2 个方案都生成 $param=(n,H_1,H_2,H_3)$ 和 $mkey$。引入滑动窗口技术重写数字 num，这种方式在一定程度上降低计算开销和存储开销。

在标签生成算法 *Der* 中，SCESW 方案生成的标签是 $token=(d_1,d_2,\cdots,d_m)$。SCE 方案生成的标签是 $token=(d_1,d_2,\cdots,d_n)$。由此可知，在标签存储时 SCE 方案大约是 SCESW 方案的 t 倍。

在密文生成算法 *Enc* 中，SCESW 方案生成密文为 $ciph=(I,(f_0,f_1,\cdots,f_{m-1}))$，其中，$(f_0,f_1,\cdots,f_{m-1})$ 转化为整数 $F^*=\sum_{i=0}^{m-1}f_i(2^{(t+1)}-1)^i$ 存储。而 SCE 方案生成密文可以表示为 $ciph=(I,(f_0,f_1,\cdots,f_{n-1}))$，其中，$(f_0,f_1,\cdots,f_{n-1})$ 转化为整数 $F=\sum_{i=0}^{m-1}f_i3^i$ 存储。因此，SCESW 方案存储开销较短。因为密文存储在云端，不考虑密文存储长度，这里主要考虑资源受限的云租户。

这 2 个方案泄露的信息仅仅是第一个不同窗口的数值。定义窗口的大小为 t（其中，n 是 t 的倍数），SCESW 方案中开窗数量为 $\frac{n}{t}=m$，本文用 $\frac{t}{n}$ 表示数字的泄露比。图 3 表示当 $n=1\,024$ 时不同开窗长度对应的信息泄露比。

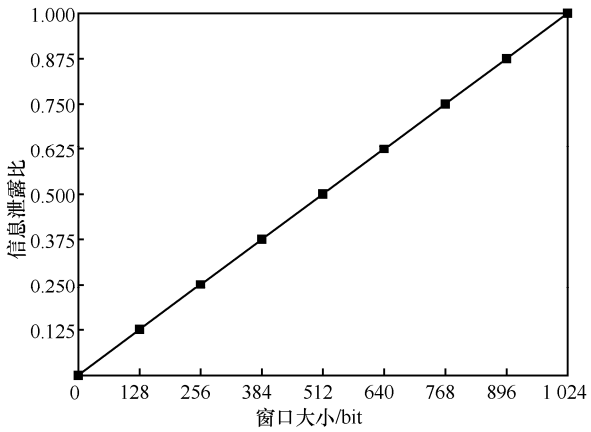

图 3　$n=1\,024$ 时不同开窗长度对应的信息泄露比

SCE 方案是 SCESW 方案的一个特例，即 $t=1$。由图 3 可知，随着窗口 t 的不断增大，敌手可以更精确地获得 2 个数字的大小关系。由于在实际中数字的长度一般都是 1 024 bit 之内的数字，当开窗值较小时，这意味着几乎不泄露任何有关数字的信息。在实际应用中，云租户可根据需求来调整窗口值的大小 t，从而实现效率和安全性的折中。

5.2　效率分析

5.1 节分析了 SCESW 方案和 SCE 方案之间的理论性能比较，本节主要分析 SCESW 方案和 SCE 方案之间的实验性能比较。本文的实现和应用均是在 Windows 操作系统中进行的，硬件配置情况为笔记本电脑（中央处理器为酷睿 i5 CPU（2.3 GHz），内存 4 GB，运行操作系统为 Ubuntu15.04）。开发平台采用 ASP.NET，开发语言采用 C++，集成开发环境为 VisualStudio2010。数据库采用 Microsoft 的 SQL Server 及 PBC（Pairing-Based Cryptography）库对 2 种方案的实际性能进行验证分析。实验结果是对 10 000 个数据进行测试 100 次，然后求平均值得到的。实验随机生成 n 为 1 024 bit 的数，k 为 160 bit。

2 种方案的标签长度对比如图 4 所示。由图 4 可以看出，随着 $\frac{m}{n}$ 比值的增大，SCE 方案标签的存储长度大致不变（约 160 000 bit），而 SCESW 方案随着 $\frac{m}{n}$ 比值的增大标签的存储长度逐渐变大。当 $\frac{m}{n}<\frac{1}{16}$ 时，SCESW 方案需要的标签存储长度远远小于 SCE 方案。实验中取 n 为 1 024 bit，k 为 160 bit。例如，当 $\frac{m}{n}=\frac{1}{4}$ 时表示开窗 4 bit，理论上 SCESW 方案存储标签长度为 mk=40 960 bit 且 SCE 方案为 $(n+1)k=164\,000$ bit。由图 4 可知，实验结果与表 2 中理论分析一致。其他情况类似，由此可知 SCESW 方案引用开窗技术在很大程度上节省了存储开销。

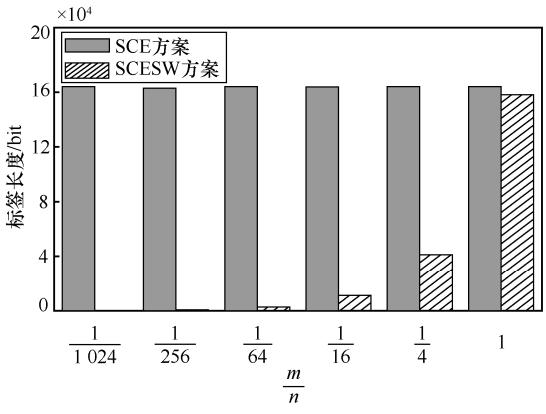

图 4　2 种方案的标签长度对比

2 种方案的加密阶段耗时对比如图 5 所示。由图 5 可以看出，随着 $\frac{m}{n}$ 比值的逐渐增大 SCE 方案在加密阶段耗时大致不变（约 9 ms），而 SCESW 方案随着 $\frac{m}{n}$ 比值的增大加密阶段耗时逐渐变大。当

$\frac{m}{n}<\frac{1}{16}$ 时，SCESW 方案需要的加密阶段耗时远远小于 SCE 方案。例如，当 $\frac{m}{n}=\frac{1}{4}$ 时表示开窗 4 bit，理论上，SCESW 方案加密阶段消耗 H 运算个数为 $3mh=768h$ 且 SCE 方案为 $3nh=3\,072h$。由图 5 可知，实验结果与表 2 中理论分析一致。其他情况类似，由此可知 SCESW 方案引用开窗技术在很大程度上提高了加密阶段的效率。

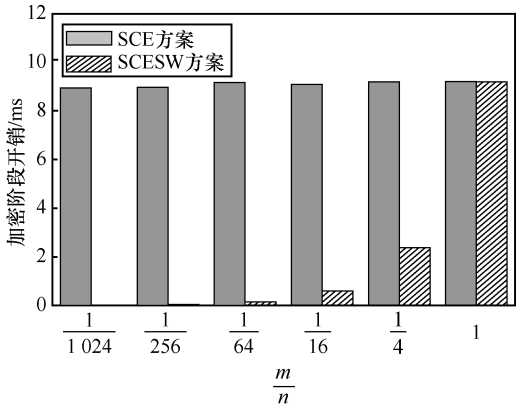

图 5 2 种方案的加密阶段耗时对比

2 种方案的标签生成阶段耗时对比如图 6 所示。由图 6 可以看出，随着 $\frac{m}{n}$ 比值的逐渐增大，SCE 方案在标签生成阶段耗时大致不变（约 3 ms），而 SCESW 方案随着 $\frac{m}{n}$ 比值的增大标签生成阶段耗时逐渐变大。当 $\frac{m}{n}<\frac{1}{16}$ 时 SCESW 方案需要的标签生成阶段耗时远远小于 SCE 方案。例如，当 $\frac{m}{n}=\frac{1}{4}$ 时表示开窗 4 bit，理论上 SCESW 方案标签生成阶段消耗 H 运算个数为 $mh=256h$ 且 SCE 方案为 $(n+1)h=1\,025h$。由图 6 可知，实验结果与表 2 中理论分析一致。其他情况类似，由此可知 SCESW 方案引用开窗技术在很大程度上提高了标签生成阶段的效率。

图 7 假设 $t=\frac{n}{m}=4$，进而研究随着 L 的变化比较阶段耗时的变化情况。实验中取 n 为 1 024 bit，L 表示数的第 $L+1$ 位。数字 $num_0^*=(\beta_0,\cdots,\beta_{n-1})$ 和 $num_1^*=(\gamma_0,\cdots,\gamma_{n-1})$，$L$ 满足 $(\beta_L,\cdots,\beta_{n-1})=(\gamma_L,\cdots,\gamma_{n-1})$ 且 $\beta_{L-1}<\gamma_{L-1}$。由图 7 可以看出，当开窗固定为 4 bit 时，SCE 方案和 SCESW 方案都是随着 L 的

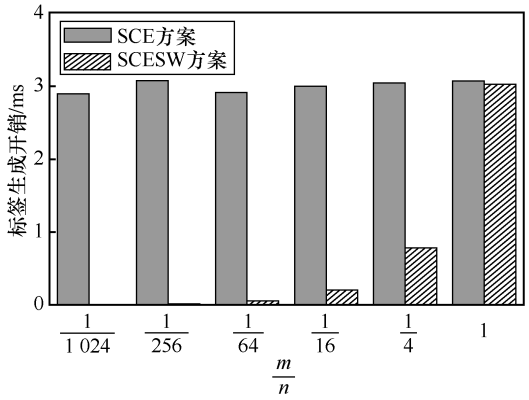

图 6 2 种方案的标签生成阶段耗时对比

增加比较阶段耗时逐渐减少。例如，$L=63$ 表示对于 1 024 bit 的数，SCE 方案从第 1024 位依次向前比较直到第 64 位才能出现不同位的值，此时比较阶段耗时大约是 6 ms。而对于 1 024 bit 的数 SCESW 方案从第 256 个窗口位依次向前比较直到第 16 个窗口才出现不同的值，此时比较阶段耗时大约是 2.5 ms。理论上 SCESW 方案比较阶段消耗 H 运算个数为 $2(m-L+1)h=386h$ 且 SCE 方案为 $2(n-L+2)h=1\,924h$。由图 7 可知，实验结果与表 2 中理论分析一致。其他情况类似，由此可知 SCESW 方案引用开窗技术在很大程度上节约了计算开销和存储开销。

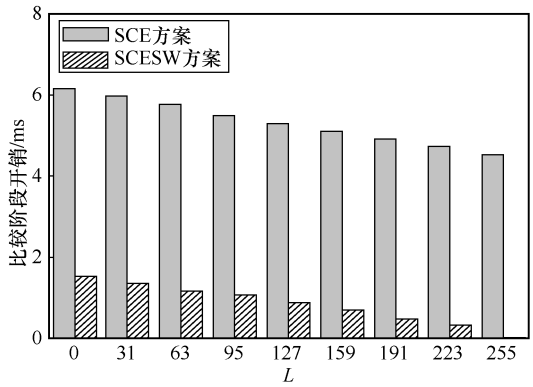

图 7 比较阶段随 L 变化时的耗时对比

表 3 给出了 2 种方案的功能比较。通过以上分析可以看出，SCESW 方案的理论性能分析和实际性能测试相一致。因为密文存储在云端，不考虑密文存储长度，这里主要考虑其他方面的优势，SCESW 方案对于资源受限的云租户来说是具有优势的，并且在实际应用中 SCESW 方案是可行的。Chen 等[18]提出的方案（以下简称 Chen 方案）采用的是 CE 方案结合滑动窗口技术，SCESW 方案采用的是 SCE 方案然后利用滑动窗口技术降低计算和

存储开销。由于 CE 方案的计算开销和存储开销很大，所以 SCE 方案才被人们研究和关注。SCESW 方案和 SCE 方案是比 Chen 方案有优势的。因此本文实验部分只对 SCESW 方案和 SCE 方案进行了分析。由于 Chen 方案具有代表性，表 3 给出了 3 种方案的功能比较。

表 3　　3 种方案功能比较

功能	SCESW 方案	SCE 方案	Chen 方案
存储开销	较小	较大	较大
计算开销	较小	较大	较大
开窗技术	支持	不支持	支持
效率性	较高	较低	较低

6　结束语

为了减小物联网加密数据比较的计算开销和存储开销的问题，本文利用滑动窗口技术提出一种 SCESW 方案。在实际应用中，对滑动窗口法做了一些优化，不再区分零窗口和非零窗口，对二进制形式统一进行开窗，且云租户可根据需求来调整窗口大小 t 从而实现效率和安全性的折中。严格的安全分析证明，SCESW 方案能够满足在标准模型下弱不可区分性，同时保证数据的完整性和机密性。实验结果表明，SCESW 方案在开销存储和效率上都要优于 SCE 方案。SCESW 方案对于物联网中资源受限的云租户来说是具有优势的。在图像加密与检索的场景中，如何运用 SCESW 方案对特征向量进行加密，从而使加密图像在云端进行查询与检索操作，这些问题都值得笔者继续深入研究。

参考文献：

[1] ATZORI L, IERA A, MORABITO G. The Internet of things: a survey[J]. Computer Networks, 2010, 54(15):2787-2805.

[2] AGRAWAL R, KIERNAN J, SRIKANT R, et al. Order preserving encryption for numeric data[C]//ACM SIGMOD International Conference on Management of Data. 2004: 563-574.

[3] AGRAWAL D, ABBADI A E, EMEKCI F, et al. Database management as a service: challenges and opportunities[C]//International Conference on Data Engineering. 2009: 1709-1716.

[4] KADHEM H, AMAGASA T, KITAGAWA H. A secure and efficient order preserving encryption scheme for relational databases[C]//The International Conference on Knowledge Management and Information Sharing Valencia. 2010: 25-35.

[5] LIU D, WANG S. Programmable order-preserving secure index for encrypted database query[C]// International Conference on Cloud Computing. 2012: 502-509.

[6] LEE S, TAE-JUN P, LEE D, et al. Chaotic order preserving encryption for efficient and secure queries on databases[J]. IEICE Transactions on Information & Systems, 2009, 92-D(11): 2207-2217.

[7] SHI E, BETHENCOURT J, CHAN T H H, et al. Multi-dimensional range query over encrypted data[C]//2007 IEEE Symposium on Security and Privacy. 2007: 350-364.

[8] POPA R A, REDFIELD C M S, ZELDOVICH N, et al. CryptDB: protecting confidentiality with encrypted query processing[C]//ACM Symposium on Operating Systems Principles. 2011: 85-100.

[9] GAO Y, MIAO M, WANG J, et al. Secure approximate nearest neighbor search over encrypted data[C]//Ninth International Conference on Broadband and Wireless Computing Communication and Applications. 2014: 578-583.

[10] TANG Q. Privacy preserving mapping schemes supporting comparison[C]//The 2010 ACM Workshop on Cloud Computing Security Workshop. 2010: 53-58.

[11] WANG C, CAO N, LI J, et al. Secure ranked keyword search over encrypted cloud data[C]//International Conference on Distributed Computing System. 2010: 253-262.

[12] DING Y, KLEIN K. Model-driven application-level encryption for the privacy of e-health data[C]//Ares'10 International Conference on Availability Reliability and Security. 2010: 341-346.

[13] FURUKAWA J. Request-based comparable encryption[M]. Computer Security–ESORICS. 2013: 129-146.

[14] FURUKAWA J. Short comparable encryption[C]//International Conference on Cryptology and Network Security. 2014: 337-352.

[15] ZOU Q, WANG J, YE J, et al. Efficient and secure encrypted image search in mobile cloud computing[J]. Soft Computing, 2016:1-11.

[16] ZHU Y, LIU L, CHEN X. Efficient first-price sealed-bid auction protocols from modified comparable encryption[C]// International Conference on Broadband and Wireless Computing, Communication and Applications. 2015:417-421.

[17] GUO Z, FU Y, CAO C. Secure first-price sealed-bid auction scheme[J]. EURASIP Journal on Information Security, 2017, 2017(1):16.

[18] CHEN P, YE J, CHEN X. A new efficient request-based comparable encryption scheme[C]// International Conference on Advanced Information Networking and Applications Workshops. 2015: 436-439.

[19] KOÇ C K. Analysis of sliding window techniques for exponentiation [J]. Computers & Mathematics with Applications, 1995, 30(10): 17-24.

[作者简介]

孟倩（1989-），女，山东菏泽人，西安电子科技大学博士生，主要研究方向为密码学、图像加密与检索等。

马建峰（1963-），男，陕西西安人，博士，西安电子科技大学教授、博士生导师，主要研究方向为信息安全、密码学与无线网络安全等。

陈克非（1959-），男，陕西西安人，博士，杭州师范大学教授、博士生导师，主要研究方向为密码理论与应用、网络与信息安全技术等。

苗银宾（1988-），男，陕西西安人，博士，西安电子科技大学讲师，主要研究方向为信息安全与无线网络安全等。

杨腾飞（1987-），男，陕西咸阳人，西安电子科技大学博士生，主要研究方向为密码学、图像加密与检索等。