具有统计特性的不经意传输协议

石润华[1],仲 红[1],崔 杰[1],许 艳[1,2],张 顺[1],黄刘生[2]
(1.安徽大学计算机科学与技术学院,安徽合肥 230601;2.中国科学技术大学苏州研究院,江苏苏州 215123)

摘 要: 不经意传输协议是安全多方计算中的基础协议,在保护用户隐私方面有着非常重要的应用.已有的不经意传输协议缺少统计分析能力.对此,引入多方安全求和及同态加密技术,并借助一种巧妙的编码方法,提出了一种具有统计特性的不经意传输协议.在保证协议原有正确性与安全性的基础上,增加了发送者的统计特性.即在一个执行周期后发送者能够统计出各个秘密消息发送出的总次数.理论分析表明该协议安全有效,在电子商务,医疗卫生等领域有着很好的应用前景.

关键词: 密码编码学; 安全多方计算; 不经意传输; 统计分析
中图分类号: TP309 **文献标识码:** A **文章编号:** 0372-2112 (2014)11-2273-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2014.11.022

A Novel Oblivious Transfer Protocol with Statistical Analysis

SHI Run-hua[1], ZHONG Hong[1], CUI Jie[1], XU Yan[1,2], ZHANG Shun[1], HUANG Liu-sheng[2]
(1. *School of Computer Science and Technology, Anhui University, Hefei, Anhui* 230601, *China*;
2. *Suzhou Institute for Advanced Study, University of Science and Technology of China, Suzhou, Jiangsu* 215123, *China*)

Abstract: Oblivious Transfer (OT) protocol is a fundamental building block for Secure Multiparty Computation and finds critically important applications in protecting privacy in various settings. Nevertheless, existing OT protocols do not include the procedure to conduct the statistical analysis in a cooperative environment. Here, a new OT protocol is proposed with a cooperative statistical analysis by pursuing secure multiparty computation of the sum, adopting the homomorphic encryption scheme, and using a clever coding method. Specifically, the sender can calculate the total number of each message sent privately within a given period. Theoretical analysis indicates that the new protocol allows for secure and efficient transfer of private messages in an oblivious way and shows great promise for practical applications in Electronic Commerce, Healthcare Management and other secure systems as well.

Key words: cryptography; secure multiparty computation; oblivious transfer; statistical analysis

1 引言

随着用户对敏感数据隐私保护的迫切需求,安全多方计算正成为研究热点[1~5].安全多方计算研究一组互不信任的参与方之间保护隐私的合作计算问题,最早由图灵奖得主姚期智先生于 1982 年提出[6],随后由著名密码学家 Oded Goldreich[7]进一步发展.

为了解决一些具体的安全多方计算问题,学者们构造了若干基础协议.例如比特承诺、秘密比较、不经意传输等.其中,不经意传输能够保证服务接收者以不经意的方式得到服务发送者的消息,确保接收者的隐私不被发送者所知.它是密码学中的一个根基性的原子协议[8,9],可作为基本组件构造其它安全协议,例如零知识证明,匿名秘密分享,不经意电路计算等.它也可直接应用于电子商务领域以保护用户的隐私[10],例如公平合同的签署,保护隐私的投票,不经意的数据库查询/网上订购/在线付费浏览等.

不经意传输(Oblivious Transfer,简记为 OT)的概念最早由 Rabin 在文献[11]中提出,随后该概念被 Even 等人在文献[12]中发展为二中选一的 OT(简记为 OT_1^2),再后来 Brassard 等[13]又提出了 n 选 1 的 OT(简记为 OT_1^n).接着,学者们又提出了更为一般的 n 选 k 的 OT 协议[14],以及具有一些特殊扩展功能的 OT 协议[15~19].例如,带承诺的,隐藏认证的,可验证的,分布式的,有价

收稿日期:2013-05-03;修回日期:2014-01-21;责任编辑:蓝红杰
基金项目:国家自然科学基金(No.61173187,No.61173188,No.11301002);安徽省自然科学基金(No.11040606M1411,No.1408085QF107);安徽省高等教育振兴计划重大教学改革资助项目(No.2013zdjy040);安徽大学搏士科研启动经费项目(No.33190187);安徽大学"信息安全"新专业项目(No.17110099)

格的不经意传输,以及量子不经意传输.

以上这些不经意传输协议均能保证发送者对于实际发送出的消息一无所知.这一方面保护了接收者的隐私,但另一方面又牺牲了发送者的部分利益.例如发送者不能对最终发送出的消息进行统计分析,更不能统计每条消息发送出的具体次数,从而对于"哪些消息最受欢迎或有最大需求"同样一无所知.实际上,这样的统计分析在一些电子商务领域尤其需要.例如,不经意的网上订购、不经意的在线付费浏览、不经意的股票交易等.它能方便发送者(即商家)作出获取更大利润的投资决策,定期订购或添加一些具有较大需求的消息(可能是情报,也可能是其它数字商品,例如图片,音乐,股票等).而在一些保护隐私的电子政务或公共医疗卫生系统中,这样的统计分析同样能便于政府或专业机构作出正确的导向性的宏观决策.

本文的主要目的是建立一种新的具有统计特性的不经意传输模型,并设计能满足发送者统计需求的不经意传输协议.在新的不经意传输协议中,除了满足已有不经意传输协议的正确性即保护发送者和接收者的隐私性外,还要求能满足发送者的统计特性.具体来说,就是在经过一个执行周期后发送者能够统计出各个秘密消息发送出的总次数.

2 预备知识

2.1 具有统计特性的不经意传输

从信息论的角度来看 OT,OT_1^2、OT_1^n 三者是等价的,而 OT_k^n 是它们的更为一般形式.因而,下面我们仅对更为一般的具有统计特性的 OT_k^n 协议进行定义.新模型中,存在一个发送方 S,以及一个涉及统计范围的接收者集合,不妨假定该集合拥有 t 个接收者 $R_1,R_2,\cdots,R_t$.

定义(具有统计特性的 $\mathbf{OT_k^n}$ 协议). 发送者 S 有 n 个秘密消息 $m_1,m_2,\cdots,m_n$;每个接收者 $R_i(1\leqslant i\leqslant t)$ 选择一组整数 $i_1,i_2,\cdots,i_k(i_j\in\{1,2,\cdots,n\},1\leqslant j\leqslant k)$.简单起见,这里要求 $i_{j_1}\neq i_{j_2}(j_1\neq j_2)$,即 R_i 选择的 k 个消息不重复.协议结果要求:

(1)每个接收者 R_i 得到期望的秘密消息 $m_{i_1},m_{i_2},\cdots,m_{i_k}$,除此以外,对于其它的秘密 $m_l(l\neq i_j,1\leqslant j\leqslant k)$ 一无所知;

(2)发送者 S 对接收者 R_i 的选择 $i_1,i_2,\cdots,i_k$ 一无所知,即她不知道 R_i 到底选择了哪 k 个秘密消息;

(3)当所有 t 个接收者得到他们所期望的消息后,发送者 S 在所有接收者的协助下能够统计出每个秘密消息发送出的总次数.

以上定义中,除了满足不经意传输协议的基本属性外,还要求满足发送者的统计特性.也即,具有统计特性的不经意传输协议必须满足如下四个属性.(1)正确性:完成协议后,接收者 R_i 能够得到他所期望的正确的消息;(2)接收者的隐私性:接收者 R_i 的选择是隐私的;(3)发送者的安全性:发送者 S 的其它消息是秘密的;(4)发送者的统计特性:经过一个周期后,发送者 S 能够统计每条秘密消息发送出的总次数.

2.2 Paillier 同态加密

1999 年 P. Paillier 提出了一个概率公钥加密方案[20],具体方案如下.

密钥生成:选取两个大的强素数 p,q,令 $N=pq$,$\lambda(N)=lcm(p-1,q-1)$.随机选取 $g\in Z_{N^2}^*$,要求 g 的阶为 N 的非零整数倍,即 $N\mid ord_{N^2}(g)$,且 $\gcd(L(g^{\lambda(N)}\bmod N^2),N)=1$(函数 $L(\cdot)$ 的定义在解密部分说明).则 (N,g) 为公钥,$\lambda(N)$ 为私钥.

加密 对于明文 $m\in Z_N$,随机选择整数 $r\in Z_N^*$,生成的密文 $c=E(m)=g^m r^N \bmod N^2$.

解密 对于密文 c,解密后明文 $m=\dfrac{L(c^{\lambda(N)}\bmod N^2)}{L(g^{\lambda(N)}\bmod N^2)}\bmod N$,其中 $L(u)=(u-1)/N$;$u\in Z_{N^2}$ 且 $u\equiv 1 \bmod N$.

Paillier 加、解密的正确性基于以下两个等式[20].对于任意 $w\in Z_{N^2}^*$,有:

$$w^{\lambda(N)}\equiv 1 \bmod N \tag{1}$$

$$w^{N\lambda(N)}\equiv 1 \bmod N^2 \tag{2}$$

Paillier 加密方案的安全基础基于"计算合数剩余类假设"[20,21]:定义函数 ε_g,

$$\varepsilon_g: Z_n\times Z_N^*\to Z_{N^2}^*$$
$$(x,y)\to g^x\cdot y^N \bmod N^2 \tag{3}$$

如果 g 在 $Z_{N^2}^*$ 中的阶为 N 的倍数,则 ε_g 是一一映射.那么对于给定的 $c\in Z_{N^2}^*$,$x\in Z_N$,则 $\exists\, y\in Z_N^*$,使得 $\varepsilon_g(x,y)=c$.这样的 $\varepsilon_g(x,y)$ 称为 c 的 N 次剩余类,用 $[[c]]_g$ 表示.目前认为,对于给定的 N,g 和 c,计算 $[[c]]_g$ 问题是困难的.

Paillier 加密方案具有加法同态性质.即给定明文 m_1,m_2,有

$$\begin{aligned}E(m_1)\cdot E(m_2)&=[g^{m_1}r_1^N \bmod N^2]\cdot[g^{m_2}r_2^N \bmod N^2]\\&=g^{m_1+m_2}(r_1\cdot r_2)^N \bmod N^2\\&\quad(\text{令 } r=r_1\cdot r_2 \bmod N)\\&=E(m_1+m_2)\end{aligned} \tag{4}$$

此外,Paillier 加密算法还有另外一个好的性质 - 对密文重加密性质[22]:随机选择一个 $s\in Z_N^*$,计算 $c'=c\cdot s^N=g^m\cdot(s\cdot r)^N \bmod N^2$,此时密文 c 及 c' 均可解密得到明文 m.

2.3 Partial-DL-ZN2P 困难假设

Partial-DL-ZN2P(Partial Discrete Logarithm problem in

$Z_{N^2}^*$)困难假设[20,23]:已知 RSA 模数 $N=pq$,要求 p,q 为两个大的强素数(即 $p=2p'+1, q=2q'+1, p', q'$ 为素数).另外已知 $g\in Z_{N^2}^*$,要求 g 的阶为 N 的非零整数陪(这里 N,g 即 Paillier 加密算法中的公钥).则不存在一个有效的概率多项式时间算法能以不可忽略的概率由 $h=g^x \bmod N^2$ 计算出 $x \bmod N$,即由 $h=g^x \bmod N^2$ 计算 $x \bmod N$ 是困难的.

3　建议的协议

假定发送者 S 有 n 个秘密消息 $m_1, m_2, \cdots, m_n$($m_i \in Z_N$),而参与统计的集合有 t 个接收者 $R_1, R_2, \cdots, R_t$.发送者和所有接收者事先约定用一个 n 维的 0、1 向量来表示消息的编号.例如,用单位向量$(0,0,1,0,,0)^{\mathrm{T}}$来表示第 3 条消息 m_3 的编号;用单位向量$(0,0,\cdots,0,1)^{\mathrm{T}}$来表示第 n 条消息m_n的编号.即,第 i 条消息用一个第 i 位分量为 1 其它所有分量为 0 的 n 维单位向量来表示它的编号.显然这些单位向量累加和的各分量就对应了各消息的累加次数.

具体协议包含三个阶段:系统初始化阶段,不经意传输阶段,统计分析阶段.执行完该协议后,每个接收者 $R_i(1\leqslant i\leqslant t)$恰好得到他所期望的 k 个消息 $m_{i_1}, m_{i_2}, \cdots, m_{i_k}$($i_k\in\{1,2,\cdots,n\}$),并且不能多得到其它的消息.另外,发送者 S 不能确定接收者 R_i 选择的是哪些消息,也即 R_i 所选择的消息的编号是隐私的.但当所有接收者同意发送者实施统计时,发送者 S 能统计每条消息秘密发送出的总次数,从而推断哪些消息在该群组内最受欢迎.

系统初始化阶段

发送者 S:随机生成两个大的强素数 p,q,计算 $N=pq$,$\lambda(N)=lcm(p-1,q-1)$(后简记为 λ,即 $\lambda=\lambda(N)$).随机选择 $g\in Z_{N^2}^*$,要求 $N\mid ord_{N^2}(g)$,且 $\gcd(L(g^{\lambda(N)} \bmod N^2), N)=1$.其中$(N,g)$为 Paillier 加密算法的公钥,$\lambda$ 为相应的私钥.另外随机选取 $h\in Z_N^*$(对于任意的 $r\in Z_N^*$,要求 $h\neq r^N \bmod N^2$,即 $h^\lambda\neq 1 \bmod N^2$).发送者公开系统参数 N,g,h;秘密保存 λ.

不经意传输阶段

①对于每个接收者 R_i,若他期望得到消息 $m_{i_1}, m_{i_2}, \cdots, m_{i_k}$($i_k\in\{1,2,\cdots,n\}$),那么根据编号规则,他首先生成这 k 个消息对应的 n 维单位向量 $\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \cdots, \boldsymbol{x}_{i_k}$.接着根据 $t+1$ 进制的转换规则,把这些单位向量转换为相应的十进制整数 $x_{i_1}, x_{i_2}, \cdots, x_{i_k}$.例如,若 $n=4$(向量的维数为 4),$t=7$,则转换前的进制为 8,转换后的进制为 10.那么某个 $\boldsymbol{x}_{ij}=(1,0,0,0)^{\mathrm{T}}$ 唯一转换为 $\boldsymbol{x}_{ij}=1\cdot 8^3+0\cdot 8^2+0\cdot 8^1+0\cdot 8^0=512$;反过来一个适当大小的十进制数也可唯一转换为一个 n 维的 $t+1$ 进制向量.注:在被统计的群组内,每条消息被秘密发送出去的次数最多为 t 次,所以这里我们引入 $t+1$ 进制,以保证最后的累加向量没有进位.

②接收者 R_i 在 Z_N 上随机生成 k 个秘密整数 $y_{i_1}, y_{i_2}, \cdots, y_{i_k}$,并计算:

$$y_i=\sum_{j=1}^{k} y_{i_j} (\bmod N) \tag{5}$$

进而,R_i 计算 $Y_{i_1}=g^{x_{i_1}} r_{i_1}^N h^{y_{i_1}} \bmod N^2$, $Y_{i_2}=g^{x_{i_2}} r_{i_2}^N h^{y_{i_2}} \bmod N^2, \cdots, Y_{i_k}=g^{x_{i_k}} r_{i_k}^N h^{y_{i_k}} \bmod N^2$,并把$(Y_{i_1}, Y_{i_2}, \cdots, Y_{i_k})$发送给 S.其中 $r_{i_j}\in {}_R Z_N^*(1\leqslant j\leqslant k)$.

③收到$(Y_{i_1}, Y_{i_2}, \cdots, Y_{i_k})$后,发送者 S 执行以下程序:

For $j=1$ to k

　对每个 Y_{i_j},发送者 S 随机选择 $s_{i_j}\in Z_N^*$,计算 $\alpha_{i_j}=h^{\lambda s_{i_j}} \bmod N^2$;

　计算 $\beta_{i_j,1}=m_1(\frac{Y_{i_j}}{g^{\sigma_1}})^{\lambda s_{i_j}} \bmod N^2$,

　$\beta_{i_j,2}=m_2(\frac{Y_{i_j}}{g^{\sigma_2}})^{\lambda s_{i_j}} \bmod N^2 \cdots$,

　$\beta_{i_j,n}=m_n(\frac{Y_{i_j}}{g^{\sigma_n}})^{\lambda s_{i_j}} \bmod N^2$.这里 $\sigma_1,\sigma_2,\cdots,\sigma_n$ 对应着单位向量$\boldsymbol{\sigma}_1=(1,0,\cdots 0)^{\mathrm{T}}, \boldsymbol{\sigma}_2=(0,1,\cdots 0)^{\mathrm{T}}, \cdots, \boldsymbol{\sigma}_n=(0,0,\cdots 1)^{\mathrm{T}}$经$(t+1)$进制转换而得到的十进制整数;

　发送者 S 把 α_{i_j}及$(\beta_{i_j,1}, \beta_{i_j,2}, \cdots, \beta_{i_j,n})$发送给接收者 R_i.

EndFor

---

④接收者 R_i 收到发送者 S 以上加密消息后,执行以下程序:

For $j=1$ to k

　IF $(\alpha_{i_j}\neq 1)$ 接收者 R_i 计算$\beta_{i_j,i_j}/(\alpha_{i_j})^{y_{i_j}}$从而得到消息 m_{i_j}.

　//下标 i_j 表示第 i 个接收者 R_i 秘密选择的第 j 个消息的编号.

EndFor

统计分析阶段

发送者最终统计的信息只包含同意被统计的接收者的信息,并不覆盖个别不愿意参与统计的接收者的信息.对于不愿意参与统计的接收者,仅仅执行不经意传输阶段协议,不需继续执行以下协议(相当于传统的不经意传输).下面不妨假定所有接收者均同意协助发送者实施统计分析.

当所有 t 个用户均得到了自己所想要的秘密消息后,协助发送者执行一个多方安全求和协议,把各自的子秘密 y_i 累加起来.有了$\sum y_i$的值,发送者 S 才能解密(即统计)每条消息 m_j 所发送出的总次数 $d_j(1\leqslant j\leqslant n)$.

①每个接收者 $R_i(1\leqslant i\leqslant t)$随机生成一个 $i-1$ 次

的秘密多项式

$$\begin{aligned}
&f_1(x) = a_{10}\\
&f_2(x) = (a_{20} + a_{21}x) \bmod N\\
&f_3(x) = (a_{30} + a_{31}x + a_{32}x^2) \bmod N\\
&\quad\vdots\\
&f_i(x) = (a_{i0} + a_{i1}x + a_{i2}x^2 + \cdots + a_{i(i-1)}x^{i-1}) \bmod N\\
&\quad\vdots\\
&f_t(x) = (a_{t0} + a_{t1}x + a_{t2}x^2 + \ldots + a_{t(t-1)}x^{t-1}) \bmod N
\end{aligned} \tag{6}$$

其中 $a_{i0} = y_i$, $a_{ij} \in {}_RZ_N^*$ $(1 \leqslant j \leqslant i-1)$. 接着 R_i 计算共享份额 $f_i(j)$ 并发送给其他接收者 $R_j(1 \leqslant j \leqslant i-1)$. 收到其它所有接收者的共享份额后,接收者 R_i 计算:

$$F(i) = \sum_{j=i}^{t} f_j(i) (\bmod N) \quad (1 \leqslant i \leqslant t) \tag{7}$$

其中 $f_1(1) = a_{10} = y_1$.

继而,所有接收者协助发送者 S 计算 $F(0)$: R_i 计算密文 $E(F(i)) = g^{F(i)} r^N \bmod N^2$, $r \in {}_RZ_N^*$. 并把 $E(F(i))$ 发送给 S. S 收到 $E(F(i))$ 后解密得到 $F(i)$:

$$F(i) = \frac{L(E(F(i))^{\lambda} \bmod N^2)}{L(g^{\lambda} \bmod N^2)} \bmod N \tag{8}$$

根据拉格朗日插值公式, S 计算:

$$F(0) = \sum_{i=1}^{t} F(i) \prod_{j=1, j\neq i}^{t} \frac{0-j}{i-j} \bmod N \tag{9}$$

根据 Shamir 秘密共享,显然有,

$$\begin{aligned}
&F(0) = \sum_{i=1}^{t} f_i(0) \bmod N\\
&F(0) = (y_1 + y_2 + \cdots + y_t) \bmod N
\end{aligned} \tag{10}$$

接着,发送者 S 计算:

$$y^* = N - F(0) \quad (\text{即 } F(0) + y^* = N) \tag{11}$$

②收集到 t 个接收者的所有加密消息后,发送者 S 计算

$$Y_i = \prod_{j=1}^{k} Y_{i_j} \bmod N^2 \tag{12}$$

$$Y = \prod_{i=1}^{t} Y_i \bmod N^2 \tag{13}$$

③发送者 S 计算

$$Y^* = Y \cdot h^{y^*} (\bmod N^2) \tag{14}$$

④发送者 S 计算

$$d = \frac{L(Y^{*\lambda} \bmod N^2)}{L(g^{\lambda} \bmod N^2)} \bmod N \tag{15}$$

⑤根据 $t+1$ 进制的编码规则,发送者 S 由 d(十进制)计算出对应的 n 维向量 $\boldsymbol{d}$($t+1$ 进制). 则 $\boldsymbol{d} = (d_1, d_2, \cdots, d_n)^{\mathrm{T}}$ 的第 j 个分量 d_j 就是第 j 条消息 m_j 被秘密发送出的总次数,也即消息 m_j 已被 t 个接收者中的 d_j 个接收者秘密获得,但至于具体是哪些接收者无从获知.

4 协议分析

4.1 正确性证明

定理 1 在不经意传输阶段,接收者 R_i 通过公式 $\beta_{i_j, i_j} / (\alpha_{i_j})^{y_{i_j}}$ 计算得到的是正确的消息 $m_{i_j}(1 \leqslant j \leqslant k)$,也即正是他所期望的第 i_j 条秘密消息.

证明 因为 $\alpha_{i_j} = h^{\lambda s_{i_j}} \bmod N^2$,所以有

$$\begin{aligned}
(\alpha_{i_j})^{y_{i_j}} &= (h^{\lambda s_{i_j}})^{y_{i_j}} \bmod N^2\\
&= (h^{y_{i_j}})^{\lambda s_{i_j}} \bmod N^2
\end{aligned} \tag{16}$$

又因为 $\beta_{i_j, i_j} = m_{i_j} (\frac{Y_{i_j}}{g^{\sigma_{i_j}}})^{\lambda s_{i_j}} \bmod N^2$,

而 $Y_{i_j} = g^{x_{i_j}} r_{i_j}^N h^{y_{i_j}} \bmod N^2$,故有,

$$\begin{aligned}
\beta_{i_j, i_j} / (\alpha_{i_j})^{y_{i_j}} &= [m_{i_j} (\frac{Y_{i_j}}{g^{\sigma_{i_j}}})^{\lambda s_{i_j}} \bmod N^2] / (h^{y_{i_j}})^{\lambda s_{i_j}} \bmod N^2\\
&= [m_{i_j} (\frac{g^{x_{i_j}} r_{i_j}^N h^{y_{i_j}}}{g^{\sigma_{i_j}}})^{\lambda s_{i_j}} \bmod N^2] / (h^{y_{i_j}})^{\lambda s_{i_j}} \bmod N^2\\
&= [m_{i_j} (r_{i_j}^N h^{y_{i_j}})^{\lambda s_{i_j}} \bmod N^2] / (h^{y_{i_j}})^{\lambda s_{i_j}} \bmod N^2\\
&\quad (\text{由 } x_{i_j} = \sigma_{i_j} \text{可得})\\
&= [m_{i_j} (r_j^{\lambda N i})^{s_{i_j}} (h^{y_{i_j}})^{\lambda s_{i_j}} \bmod N^2] / (h^{y_{i_j}})^{\lambda s_{i_j}} \bmod N^2\\
&= [m_{i_j} (h^{y_{i_j}})^{\lambda s_{i_j}} \bmod N^2] / (h^{y_{i_j}})^{\lambda s_{i_j}} \bmod N^2\\
&\quad (\text{由公式(2)可得})\\
&= m_{i_j}
\end{aligned} \tag{17}$$

即,接收者 R_i 通过公式 $\beta_{i_j, i_j} / (\alpha_{i_j})^{y_{i_j}}$ 计算而得到的是正确的消息 $m_{i_j}(1 \leqslant j \leqslant k)$.

定理 2 在统计分析阶段,发送者 S 能够正确得到各个消息发送出的总次数.

证明 首先,根据 Shamir 秘密共享, $F(0) = (y_1 + y_2 + \cdots + y_t) \bmod N$.

其次,又因为 $Y_{i_j} = g^{x_{i_j}} r_{i_j}^N h^{y_{i_j}} \bmod N^2$,而 $Y_i = \prod_{j=1}^{k} Y_{i_j} \bmod N^2$,所以有,

$$\begin{aligned}
Y_i &= \prod_{j=1}^{k} Y_{i_j} \bmod N^2\\
&= \prod_{j=1}^{k} (g^{x_{i_j}} r_{i_j}^N h^{y_{i_j}}) \bmod N^2\\
&= g^{\sum_{j=1}^{k} x_{i_j}} (\prod_{j=1}^{k} r_{i_j})^N h^{\sum_{j=1}^{k} y_{i_j}} \bmod N^2\\
&= g^{\sum_{j=1}^{k} x_{i_j}} (\prod_{j=1}^{k} r_{i_j})^N h^{y_i} \bmod N^2 \quad (\text{根据公式(5)})
\end{aligned} \tag{18}$$

因而,

$$\begin{aligned}
Y &= \prod_{i=1}^{t} Y_i \bmod N^2\\
&= \prod_{i=1}^{t} (g^{\sum_{j=1}^{k} x_{i_j}} (\prod_{j=1}^{k} r_{i_j})^N h^{y_i}) \bmod N^2
\end{aligned}$$

$$= g^{\sum_{i=1}^{t}\sum_{j=1}^{k}x_{i_j}}\left(\prod_{i=1}^{t}\prod_{j=1}^{k}r_{i_j}\right)^N h^{\sum_{i=1}^{t}y_{i_j}} \bmod N^2$$

$$= g^{\sum_{i=1}^{t}\sum_{j=1}^{k}x_{i_j}}\left(\prod_{i=1}^{t}\prod_{j=1}^{k}r_{i_j}\right)^N h^{F(0)} \bmod N^2 \text{(根据式(10))}$$

$$Y^* = Y \cdot h^{y^*} \bmod N^2$$

$$= g^{\sum_{i=1}^{t}\sum_{j=1}^{k}x_{i_j}}\left(\prod_{i=1}^{t}\prod_{j=1}^{k}r_{i_j}\right)^N h^{F(0)} \cdot h^{y^*} \bmod N^2$$

$$= g^{\sum_{i=1}^{t}\sum_{j=1}^{k}x_{i_j}}\left(\prod_{i=1}^{t}\prod_{j=1}^{k}r_{i_j}\right)^N h^{(F(0)+y^*)} \bmod N^2$$

$$= g^{\sum_{i=1}^{t}\sum_{j=1}^{k}x_{i_j}}\left(\prod_{i=1}^{t}\prod_{j=1}^{k}r_{i_j}\right)^N h^N \bmod N^2 \quad \text{(根据式(11))}$$

$$= g^{\sum_{i=1}^{t}\sum_{j=1}^{k}x_{i_j}}\left(\left(\prod_{i=1}^{t}\prod_{j=1}^{k}r_{i_j}\right)\cdot h\right)^N \bmod N^2 \tag{19}$$

令 $r = \left(\prod_{i=1}^{t}\prod_{j=1}^{k}r_{i_j}\right)\cdot h \bmod N$,则有,

$$Y^* = g^{\sum_{i=1}^{t}\sum_{j=1}^{k}x_{i_j}} r^N \bmod N^2 \tag{20}$$

从而,

$$d = \frac{L(Y^{*\lambda} \bmod N^2)}{L(g^\lambda \bmod N^2)} \bmod N \tag{21}$$

$$= \sum_{i=1}^{t}\sum_{j=1}^{k}x_{i_j} \quad \text{(根据 Paillier 加密原理)}$$

根据事先约定的消息编号规则,每个消息对应着一个 n 维的 0、1 单位向量,显然所有这些单位向量相加之和(即 $\sum_{i=1}^{t}\sum_{j=1}^{k}\boldsymbol{x}_{i_j}$)就等于统计向量 $\boldsymbol{d}$.反过来,根据 $t+1$ 进制的编码性质,发送者 S 由 d 计算出的 n 维向量 $\boldsymbol{d}$ 就是 $\sum_{i=1}^{t}\sum_{j=1}^{k}\boldsymbol{x}_{i_j}$,即统计向量是正确的.例如,发送者总的有 4 个消息 m_1, m_2, m_3, m_4.另有 3 个接收者 R_1, R_2 和 R_3.其中,R_1 期望得到 m_1 和 m_3,因而生成单位向量 $\boldsymbol{x}_{1_1} = (1,0,0,0)^{\mathrm{T}}$ 和 $\boldsymbol{x}_{1_2} = (0,0,1,0)^{\mathrm{T}}$,继而得到 $x_{1_1} = 64, x_{1_2} = 4$;$R_2$ 期望得到 m_3 和 m_4,因而生成单位向量 $\boldsymbol{x}_{2_1} = (0,0,1,0)^{\mathrm{T}}$ 和 $\boldsymbol{x}_{2_2} = (0,0,0,1)^{\mathrm{T}}$,继而得到 $x_{2_1} = 4, x_{2_2} = 1$;R_3 期望得到 m_1, m_2 和 m_3,因而生成单位向量 $\boldsymbol{x}_{3_1} = (1,0,0,0)^{\mathrm{T}}, \boldsymbol{x}_{3_2} = (0,1,0,0)^{\mathrm{T}}$ 和 $\boldsymbol{x}_{3_3} = (0,0,1,0)^{\mathrm{T}}$,继而得到 $x_{3_1} = 64, x_{3_2} = 16$ 和 $x_{3_3} = 4$.则

$$d = (x_{1_1} + x_{1_2}) + (x_{2_1} + x_{2_2}) + (x_{3_1} + x_{3_2} + x_{3_3})$$
$$= 157$$

根据 4 进制编码规则,反过来由 d 可计算得到 $\boldsymbol{d}' = (2,1,3,1)^{\mathrm{T}}$.另一方面,

$$\boldsymbol{d} = (\boldsymbol{x}_{1_1} + \boldsymbol{x}_{1_2}) + (\boldsymbol{x}_{2_1} + \boldsymbol{x}_{2_2}) + (\boldsymbol{x}_{3_1} + \boldsymbol{x}_{3_2} + \boldsymbol{x}_{3_3})$$
$$= (2,1,3,1)^{\mathrm{T}}$$

显然有 $\boldsymbol{d}' = \boldsymbol{d}$.所以由 $\boldsymbol{d}'$ 可知 m_1, m_2, m_3, m_4 发送出的总次数分别为:2,1,3,1,其中消息 m_3 最受欢迎.

4.2 安全性分析

定理 3　在半诚实模型下,接收者 $R_i(1 \leqslant i \leqslant N)$ 的选择 i_j 是无条件安全的,发送者 S 得不到 i_j 的任何信息.

证明　通过引入随机数 y_{i_j},接收者 $R_i(1 \leqslant i \leqslant N)$ 把他的选择 x_{i_j} 加密后隐藏于 Y_{i_j} 之中($Y_{i_j} = g^{x_{i_j}} r_{i_j}^N h^{y_{i_j}} \bmod N^2$).除非发送者 S 选择的 h 存在以下关系:对于某个 $r \in Z_N^*$,有 $h = r^N \bmod N^2$,也即 $Y_{i_j} = g^{x_{i_j}}(r_{i_j} \cdot r^{y_{i_j}})^N \bmod N^2$,从而她能解密得到 x_{i_j},继而得到 i_j.但此时必定 $h^\lambda = 1 \bmod N^2$,且所有 $\alpha_{i_j} = 1(\alpha_{i_j} = h^{\lambda s_{i_j}} \bmod N^2, 1 \leqslant i \leqslant t, 1 \leqslant j \leqslant k)$.这样的 h 以及 α_{i_j} 势必能被接收者发现.

不妨设 $c_{i_j} = g^{x_{i_j}} r_{i_j}^N \bmod N^2$,根据 Paillier 加密原理可知,此时 c_{i_j} 与 (x_{i_j}, r_{i_j}) 存在一一映射,若已知 c_{i_j},那么根据私钥 λ 可以解密出 x_{i_j}.但是在建议的方案中发送者并不知道 c_{i_j} 的值,因为他所获取的仅仅是 $Y_{i_j} = g^{x_{i_j}} r_{i_j}^N h^{y_{i_j}} \bmod N^2$,而 $Y_{i_j} = c_{i_j} \cdot h^{y_{i_j}} \bmod N^2$. y_{i_j} 是在 Z_N 上随机选取的,因而对发送者来说,仅由 Y_{i_j} 得不到 c_{i_j} 的任何信息,当然也就不能解密得到 x_{i_j} 进而获知 i_j.

实际上,若不满足 $h = r^N \bmod N^2$ 条件,在等式 $Y_{i_j} = g^{x_{i_j}} r_{i_j}^N h^{y_{i_j}} \bmod N^2$ 中,有三个未知数 x_{i_j}, y_{i_j} 和 r_{i_j}.对任意 x_{i_j}',理论上都可以找到一个 y_{i_j}' 满足条件 $g^{x_{i_j}} h^{y_{i_j}} = g^{x_{i_j}'} h^{y_{i_j}'}$,也即 $g^{x_{i_j}} r_{i_j}^N h^{y_{i_j}} = g^{x_{i_j}'} r_{i_j}^N h^{y_{i_j}'}$.因此即使发送者有无限的计算能力,接收者的选择也是无条件安全的.

定理 4　接收者 $R_i(1 \leqslant i \leqslant N)$ 试图通过发送者 S 其余的加密信息得到对应的明文消息(没有经过事先选择的),其难度等同于计算 Partial-DL-ZN2P 问题,即计算上是不可行的.

证明　假定恶意接收者 R_i 能以不可忽略的概率 ε 计算得到 m_l,下证他也能以不可忽略的概率 ε 计算 Partial-DL-ZN2P 问题(反证法). m_l 是没有经过事先选择的消息,对应的加密信息为 $m_l\left(\frac{Y_{i_j}}{g^{\sigma_l}}\right)^{\lambda s_{i_j}} \bmod N^2$.这里 $l \neq i_j$(i_j 是事先选择的消息编号).根据 $Y_{i_j} = g^{x_{i_j}} r_{i_j}^N h^{y_{i_j}} \bmod N^2$,则有,

$$m_l\left(\frac{Y_{i_j}}{g^{\sigma_l}}\right)^{\lambda s_{i_j}} \bmod N^2 = m_l\left(\frac{g^{x_{i_j}} r_{i_j}^N h^{y_{i_j}}}{g^{\sigma_l}}\right)^{\lambda s_{i_j}} \bmod N^2$$
$$= m_l\left(g^{x_{i_j} - \sigma_l} r_{i_j}^N h^{y_{i_j}}\right)^{\lambda s_{i_j}} \bmod N^2 \quad (\text{令 } x_{i_j}^* = x_{i_j} - \sigma_l)$$
$$= m_l\left(g^{x_{i_j}^*} r_{i_j}^N h^{y_{i_j}}\right)^{\lambda s_{i_j}} \bmod N^2$$
$$= m_l\left(g^{x_{i_j}^*}\right)^{\lambda s_{i_j}} \left(r^{s_{i_j} i_j}\right)^{\lambda N} \left(h^{\lambda s_{i_j}}\right)^{y_{i_j}} \bmod N^2$$
$$= m_l\left(g^{x_{i_j}^*}\right)^{\lambda s_{i_j}} \left(h^{\lambda s_{i_j}}\right)^{y_{i_j}} \bmod N^2 \quad \left(\left(r^{s_{i_j} i_j}\right)^{\lambda N} = 1 \bmod N^2\right) \tag{22}$$

其中 $\left(h^{\lambda s_{i_j}}\right)^{y_{i_j}} = \left(\alpha_{i_j}\right)^{y_{i_j}}$ 是已知的.根据等式(22),若恶意接收者 R_i 能以不可忽略的概率 ε 计算出 m_l,则等价于他能以不可忽略的概率 ε 计算出 $\left(g^{x_{i_j}^*}\right)^{\lambda s_{i_j}}$.

一方面,令 $g' = g^{x_{i_j}^*}$,则 $\left(g^{x_{i_j}^*}\right)^{\lambda s_{i_j}} = \left(g'\right)^{\lambda s_{i_j}}$.若恶意接

收者 R_i 能以不可忽略的概率 ε 计算出 $(g')^{\lambda s_{i_j}}$，则等价于他能以不可忽略的概率 ε 计算出 λs_{i_j}（g 和 $x_{i_j}^*$ 已知，因而 g' 已知）.实际上 λs_{i_j} 秘密隐藏于 $\alpha_{i_j}=h^{\lambda s_{i_j}} \bmod N^2$ 中.也即他能以不可忽略的概率 ε 计算 Partial-DL-ZN2P 难题（由 $\alpha_{i_j}=h^{\lambda s_{i_j}} \bmod N^2$ 计算 $\lambda s_{i_j} \bmod N$）.

另一方面，令 $g=h^x \bmod N^2$，则有 $(g^{x_{i_j}^*})^{\lambda s_{i_j}}=((h^x)^{x_{i_j}^*})^{\lambda s_{i_j}}=((h^{\lambda s_{i_j}})^{x_{i_j}^*})^x=(\alpha_{i_j}^{x_{i_j}^*})^x$（$\alpha_{i_j}$ 和 $x_{i_j}^*$ 已知），继而令 $a=\alpha_{i_j} x_{i_j}^*$（已知），则有 $(g^{x_{i_j}^*})^{\lambda s_{i_j}}=a^x$.因而若恶意接收者 R_i 能以不可忽略的概率 ε 计算出 $(g^{x_{i_j}^*})^{\lambda s_{i_j}}$，等价于他能以不可忽略的概率 ε 计算出 x，也即他能以不可忽略的概率 ε 计算 Partial-DL-ZN2P 难题（由 $g=h^x \bmod N^2$ 计算 $x \bmod N$）.

综上所述，若恶意接收者 R_i 能以不可忽略的概率 ε 计算得到 $m_l(l \neq i_j)$，则他也能以不可忽略的概率 ε 计算 Partial-DL-ZN2P 难题.反过来，根据 Partial-DL-ZN2P 困难假设，接收者 R_i 不能以不可忽略的概率 ε 计算出其它秘密消息 m_l，也即保证了发送者的安全性.

4.3 效率分析

在系统初始化阶段，对于服务发送者，她最主要的开销是生成系统参数，与 Paillier 加密体制的初始化开销一致.

在不经意传输阶段，每发送一个消息，发送者 S 的主要运算为 $n+1$ 次模幂运算，n 次模乘和 n 次模除运算；而接收者 R_i 的主要运算为 k 次 Paillier 加密和 1 次模幂和 1 次模除运算.至于其它计算，例如整数与向量的转换，整数与整数的加法等，基于目前的计算能力均是轻量级的.另外，每发送一个消息，发送者 S 需要发送 $n+1$ 个信息.而每个接收者，在不经意传输阶段仅仅需要发送 k 个信息.

在统计分析阶段，对于服务接收者，最主要的开销是联合进行共享份额的计算和分发.其中单个接收者 R_i 计算共享份额的次数和通信次数均为 $O(i)$.为了节省开销，也可以采用其它更有效的秘密拆分方法[24].对于服务发送者 S，主要运算为 $\sum_{i=1}^{n} d_i$ 次模乘和 1 次模幂，以及 $t+1$ 次 Paillier 解密运算.其它的整数与向量间转换运算也是轻量级的.

由以上分析可知，在系统初始化阶段和不经意传输阶段，新方案的计算与通信开销与主流方案（利用公钥加密算法实现不经意传输）相当.而在统计分析阶段，为了协助发送者计算各子秘密 y_{i_j} 的和，增加了接收者的部分开销（用于子秘密的拆分）.另外，为了统计各秘密消息发送出的总次数，发送者增加了 $\sum_{i=1}^{n} d_i$ 次模乘，1 次模幂，以及 $t+1$ 次 Paillier 解密运算.基于现有的计算和通信技术，这些均是可行的.

尽管增加了一些计算和通信开销，但建议的协议不仅能实现不经意传输，还能实现安全的统计分析.而这样的统计分析在诸如电子投票、电子拍卖等电子商务系统中尤为重要，例如电子投票中的安全计票实质上就是一种安全统计.一般地，统计属性只有针对一定范围一定周期才有意义.例如对于不同的客户群，可能关注的对象或重点不一样，因而需要统计不同类别的秘密被发送出的次数.实际上，建议的方案是一种自适应的可动态调整统计范围和周期的方案.即可以针对不同群体，不同时间周期，分别加以统计，而系统初始化阶段和不经意传输阶段协议不变.

5 结束语

在很多应用中，服务发送方希望能统计每个消息被秘密发送出的次数.但在现有的不经意传输协议中，为了保护服务接收方的隐私，接收方的选择对发送方是不可见的，进而也没有考虑发送方的统计需求.针对此问题，本文提出了一种新的具有统计特性的不经意传输协议.在新的协议中，引入了多方安全求和与同态加密的思想，借助一种较为巧妙的编码方式，不仅能够满足原有不经意传输协议的正确性和安全性，还能够满足发送方的统计需求.尽管扩展了新功能，但其主要阶段（不经意传输阶段），与已有主流协议相比，效率相当.而基于它的统计特性，建议的协议在诸如数据库隐私查询，安全数据挖掘，电子商务，医疗卫生等领域有着更好的应用前景.

参考文献

[1] 张锋，孙雪冬，等.两方参与的隐私保护协同过滤推荐研究[J].电子学报，2009，37(1)：84－89.
Zhang Feng，Sun Xue-dong，et al.Research on privacy-preserving two-party collaborative[J].Acta Electronica Sinica，2009，37(1)：84－89.(in Chinese)

[2] 韩建民，于娟，等.面向敏感值的个性化隐私保护[J].电子学报，2010，38(7)：1723－1728.
Han Jian-min，Yu Juan，et al.Individuation privacy preservation oriented to sensitive values[J].Acta Electronica Sinica，2010，38(7)：1723-1728.(in Chinese)

[3] 刘华玲，郑建国，等.基于贪心扰动的社交网络隐私保护研究[J].电子学报，2013，41(8)：1586－1591.
Liu Hua-ling，Zheng Jian-guo，et al.Privacy preserving in social networks based on greedy perturbation[J].Acta Electronica Sinica，2013，41(8)：1586-1591.(in Chinese)

[4] 王丽娜，彭瑞卿，等.个人移动数据收集中的多维轨迹匿名方法[J].电子学报，2013，41(8)：1653－1659.

Wang Li-na, Peng Rui-qing, et al. Multi-dimensional trajectory anonymity in collecting personal mobility data[J]. Acta Electronica Sinica, 2013, 41(8): 1653 – 1659. (in Chinese)

[5] 刘文,王永滨.安全多方信息比较相等协议及其应用[J].电子学报,2012,40(5):871 – 876.
Liu Wen, Wang Yong-bin. Secure multi-party comparing protocol and its applications[J]. Acta Electronica Sinica, 2012, 40(5): 871 – 876. (in Chinese)

[6] A C Yao. Protocols for secure computation[A]. Proc of 23rd IEEE Symposium on the Foundation of Computer Science[C]. Los Alamitos: IEEE Computer Society Press, 1982. 160 – 164.

[7] O Goldreich. Secure Multi-party Computation [EB/OL]. http://www.wisdom.weizmann.ac.il/ ~ oded /PS/prot.ps. 1998.

[8] 冯涛,马建峰,等.UC 安全的高效不经意传输协议[J].电子学报,2008,36(1):17 – 23.
Feng Tao, Ma Jian-feng, et al. Efficient and universally composable security oblivious transfer[J]. Acta Electronica Sinica, 2008, 36(1): 17 – 23. (in Chinese)

[9] 田有亮,马建峰,等.秘密共享体制的博弈论分析[J].电子学报,2011,39(12):2790 – 2795.
Tian You-liang, Ma Jian-feng, et al. Game-theoretic analysis for the secret sharing scheme[J]. Acta Electronica Sinica, 2011, 39(12): 2790 – 2795 (in Chinese).

[10] 赵春明.可保护授权隐私性的不经意传输[D].西安:西安电子科技大学,2006.
Zhao Chun-ming. Oblivious transfer with privacy-preserving authorization[D]. Xian: Xian Electronics Science and Technology University, 2006. (in chines)

[11] M O Rabin. How to exchange secrets with oblivious transfer [EB/OL]. http://eprint.iacr.org /2005/187, Harvard University Technical Report 81, 1981.

[12] S Even, O Goldreich, et al. A randomized protocol for signing contracts[A]. Proc CRYPTO'82[C]. New York: Plenum Press, 1983. 205 – 210.

[13] G Brassard, C Crépeau, et al. All-or-nothing disclosure of secrets[A]. Advances in Cryptology-Crypto86[C]. Santa Barbara, California, USA: Springer – Verlag, 1987. 234 – 238.

[14] M Naor, B Pinkas. Efficient oblivious transfer protocols[A]. In Proceedings of the 12th Annual ACM/SIAM Symposium on Discrete Algorithms [C]. Washington, DC, USA: ACM/SIAM, 2001. 448 – 457.

[15] H Javier. Restricted adaptive oblivious transfer[J]. Theoretical Computer Science, 2011, 412 (46): 6498 – 6506.

[16] Chang Chin-chen, Lee Jung-san. Robust t-out-of-n oblivious transfer mechanism based on CRT[J]. Journal of Network and Computer Applications, 2009, 32(1): 226 – 235.

[17] Y Lindell, H Zarosim. Adaptivezero-knowledge proofs and adaptively secure oblivious transfer[J]. Journal of Cryptology, 2011, 24(4): 761 – 799.

[18] S Halevi, Y T Kalai. Smooth projective hashing and two-message oblivious transfer[J]. Journal of Cryptology, 2012, 25(1): 158 – 193.

[19] T Tassa. Generalized oblivious transfer by secret sharing[J]. Designs, Codes and Cryptography, 2011, 58(1): 11 – 21.

[20] P Paillier. Public-key cryptosystems based on composite degree residuosity class[A]. Proc of Advances in Cryptology-EUROCRYPTO'99[C]. Prague, Czech Republic: Springer-Verlag, 1999. 223 – 238.

[21] 苏东,王克,等.Paillier 陷门函数的两个变体的比特安全性分析[J].计算机学报,2010,33(6):1050 – 1059.
Su Dong, Wang Ke, et al. The bit security of two variants of paillier trapdoor function[J]. Chinese Journal of Computers, 2010, 33 (6): 1050 – 1059. (in Chinese)

[22] P Y A Ryan. Pret a voter with paillier encryption[J]. Mathematical and Computer Modelling, 2008, 48 (9 – 10): 1646 – 1662.

[23] E Bresson, D Catalano, et al. A simple public key cryptosystem with a double trapdoor decryption mechanism and its application[A]. Proc of Advances in Cryptology-Asiacrypt 2003 [C]. Taipei, Taiwan: Springer-Verlag, 2003. 37 – 54.

[24] 石润华,黄刘生等.新型有效的秘密共享方案[J].通信学报,2012,33(1):10 – 16.
Shi Run-hua, Huang Liu-sheng, et al. Novel and effective secret sharing scheme[J]. Journal on Communications, 2012, 33(1): 10 – 16. (in Chinese)

作者简介

石润华　男,1974 年生于安徽宿松.安徽大学计算机科学与技术学院教授,博士生导师.研究方向为现代密码学,安全多方计算,量子计算,量子密码.
E-mail: shirh@ahu.edu.cn

仲　红　女,1965 年生于安徽固镇.安徽大学计算机科学与技术学院教授,博士生导师.研究方向为网络与信息安全,智能计算.
E-mail: zhongh@ahu.edu.cn