

一个基于 LPN 问题的网络编码同态 MAC 加密方案

梁 满

(上海市信息安全测评认证中心 上海 200011)

摘要 基于网络编码的应用系统很容易遭受到污染攻击。在目前解决污染攻击的对称密钥方法中,中间节点往往不具备防御能力。针对以上问题,提出一个基于 LPN 问题难度的 HB 协议网络编码同态 MAC 加密方案,并在基本模型下证明方案的安全性。与以往的基于对称密钥的方法相比,该方案允许网络中间节点验证所收到数据包的合法性,可以尽早地发现并过滤掉被污染的数据包。同时,由于具有较低的计算开销和带宽开销,该方案非常适用于实时性较强的网络编码应用。

关键词 网络编码 污染攻击 LPN 问题 同态消息鉴别码

中图分类号 TN925.93 TP393.08 **文献标识码** A **DOI**:10.3969/j.issn.1000-386x.2019.01.054

A HOMOMORPHIC MAC ENCRYPTION SCHEME FOR NETWORK CODING BASED ON LPN

Liang Man

(Shanghai Information Security Testing Evaluation and Certification Center, Shanghai 200011, China)

Abstract Application system based on network coding is vulnerable to pollution attacks. In the current symmetric encryption to solve pollution attacks, intermediate nodes usually do not have the capability to defend the attacks. To solve this problem, we proposed a homomorphic MAC encryption scheme for network coding with HB protocol based on LPN problem, and proved that the proposed scheme was secure under the basic model. Compared with the previous schemes based on symmetric encryption, the proposed scheme allowed intermediate nodes in the network to verify the legality of the received packets, and detected and filtered the contaminated packets as early as possible. Duo to the lower computational overhead and width overhead, our scheme is ideally suitable for some real-time network coding applications.

Keywords Network coding Pollution attack LPN Homomorphic MAC

0 引 言

网络编码为传统的基于“存储-转发”机制的通信网络的数据传输方式带来了一种新的替代方案^[1]。然而,基于网络编码的应用系统很容易遭受到污染攻击^[2-6]。在这种攻击中,网络中的一些恶意节点将一些非法数据包注入通信网络,由于网络编码中间节点有组合数据包的特性,少量的被污染的数据包会造成大规模的污染扩散,导致接收节点无法正常解码。

为了解决网络编码的污染攻击问题,研究者陆续提出了许多基于密码学方法的解决方案^[7-9]。这些解决方案主张使用同态哈希或者同态签名来阻止污染攻

击,允许网络的中间节点验证所接收到数据包的合法性,可以尽早地发现并过滤掉被污染的数据包,属于主动的防御方法。然而,在这些解决方案中,签名的生成和验证都需要使用计算耗时的双线性群配对操作或者模幂运算,对于在线实时的网络编码应用而言相对较慢。这些签名方案的另一个缺点是将随机网络编码^[10]经常使用的一个较小的有限域(8 bit)替换成了一个非常大的适合密码构造使用的有限域(160 bit),这无疑增加了中间节点的计算开销和带宽开销^[11]。

为了解决网络编码签名方案的效率问题,研究者提出了许多基于对称密钥的方法来抵抗污染攻击^[11-12]。这些方法要求源节点与接收节点共享一个密钥,用于为数据包生成合法的同态消息鉴别码(MAC)。中间

节点利用同态的性质可以为数据包生成合法的 MAC 标签,而接收节点通过验证数据包 MAC 标签的合法性来过滤掉非法的数据包。这些基于对称密钥的方案相对于签名方案而言有着更快的验证效率,非常适合实时性较强的在线网络编码应用。然而,在基于对称密钥方法的解决方案中,网络的中间节点往往不具备检测数据包的能力,导致这些方案并不能尽早地发现并阻击污染攻击的进一步扩大,极大地限制了这些解决方案的实用性^[11]。

为了解决对称密钥加密方案的中间节点不能验证数据包合法性的问题,本文提出了一个基于 LPN 问题^[13]难度的 HB 协议^[14]网络编码同态 MAC 加密方案。与以往的基于对称密钥的方法相比,本文所提出的解决方案允许网络中间节点验证所收到数据包的合法性,可以尽早地发现并过滤掉被污染的数据包,以阻击污染攻击的进一步扩大。此外,由于中间节点的验证过程只需要进行有限域的内积和异或操作,并不需要进行昂贵双线性群配对操作或者模幂运算。与以往的网络编码签名方案相比,只需要使用的一个较小的有限域(8 bit),因此具有更低的计算开销和带宽开销,非常适用于实时性较强的网络编码应用。

1 问题陈述

1.1 符号定义

在本文中,使用 \mathbb{Z}_q 表示集合 $\{0, 1, \dots, q-1\}$, 即规定加法总是模 q 的,其中 $q = 2^r$ 为 2 的幂次。特别地,全用 $\mathbb{Z}_2 = \{0, 1\}$ 表示位,同时规定这也是一个有限域。我们用 \oplus 表示按位异或操作。用小写黑体 x 表示向量,大写黑体字母 X 表示矩阵,花体字母 \mathcal{X} 表示集合。对于一个集合 \mathcal{X} 而言,使用 $x \xleftarrow{R} \mathcal{X}$ 表示 x 是从 \mathcal{X} 中均匀一致选取的。对于一个分布 D 来说,使用 $x \leftarrow D$ 则表示 x 从分布 D 中取样的。定义 Ber_τ 表示参数为 τ 的伯努利分布,也就是说 $Pr[x = 1; x \leftarrow Ber_\tau] = \tau$ 。对于 $m \in \mathbb{N}$,使用 U_m 表示 \mathbb{Z}_2^m 上的一致分布。定义 $X \sim D$ 表示 X 是一个服从分布 D 的随机变量。使用 $\langle a, b \rangle = \sum_{i=1}^n a[i] \cdot b[i] \bmod P$ 表示 $a, b \in \mathbb{Z}_p^n$ 的内积。定义 $HW(e)$ 是 e 的汉明重量,也就是在 e 中比特为“1”的个数。

1.2 网络模型

我们假设一个使用随机线性网络编码进行网络多播的网络模型^[10]。在这个模型中,网络中的节点被划分为 3 种角色:一个源节点,一些中间节点和一些接收

节点。源节点通过中间节点发送一些数据包给接收节点。为了实现这个目的,源节点把需要发送的数据包划分为很多“代”^[9],每一代都包含 m 个数据包。源节点把每一代中的 m 个数据包都表示为有限域 \mathbb{F}_q 上的一个 n 维的向量空间中的 m 个向量 $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_m$ 。其中, $q = 2^r$ 为 2 的幂次,以方便在中间节点和源节点之间应用 HB 协议^[14]。为了方便接收节点解码,源节点将这 m 个向量充成 m 个增广向量,它们具有如下形式:

$$v_i = (\underbrace{0, \dots, 0}_{i-1}, \underbrace{1, 0, \dots, 0}_{m-i}, \bar{v}_i) \in \mathbb{F}_q^{n+m}$$

也就是说,增广向量 v_i 的前 m 个系数构成一个单位向量,而这些单位向量的第 i 个位置为“1”,其他的位置均为“0”。如果我们把一个向量 $y \in \mathbb{F}_q^{n+m}$ 看作是增广向量 v_1, v_2, \dots, v_m 的一个线性的组合,那么向量 y 的前 m 个位置正好就是这个线性组合使用的系数。源节点以一代接一代的形式将所有数据包发送进网络中。

网络中的每个中间节点都进行随机线性网络编码操作^[10]。中间节点只编码那些属于同一代的数据包,而不会将属于不同代的数据包编码到一起。举例来说,我们假设一个中间节点从它的入边接收到了第 k 代的 j 个包 p_1, p_2, \dots, p_j 。接下来这个节点会向它下游的每个节点发送一个数据包 $y = \sum_{i=1}^j c_i p_i$, 其中每个系数 $c_i \in \mathbb{F}_q$ 都是从有限域 \mathbb{F}_q 里随机均匀地选取的。假如网络中的传输过程没有出错的话,那么所有网络中发送的包实际上都是 m 个增广向量 v_1, v_2, \dots, v_m 的线性组合。当网络中的接收节点收到第 k 代的 m 个线性无关的增广向量(数据包)以后,就可以对一个 $m \times (m+n)$ 的矩阵进行高斯消元来解码第 k 代的原始数据包 $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_m$ 。

我们在建立基于网络编码多播的网络模型的时候,假定所有的网络节点都是合法的且不存在攻击者的情况,但实际上网络中存在着潜在的攻击者。

1.3 攻击者模型

为了不失一般性,在建立攻击者模型时假定源节点和接收节点是可以信赖的,而所有的中间节点有可能是潜在的攻击者。举例来说,攻击者可能伪造一些数据包,并把这些伪造好的数据包注入到传输网络中,目的在于阻止接收节点解码出原始数据。攻击者也可能篡改流经某些网络节点的数据包,例如攻击者可能篡改一个合法数据包所携带的消息鉴别码或者签名,又或者收集前一代传输中的合法数据包,并将这些数据包伪装成下一代的合法数据包,注入到下一代的数

据传输之中等。我们正式给出被污染的数据包的定义:

定义 1 假定用 V 表示某一代的 m 个源节点增广向量 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ 张成的线性空间, 那么一个向量 (或者数据包) $\mathbf{y} = (y_1, y_2, \dots, y_m, y_{m+1}, \dots, y_{m+n})$ 相对于线性空间 V 是一个被污染的数据包, 如果存在 $\mathbf{y} \notin V$ 或者等价地说:

$$\mathbf{y} \neq \sum_{i=1}^m y_i \mathbf{v}_i \Leftrightarrow \mathbf{y} \notin V$$

式中: 元素 y_i 是向量 \mathbf{y} 的前 m 坐标点。

当建立攻击者模型的时候, 本文假设攻击者了解本文提出的方案构造, 并且有能力执行多项式时间算法 (PPT) 来实施攻击行为, 但是本文不对攻击者如何使用其能力和策略进行假设^[15]。

2 基于 LPN 的同态 MAC 加密方案定义

假设 $V = \text{Span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m)$ 表示由 m 个源节点增广向量 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ 所张成的线性空间。本文提出的基于 LPN 的同态 MAC 加密方案要求每个源节点线性空间 V 都必须有唯一的向量空间标识符 id , 这个标识符是从一个集合 $id \in \mathcal{I}$ 中随机均匀选取的一个不重复的元素。源节点随机地从密钥空间 \mathcal{K} 中选择两个密钥 $k_1 \xleftarrow{R} \mathcal{K}$ 和 $k_2 \xleftarrow{R} \mathcal{K}$ 用于生成和验证 MAC 标签, 并将两个密钥通过安全通道共享给所有的接收节点。源节点为每个增广向量 \mathbf{v}_i 计算一个 MAC 标签 t_i 并将这些数据包以三元组的形式 (id, \mathbf{v}_i, t_i) 发送到网络之中。当网络的中间节点收到数据包后, 可以利用 MAC 标签同态的性质为数据包生成合法的 MAC 标签, 并利用 HB 协议^[14] 验证数据包的合法性。接收节点使用密钥 k_1 和 k_2 为数据包生成合法 MAC 标签, 并通过与数据包携带的 MAC 标签相比较来验证数据包的合法性。

本文使用 \mathcal{I} 表示空间标识符的集合, 使用 \mathcal{K} 表示密钥空间的集合, 正式给出基于 LPN 问题的同态 MAC 加密方案的定义。

2.1 方案的定义

定义 2 一个参数为 (q, n, m, r, l) (其中 $q = 2^r$, 且 $r, l \in \mathbb{N}$ 为正整数) 的基于 LPN 问题的网络编码同态 MAC 加密方案可以被定义为一个概率多项式时间算法构成的五元组 (Generate, Sign, Combine, Verify_1, Verify_2), 并且满足以下条件:

系统初始化算法 Generate (id, V):

输入: 一个空间标识符 id , 一个向量空间 V 。

输出: 两个密钥 k_1 和 k_2 , 一个矩阵 $A \in \mathbb{Z}_2^{l \times r(n+1)}$ 。

签名算法 Sign ($id, k_1, k_2, \mathbf{v}, i$):

输入: 一个空间标识符 id , 两个密钥 k_1 和 k_2 , 一个源节点增广向量 $\mathbf{v} \in \mathbb{F}_q^{m+n}$, 其中 $i = 1, 2, \dots, m$ 表示向量 \mathbf{v} 在源节点线性空间基中的位置。

输出: 一个向量 \mathbf{v} 的 MAC 标签 $t \in \mathbb{F}_q$ 。

组合算法 Combine ($id, (y_1, t_1, c_m), (y_2, t_2, c_m), \dots, (y_m, t_r, c_m)$):

输入: 一个向量空间标识符 id , m 个向量 $y_1, y_2, \dots, y_m \in \mathbb{F}_q^{m+n}$, 和 m 向量对应的 MAC 标签 $t_1, t_2, \dots, t_m \in \mathbb{F}_q$, 以及 m 个编码系数 $c_1, c_2, \dots, c_m \in \mathbb{F}_q$ 。

输出: 一个编码后向量 $\mathbf{y} = \sum_{i=1}^m c_i y_i$ 和编码后向量 \mathbf{y} 对应的 MAC 标签 $t \in \mathbb{F}_q$ 。

验证算法 1 Verify_1 ($id, k_1, k_2, \mathbf{y}, t$):

输入: 一个向量空间标识符 id , 两个密钥 k_1 和 k_2 , 一个向量 $\mathbf{y} \in \mathbb{F}_q^{m+n}$ 以及向量 \mathbf{y} 对应的 MAC 标签 $t \in \mathbb{F}_q$ 。

输出: “1” (表示接受) 或者 “0” (表示拒绝)。

验证算法 2 Verify_2 (id, \mathbf{y}, A, t):

输入: 一个空间标识符 id , 一个向量 $\mathbf{y} \in \mathbb{F}_q^{m+n}$ 及其 MAC 标签 $t \in \mathbb{F}_q$, 矩阵 A 。

输出: “1” (表示接受) 或者 “0” (表示拒绝)。

2.2 正确性条件

本文提出的基于 LPN 问题的同态 MAC 加密方案必须满足如下正确性条件, 即要求对于所有由系统的初始化算法 Generate (id, V) 输出的 (k_1, k_2, A) , 下面的正确性条件必须满足:

假设用 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in \mathbb{F}_q^{m+n}$ 表示 m 个源节点增广向量, 并且对于所有的 $1 \leq i \leq m$, 如果存在 $t_i \leftarrow \text{Sign}(id, k_1, k_2, \mathbf{v}_i, i)$, 其中 i 表示向量在源节点线性空间基中的位置。同时, 对于所有的 $1 \leq i \leq m$ 和编码系数 $c_1, c_2, \dots, c_m \in \mathbb{F}_q$, 如果存在:

$t \leftarrow \text{Combine}(id, (\mathbf{v}_1, t_1, c_1), (\mathbf{v}_2, t_2, c_2), \dots, (\mathbf{v}_m, t_m, c_m))$

那么, 下面两个正确性条件必须满足:

Verify_1 ($id, k_1, k_2, \sum_{i=1}^m c_i \mathbf{v}_i, \text{Combine}(id, (\mathbf{v}_1, t_1, c_1), \dots, (\mathbf{v}_m, t_m, c_m))$) = 1

Verify_2 ($id, \sum_{i=1}^m c_i \mathbf{v}_i, A, \text{Combine}(id, (\mathbf{v}_1, t_1, c_1), \dots, (\mathbf{v}_m, t_m, c_m))$) = 1, 且验证失败的概率可以忽略。

2.3 安全性条件

定义安全性条件时, 我们允许攻击者可以获得任

意向量和其对应的 MAC 签名^[15]。同时,我们假定每个攻击者提交的线性空间 V_g 都有一个对应的向量空间标识符 id_g 。基于 LPN 问题的同态 MAC 加密方案的安全性定义如下:

定义 3 假设 $\mathcal{T} = (\text{Generate}, \text{Sign}, \text{Combine}, \text{Verify}_1, \text{Verify}_2)$ 是一个基于 LPN 问题的同态 MAC 加密方案,则我们使用一个攻击者游戏来定义方案 \mathcal{T} 的安全性。攻击者游戏有两个参与者,使用 \mathcal{A} 表示攻击者,使用 \mathcal{C} 表示挑战者。如果多项式时间攻击者 \mathcal{A} 在这个游戏中胜利的概率是可以忽略的,则我们说这个基于 LPN 问题的同态 MAC 加密方案 \mathcal{T} 是安全的。攻击者游戏的定义如下:

系统初始化阶段:挑战者 \mathcal{C} 随机地从密钥空间的集合 \mathcal{K} 中选择两个密钥 k_1 和 k_2 。

查询阶段:攻击者 \mathcal{A} 适应性地向挑战者 \mathcal{C} 提交查询的请求,且每个查询的形式为 (id_g, V_g) 。其中 id_g 表示一个向量空间标识符, V_g 表示 m 个基 $v_1, v_2, \dots, v_m \in \mathbb{F}_q^{m+n}$ 所张成的线性空间,攻击者 \mathcal{A} 每次提交的查询请求的向量空间标识符 id_g 均不同。当挑战者 \mathcal{C} 收到攻击者 \mathcal{A} 的查询请求后,计算 $t_i \leftarrow \text{Sign}(id_g, \text{Generate}(id_g, k, V_g), v_i, i)$, 然后将 MAC 标签 (t_1, t_2, \dots, t_m) 发还给攻击者 \mathcal{A} 。

输出阶段:攻击者 \mathcal{A} 向挑战者 \mathcal{C} 提交多次查询的请求后,攻击者 \mathcal{A} 输出一个三元组 (id_*, y_*, t_*) , 如果存在算法 $\text{Verify}_1(id_*, k_1, k_2, y_*, t_*) = 1$ 或者算法 $\text{Verify}_2(id_*, y_*, \mathcal{A}, t_*) = 1$, 且下面两个条件之一成立:(1) 对于所有 g , 存在 $id_* \neq id_g$, 并且 $y_* \neq 0$ (称为类型 1 伪造);(2) 对于某些 g , 存在 $id_* = id_g$, 并且 $y_* \notin V_g$ (称为类型 2 伪造)。那么,我们认为攻击者 \mathcal{A} 赢得了攻击者游戏。

为了叙述方便,我们使用 $\text{NC-Adv}[\mathcal{A}, \mathcal{T}]$ 表示攻击者 \mathcal{A} 在攻击者游戏中使得算法 $\text{Verify}_1(id_*, k_1, k_2, y_*, t_*) = 1$ 的概率,使用 $\text{HB-Adv}[\mathcal{A}, \mathcal{T}]$ 表示攻击者 \mathcal{A} 在攻击者游戏中使得算法 $\text{Verify}_2(id_*, y_*, \mathcal{A}, t_*) = 1$ 的概率。

3 LPN 问题和 HB 协议

本文提出同态 MAC 加密方案将依赖于基于 LPN 问题^[13]难度的 HB 协议^[14]。

3.1 LPN 问题

定义 4 (搜索/判定 LPN 问题) 对于 $\tau \in [0, 1/2]$, $l \in \mathbb{N}$, 我们定义一个判定 LPN 问题 $LPN_{\tau, l}$ 是 (q, t, ε) -难的, 如果对任意运行时间为 t 的区分者 D

都满足如下等式:

$$|Pr_{s, A, e}[D(A, As \oplus e) = 1] - Pr_{r, A}[D(A, r) = 1]| \leq \varepsilon$$

式中: $s \xleftarrow{R} \mathbb{Z}_2^l$, $A \xleftarrow{R} \mathbb{Z}_2^{q \times l}$, $e \leftarrow \text{Ber}_\tau^q$, $r \xleftarrow{R} \mathbb{Z}_2^q$ 。我们定义一个搜索 LPN 问题是 (q, t, ε) -难的, 如果对所有时间 t 内的区分者 D 满足:

$$Pr_{s, A, e}[D(A, As \oplus e) = s] \leq \varepsilon$$

LPN 问题中的判定和搜索版本是多项式等价的^[16-17]。对于 LPN 判定问题攻击时间为 t 的攻击者对于搜索版本 LPN 的攻击事件为 $\text{poly}(t)$ 。因此, 当搜索 LPN 问题为困难的时候, 基于判定 LPN 问题的密码系统就是安全的。虽然由搜索问题归约到判定问题并不是一个紧的归约, 但是实际上并没有比解决搜索 LPN 问题的算法更快的解决判定 LPN 问题的算法^[17]。

定理 1 (文献[17]中引理 1) 如果判定 LPN 问题 $LPN_{\tau, l}$ 不是 (q, t, ε) -安全的, 那么搜索 LPN 问题 $LPN_{\tau, l}$ 不是 $O(q', t', \varepsilon')$ -安全的, 其中:

$$\begin{aligned} q' &= O(q \cdot \log_2 l / \varepsilon^2), \\ t' &= O(t \cdot \log_2 l / \varepsilon^2), \\ \varepsilon' &= (\varepsilon/4) \end{aligned}$$

目前的复杂性研究无法证明不存在有效地攻击 LPN 问题的算法。搜索 LPN 问题可以解释为解码随机线性码的问题, 这是一个 NP 问题^[18]。假设 A 是生成矩阵, s 是要传输的消息。解码问题要求由有噪声的码字 $A \cdot s \oplus e$ 恢复 s , 这就是搜索 LPN 问题。目前已知最好的恢复 l 位消息的算法需要规模为 $2^{\Theta(l/\log_2 l)}$ 的时间和样本数^[13]。如果只有多项式个样本 $q = \text{poly}(l)$, 那么最好算法的运行时间为 $2^{\Theta(l/\log_2 \log_2 l)}$ ^[19]。如果只有线性个样本 $q = \Theta(l)$, 那么最好的算法为指数时间 $2^{\Theta(l)}$ ^[20]。到目前为止, 并没有针对 LPN 问题的快速量子算法。

3.2 HP 协议

HB 协议是 Hopper 和 Blum 在 2001 年提出了第一个基于 LPN 问题的认证方案^[14]。如图 1 所示, 密钥 $s \in \mathbb{Z}_2^l$, 其中 l 的选取需要保证 $LPN_{\tau, l}$ 问题是难的。

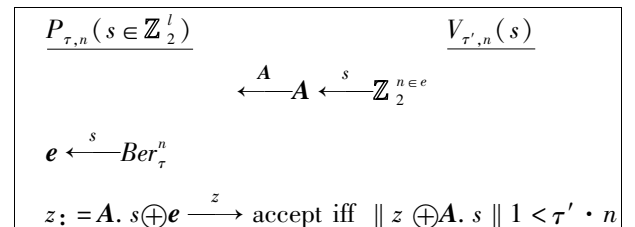


图 1 HB 协议示意图

验证者 V 首先发送一个挑战矩阵 $A \in \mathbb{Z}_2^{n \times l}$ (n 是一个统计安全参数)。证明者 P 向验证者 V 发送一个 LPN 样本 $A \cdot s \oplus e$ 。如果证明者 P 的回答 y 是 $y =$

$A \cdot s \oplus e$ 的形式,并且 e 是一个低权重的向量,那么 V 则接受 P 。一个正确生成的 e 的权重为 $n\tau$,接受的权重阈值可以设定为 $n\tau'$,其中 $\tau' < \tau < 1/2$ 。比如在 $(1/2)\tau \leq \tau' \leq (3/2)n\tau$ 这种情况下,一个正确生成的 $e \leftarrow \text{Ber}_\tau^n$ 的权重应该 $\geq n\tau'$,验证者 V 拒绝一个证明者 P 的概率对于 n 是指数级小的。HB 协议在 LPN 问题是难的这个假设下是安全的。HB 协议具有如下 3 个特点:

(1) 在 HB 协议的证明者 P 和验证者 V 交互的过程中,如果存在一个被动攻击者窃听他们的交互,那么被动攻击者不能恢复 s 。这是因为如果被动攻击者可以恢复 s 意味着攻击者可以解决 LPN 问题。

(2) 证明者 P 正确地回答出 $A \cdot s \oplus e$,但是由于 $HW(e)$ 不在 $[(1/2)\tau n, (3/2)\tau n]$ 范围内,因此验证出错的概率为:

$$Pr_{e \sim \text{Ber}_\tau^n} \left[|HW(e) - n\tau| > \frac{n\tau}{2} \right] < 2^{-\frac{m^2}{4}}$$

因此,当 $n = \omega(\log_2 l)$ 足够大时,上式出错概率可以忽略。

(3) 证明者 P 正确地回答了 $As' \oplus e'$ 满足 $As' \oplus e' = As \oplus e$ 且 $(s', e') \neq (s, e)$ 的概率。事实上,当 $n = \Omega(l)$ 时,对于任意满足 $(1/2)\tau n < HW(e) < (3/2)\tau n$ 的 e 有:

$$Pr_{A \sim U_m} \left[\exists e' \neq e : (1/2)\tau n \leq HW(e') \leq (3/2)\tau n \wedge As' \oplus e' = As \oplus e \right] < 2^{-\Omega(l)}$$

这是因为首先攻击者无法获得 s 或者 e 的值,并且这两个数值都是随机选取的,也就是说,在多项式时间内, $As \oplus e$ 和随机的 l 维向量是无法区分的。因此,随机取另一个 $As' \oplus e' = As \oplus e$ 的概率为 $1/2^n = 1/2^{\Omega(l)}$ 。这意味着当证明者 P 没有密钥 s 而想说服验证者 V 它有密钥 s 时,成功的概率为 $2^{-\Omega(l)}$,可以忽略不计^[17]。

4 方案的构造

本文所提出的加密方案的构造将使用一个伪随机生成器 (PRG) 和一个伪随机函数 (PRF)。如果假定 G 为一个伪随机生成器,并且满足 $G: \mathcal{K}_G \rightarrow \mathbb{F}_q^{m+n}$,假定 F 为一个伪随机函数,并且满足 $\mathcal{K}_F \times (\mathcal{I} \times [m]) \rightarrow \mathbb{F}_q$,那么给出加密方案的构造如下:

系统初始化算法 $\text{Generate}(id, V)$:

算法输入一个空间标识符 $id \in \mathcal{I}$ 和一个向量空间 V ,算法选择 $k_1 \xleftarrow{R} \mathcal{K}_G$, $k_2 \xleftarrow{R} \mathcal{K}_G$, $A \xleftarrow{R} \mathbb{Z}_2^{l \times r(n+1)}$ 。算法输出两个密钥 k_1 和 k_2 , 矩阵 $A \in \mathbb{Z}_2^{l \times r(n+1)}$, 矩阵 A

是一个公共参数,将被发送给网络中所有节点。

签名算法 $\text{Sign}(id, k_1, k_2, v, i)$:

算法输入一个空间标识符 $id \in \mathcal{I}$, 一对密钥 (k_1, k_2) , 一个源节点增广向量 $v \in \mathbb{F}_q^{m+n}$, 标识向量 v 在线性空间基中的位置 i , 算法计算:

$$\begin{aligned} u &\leftarrow G(k_1) \in \mathbb{F}_q^{m+n} \\ b &\leftarrow F(k_2, id, i) \in \mathbb{F}_q \\ t &\leftarrow (u \cdot v) + b \in \mathbb{F}_q \end{aligned}$$

算法输出向量 v 的 MAC 标签 $t \in \mathbb{F}_q$ 。

组合算法 $\text{Combine}(id, (y_1, t_1, c_1), (y_2, t_2, c_2), \dots, (y_m, t_m, c_m))$:

算法输入一个空间标识符 $id \in \mathcal{I}$, m 个向量 $y_1, y_2, \dots, y_m \in \mathbb{F}_q^{m+n}$, m 个向量对应的 MAC 标签 $t_1, t_2, \dots, t_m \in \mathbb{F}_q$, 以及 m 个编码系数 $c_1, c_2, \dots, c_m \in \mathbb{F}_q$, 算法计算 $y = \sum_{i=1}^m c_i y_i \in \mathbb{F}_q^{m+n}$ 以及向量 y 对应的 MAC 标

签 $t = \sum_{i=1}^m c_i t_i \in \mathbb{F}_q$ 。

算法输出向量 y 以及 y 对应的 MAC 标签 t 。

验证算法 1 $\text{Verify}_1(id, k_1, k_2, y, t)$:

此算法为接收节点的验证算法: 算法输入一个空间标识符 $id \in \mathcal{I}$, 一对密钥 (k_1, k_2) , 一个向量 $y \in \mathbb{F}_q^{m+n}$ 和其对应的 MAC 标签 $t \in \mathbb{F}_q$, 算法计算:

$$\begin{aligned} u &\leftarrow G(k_1) \in \mathbb{F}_q^{m+n} \\ b &\leftarrow \sum_{i=1}^m [y_i F(k_2, id, i)] \in \mathbb{F}_q \\ t' &= a + b \in \mathbb{F}_q \end{aligned}$$

如果存在 $t' = t$, 验证算法一输出“1”表示接受; 否则输出“0”表示拒绝。

验证算法 2 $\text{Verify}_2(id, y, A, t)$:

此算法为中间节点的验证算法, 算法输入一个空间标识符 $id \in \mathcal{I}$, 一个向量 $y \in \mathbb{F}_q^{m+n}$ 及其对应的 MAC 标签 $t \in \mathbb{F}_q$, 以及矩阵 A , 算法进行如下操作:

(1) 首先随机选择一个 $e \leftarrow \text{Ber}_\tau^l$;

(2) 中间节点将 y 的前 m 个分量 (即编码系数) 传输给源节点。因为向量 $y \in \mathbb{F}_q^{m+n}$, 而且有 $q = 2^r$, 所以可以将 y 写成 \mathbb{Z}_2 上的 $r(m+n)$ 维向量, 而 t 则可以写成 r 维向量。算法将 y 的系数部分以外的部分和 t 放在一起记为 $s \in \mathbb{Z}_2^{r(n+1)}$, 则中间节点计算 $z = A \cdot s \oplus e \in \mathbb{Z}_2^l$ 并传输给源节点;

(3) 源节点利用收到的系数计算得出 $y' = \sum_{i=1}^m c_i v_i$ 和 $t' = \sum_{i=1}^m c_i t_i$ 。将 y' 的系数部分以外的部分和 t' 放在一起, 记为 $s' \in \mathbb{Z}_2^{r(n+1)}$ 并计算 $z \oplus A \cdot s' \in \mathbb{Z}_2^l$ 。如果存在 $HW(e) \in [(1/2)\tau l, (3/2)\tau l]$, 验证算

法二输出“1”表示接受;否则输出“0”表示拒绝。

5 方案的正确性证明

我们将证明本文提出的基于 LPN 问题的网络编码同态 MAC 加密方案构造满足正确性条件。为了验证方案的正确性,我们假定一个向量 $\mathbf{y} = \sum_{i=1}^m c_i \mathbf{v}_i$, 其中 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in \mathbb{F}_q^{m+n}$ 是 m 个源节点增广向量,而 $c_1, c_2, \dots, c_m \in \mathbb{F}_q$ 为 m 个编码系数。向量 \mathbf{y} 对应的 MAC 标签由组合算法计算可得 $t = \sum_{i=1}^m c_i t_i$, 其中 t_1, t_2, \dots, t_m 是 m 个源节点增广向量的 MAC 标签。则由验证算法 Verify_1 计算得到一个 MAC 标签 t' 如下:

$$t' = G(k_1) \cdot \mathbf{y} + \sum_{i=1}^m [y_i F(k_2, id, i)] =$$

$$G(k_1) \cdot \sum_{i=1}^m y_i \mathbf{v}_i + \sum_{i=1}^m [y_i F(k_2, id, i)] = t$$

即验证算法 1 输出的 MAC 标签 t' 与组合算法输出的 MAC 标签 t 相等。同理可证 Verify_2 也是正确的。如果源节点和网络中间节点分别计算 s' 和 s , 其不相等的概率小于 $2^{-\frac{lt^2}{4}}$ 。因此,当 l 足够大的时候,算法 Verify_2 验证失败的概率可以忽略。因此,本文提出的加密方案满足正确性条件。

6 方案的安全性证明

为了证明方案的安全性,我们在假设 G 是一个安全的伪随机生成器(PRG),而 F 是一个安全的伪随机函数(PRF)。对于一个 PRF 攻击者 \mathcal{B}_1 ,使用 PRF-Adv $[\mathcal{B}_1, F]$ 表示攻击者 \mathcal{B}_1 赢得一个 PRF 游戏的概率。同样地,对于一个 PRG 攻击者 \mathcal{B}_2 ,使用 PRG-Adv $[\mathcal{B}_2, G]$ 表示攻击者 \mathcal{B}_2 赢得一个 PRG 游戏的概率。

定理 2 假设 $\mathcal{T} = (\text{Generate}, \text{Sign}, \text{Combine}, \text{Verify}_1, \text{Verify}_2)$ 是一个基于 LPN 问题的同态 MAC 加密方案。对于给定的参数 q, n, m, r, l , 只要 G 是一个安全的 PRG 以及 F 是一个安全的 PRF,那么方案在 l 足够大的情况下(如 $l = \omega(\log_2 r(n+1))$)就是安全的。对于每一个该方案 \mathcal{T} 的攻击者 \mathcal{A} 来说,都存在一个 PRF 攻击者 \mathcal{B}_1 以及 PRG 攻击者 \mathcal{B}_2 与攻击者 \mathcal{A} 具有相同的运行时间,且满足:

$$\text{NC} - \text{Adv}[\mathcal{A}, \mathcal{T}] \leq \text{PRF} - \text{Adv}[\mathcal{B}_1, F] +$$

$$\text{PRG} - \text{Adv}[\mathcal{B}_2, G] + (1/q)$$

并且:

$$\text{HB} - \text{Adv}[\mathcal{A}, \mathcal{T}] < \text{NC} - \text{Adv}[\mathcal{A}, \mathcal{T}]$$

证明:证明将使用三个游戏来进行。对于 $i = 0, 1, 2$, 我们使用 W_i 表示攻击者 \mathcal{A} 在游戏 i 中赢得安全游戏的事件,则有:

$$\Pr[W_0] = \text{NC} - \text{Adv}[\mathcal{A}, \mathcal{T}] \quad (1)$$

在游戏 1 中,伪随机生成器 G 被一个真随机字符串代替。也就是说,游戏 1 和游戏 0 完全相同,除了在响应 MAC 查询时,挑战者 C 在 Sign 阶段计算 $u \leftarrow \mathbb{F}_q^{m+n}$ 而不是计算 $u \leftarrow G(k_1)$, 那么存在一个 PRG 攻击者 \mathcal{B}_2 满足:

$$|\Pr[W_0] - \Pr[W_1]| = \text{PRG} - \text{Adv}[\mathcal{B}_2, G] \quad (2)$$

在游戏 2 中,伪随机函数 F 被一个真随机字符串代替。也就是说,游戏 2 和游戏 1 完全相同,除了在响应 MAC 查询时挑战者 C 在 Sign 阶段计算 $b \leftarrow \mathbb{F}_q$ 而不是计算 $b \leftarrow F(k_2, (id_i, j))$, 那么存在一个 PRF 攻击者 \mathcal{B}_1 满足:

$$|\Pr[W_1] - \Pr[W_2]| = \text{PRF} - \text{Adv}[\mathcal{B}_1, F] \quad (3)$$

在游戏 2 中,挑战者 C 做如下工作:

系统初始化阶段: $u \leftarrow \mathbb{F}_q^{n+m}$

查询阶段:攻击者 \mathcal{A} 提交查询请求 (V_i, id_i) , 其中,

$V_i = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) \subseteq \mathbb{F}_q^{n+m}$ 为 m 个源节点向量所张成的线性空间; id_i 为线性空间所对应的标识。对于每 $i = 1, 2, \dots, m$ 个查询,挑战者都计算 $b_{i,j} \xleftarrow{R} \mathbb{F}_q$ 和 $t_{i,j} \leftarrow (u \cdot \mathbf{v}_j) + b_{i,j} \mathbb{F}_q$, 并发送 $(t_{i,1}, t_{i,2}, \dots, t_{i,m})$ 给攻击者 \mathcal{A} 。最终攻击者 \mathcal{A} 输出一个三元组 (id^*, t^*, y^*) 。

为了确定攻击者 \mathcal{A} 是否赢得了游戏,我们计算:

如果 $id^* = id$, 那么 $(b_1^*, b_2^*, \dots, b_m^*) \leftarrow (b_{i,1}, b_{i,2}, \dots, b_{i,m})$ (称为类型 1 伪造)。

如果 $id^* \neq id$, 对于 $i = 1, 2, \dots, m$ 令 $b_j^* \xleftarrow{R} \mathbb{F}_q$ (称为类型 2 伪造)。

假设 $y^* = (y_1^*, y_2^*, \dots, y_{n+m}^*)$ 。我们说攻击者赢得游戏胜利,如果存在:

$$t^* = (u \cdot y^*) + \sum_{j=1}^m (y_j^* \cdot b_j^*) \quad (4)$$

对于类型 2 伪造来说 $y^* \notin V_i$, 并且 $(y_1^*, y_2^*, \dots, y_m^*)$ 非全零的。

接下来我们需要证明在游戏 2 中攻击者 \mathcal{A} 赢得安全游戏的概率为 $\Pr[W_2] = 1/q$ 。我们使用 T 来表示攻击者 \mathcal{A} 输出类型 1 伪造这一事件。

当类型 1 伪造发生(即事件 T 发生)时,我们需要限制 $\Pr[W_2 \wedge T]$ 的大小。在类型 1 伪造中,式的右边是一个独立于攻击者 \mathcal{A} 视图的 \mathbb{F}_q 中的随机变量。因此,当事件 T 发生的时候,式(4)成立的概率为 $1/q$, 所以有 $\Pr[W_2 \wedge T] = (1/q) \cdot \Pr[T]$ 。

类型2伪造发生(即事件T不发生)时,我们需要限制T发生时攻击者 \mathcal{A} 赢得游戏2胜利的概率。在一个类型2伪造中,攻击者 \mathcal{A} 使用一个在以前的MAC查询中使用过的向量空间标识 id^* ,那么对某个 i 来说 $id^* = id_i$ 。事件 W_2 发生仅当 $y^* \notin V_i$ 且式(4)成立时。

假设 $(t'_1, t'_2, \dots, t'_m)$ 为线性空间 V_i 的基向量 (v_1, v_2, \dots, v_m) 所对应的MAC标签,则:

$$y' = \sum_{j=1}^m y_j^* v_j \in V_i$$

对应的MAC标签为:

$$t' = \sum_{j=1}^m y_j^* t'_j \in \mathbb{F}_q$$

则标签 t' 是向量 y' 的一个合法标签。因此,我们知道下面两个等式和成立:

$$(u \cdot y^*) + \sum_{j=1}^m y_j^* b_{i,j} = t \quad (5)$$

$$(u \cdot y') + \sum_{j=1}^m y_j' b_{i,j} = t' \quad (6)$$

用式(5)减去式(6)可以得到:

$$(u \cdot (y^* - y')) = t - t' \quad (7)$$

这说明如果攻击者 \mathcal{A} 如果能输出一个伪造,就意味着找到可以满足式(7)的 y^* 和 t 。 $y^* \notin V_i$ 且 $y' \in V_i$,因此可以知道 $y^* \neq y'$ 。然而从攻击者 \mathcal{A} 的视图看, u 和一个 \mathbb{F}_q^{n+m} 中的随机变量是无法区分的,因此满足式(7)的概率为 $1/q$ 。因此,存在概率有 $\Pr[W_2 \wedge \neg T] = (1/q) \cdot \Pr[\neg T]$ 。我们联合概率 $\Pr[W_2 \wedge T]$ 和概率 $\Pr[W_2 \wedge \neg T]$ 的发生可能性,可以得到以下等式:

$$\begin{aligned} \Pr[W_2] &= \Pr[W_2 \wedge T] + \Pr[W_2 \wedge \neg T] = \\ &= 1/q(\Pr[T] + \Pr[\neg T]) = \\ &= 1/q \end{aligned} \quad (8)$$

结合式(1) - 式(3)和式(8),我们可以得到:

$$\begin{aligned} \text{NC} - \text{Adv}[\mathcal{A}, \mathcal{T}] &\leq \text{PRF} - \text{Adv}[\mathcal{B}_1, \mathbb{F}] + \\ &\text{PRG} - \text{Adv}[\mathcal{B}_2, \mathbb{G}] + (1/q) \end{aligned}$$

因此,概率 $\text{HB} - \text{Adv}[\mathcal{A}, \mathcal{T}]$ 的边界为:

$$\begin{aligned} \text{HB} - \text{Adv}[\mathcal{A}, \mathcal{T}] &= \text{NC} - \text{Adv}[\mathcal{A}, \mathcal{T}] (1 - 2^{-\frac{b_2}{4}}) + \\ &(1 - \text{NC} - \text{Adv}[\mathcal{A}, \mathcal{T}]) 2^{-\Omega(r(n+1))} < \\ &\text{NC} - \text{Adv}[\mathcal{A}, \mathcal{T}] \end{aligned}$$

定理证毕。

7 方案的性能分析

我们主要从带宽开销和计算开销的角度来评估一

下本文所提出方案的性能。

在评估带宽开销时,我们忽略在系统初始化阶段的带宽开销,因为这些操作可以在离线的时候进行。网络中传输的每个数据包的带宽开销包含一个空间标识符 $id \in \mathcal{I}$ 和一个MAC标签。假设空间标识符集合的大小是 \mathcal{I} ,因此空间标识符的大小是 $\log_2 \mathcal{I}$ 。数据包的大小是 $2 \times \log_2 q$ 。网络中传输的每个增广向量 $v \in \mathbb{F}_q^{m+n}$ 的大小是 $m+n$ 个有限域 \mathbb{F}_q 上元素,源节点增广向量大小是 $(m+n)\log_2 q$,因此带宽开销共为 $\frac{2\log_2 q + \log_2 \mathcal{I}}{(m+n)\log_2 q}$ 。此外,本文提出的加密方案的中间节点验证的带宽开销为 l 个比特。

在方案的算法签名阶段,对于每个要发送的数据包,源节点需要计算 $m+n$ 次有限域 \mathbb{F}_q 上的乘法运算以及 $m+n$ 次有限域 \mathbb{F}_q 上的加法运算来生成合法的MAC标签。在组合算法阶段,假设使用 ρ 表示网络中间节点每轮编码的数据包,为了生成一个合法MAC标签,中间节点需要进行 ρ 次有限域 \mathbb{F}_q 上的乘法和 $\rho-1$ 次有限域 \mathbb{F}_q 上的加法运算。而对于 ρ 个数据包的网络编码基本操作则需要 $\rho(m+n)$ 次有限域 \mathbb{F}_q 上的乘法以及 $\rho(m+n-1)$ 次加法运算。在验证算法1阶段,本文提出的方案不需要有限域中的指数运算,为了验证一个数据包的合法性,接收节点只需要进行 $m+n$ 次有限域 \mathbb{F}_q 上的乘法运算以及 $m+n$ 次有限域 \mathbb{F}_q 上的加法运算。在验证算法2阶段,本文提出的方案中,中间节点的计算开销为两次 $r(n+1) \times l$ 维矩阵和一个 l 维 \mathbb{Z}_2 上的向量相乘以及 \mathbb{Z}_2^l 上的两次加法运算。

8 结语

本文提出了一个基于LPN问题难度的HB协议网络编码同态MAC加密方案,并在基本模型下证明了所提出方案的安全性。与以往的基于对称密钥的方法的解决方案相比,本文所提出的解决方案允许网络中间节点验证所收到数据包的合法性,可以尽早地发现并过滤掉被污染的数据包,以阻击污染攻击的进一步扩大。与以往的网络编码签名方案相比,本文提出的方案不需要网络节点进行双线性群配对操作或者模幂运算,且可以应用在一个较小的有限域,因此具有更低的计算开销和带宽开销,非常适用于实时性较强的网络编码应用。

参 考 文 献

- [1] Ahlswede R, Cai N, Li S Y R, et al. Network information flow[J]. *IEEE Transactions on Information Theory*, 2000, 46(4):1204 – 1216.
- [2] Liu G, Xiao S. Scalable security solution against wiretapping for network coding based priority encoding transmission[J]. *Chinese Journal of Electronics*, 2018, 27(1):191 – 197.
- [3] Anglano C, Gaeta R, Grangetto M. Securing coding-based cloud storage against pollution attacks[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2017, 28(5):1457 – 1469.
- [4] Wei T, Sheng Z. A unified resource allocation framework for defending against pollution attacks in wireless network coding systems[J]. *IEEE Transactions on Information Forensics and Security*, 2017, 11(10):2255 – 2267.
- [5] Guo H, Wang X, Chang K. exploiting path diversity for thwarting pollution attacks in named data networking[J]. *IEEE Transactions on Information Forensics and Security*, 2016, 11(9):2077 – 2090.
- [6] Shang T, Pei Z, Zhao X, et al. Quantum network coding against pollution attacks[J]. *IEEE Communications Letters*, 2016, 20(7):1369 – 1372.
- [7] Boneh D, Freeman D, Katz J, et al. Signing a linear subspace: signature schemes for network coding[C]//*Public Key Cryptography*. Springer Berlin Heidelberg, 2009: 68 – 87.
- [8] Yun A, Cheon J H, Kim Y. On homomorphic signatures for network coding[J]. *IEEE Transactions on Computers*, 2010, 59(9):1295 – 1296.
- [9] Liang M, Kan H. An efficient hybrid cryptographic scheme for wireless sensor network with network coding[J]. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 2013, 96(9):1889 – 1894.
- [10] Ho T, Médard M, Koetter R. A random linear network coding approach to multicast[J]. *IEEE Transactions on Information Theory*, 2006, 52(10):4413 – 4430.
- [11] Agrawal S, Boneh D. Homomorphic MACs: MAC-based integrity for network coding[C]//*Applied Cryptography and Network Security*. Springer Berlin Heidelberg, 2009: 292 – 305.
- [12] Cheng C, Jiang T. A novel homomorphic MAC scheme for authentication in network coding[J]. *IEEE Communications Letters*, 2011, 15(11):1228 – 1230.
- [13] Blum A, Kalai A, Wasserman H. Noise-tolerant learning, the parity problem, and the statistical query model[J]. *Journal of the ACM*, 2003, 50(4):506 – 519.
- [14] Hopper N J, Blum M. Secure human identification protocols [C]//*Advances in cryptology—ASIACRYPT 2001*. Springer Berlin Heidelberg, 2001:52 – 66.
- [15] Katz J, Lindell Y. Introduction to modern cryptography: principles and protocols[M]. CRC Press, 2007.
- [16] Blum A, Furst M, Kearns M, et al. Cryptographic primitives based on hard learning problems [C]//*Advances in Cryptology — CRYPTO' 93*. Springer Berlin Heidelberg, 1993:278 – 291.
- [17] Katz J, Ji S S, Smith A. Parallel and concurrent security of the HB and HB[J]. *Journal of Cryptology*, 2010, 23(3):402 – 421.
- [18] Berlekamp E R, McEliece R J, Van Tilborg H C A. On the inherent intractability of certain coding problems[J]. *IEEE Transactions on Information Theory*, 1978, 24(3):384 – 386.
- [19] Lyubashevsky V. The parity problem in the presence of noise, decoding random linear codes, and the subset sum problem [J]. *Lecture Notes in Computer Science*, 2005, 3624:378 – 389.
- [20] Stern J. A method for finding codewords of small weight [C]//*Coding Theory and Applications*. Springer Berlin Heidelberg, 1989:106 – 113.

(上接第 250 页)

- [8] 杨立娜. 基于相位相关理论的最大互信息图像配准[D]. 西安:西安电子科技大学, 2010.
- [9] 黄志勇, 陈一民. 基于频域相位相关的自适应光学图像配准算法[J]. *计算机应用与软件*, 2016, 33(5):166 – 168.
- [10] 许雷, 俞锋. 一种基于相位相关法及数学形态学方法的眼底血管图像自动拼[J]. *生物医学工程学杂志*, 1998(3):286 – 290.
- [11] 韩峻峰, 王帅. 基于双目立体视觉技术的汽车测距系统实现[J]. *计算机应用与软件*, 2016, 33(9):227 – 230.
- [12] 王静, 张军华, 单联瑜, 等. 相位相关速度分析方法研究及效果比较[J]. *物探化探计算技术*, 2010, 32(2):164 – 167.
- [13] 王志强, 程红, 孙文邦, 等. 一种基于梯度最大值相位相关的航空影像自动配准算法[J]. *地理与地理信息科学*, 2008, 24(6):111 – 112.
- [14] 陈敏, 许雪林. 基于 OpenCV 的互相关图像测速方法研究[J]. *科学技术与工程*, 2010, 10(15):3570 – 3573.
- [15] 王宏兴, 霍玉洪. 奇异值分解的教与学[J]. *淮南师范学院学报*, 2015(3):125 – 127.
- [16] 牛亚坤, 玉振明, 李陶深. 一种近似奇异值分解的观测矩阵优化方法[J]. *计算机应用与软件*, 2016, 33(1):195 – 197, 262.