基于模分量的同态加密方法研究与应用

于浩洋[1,2], 封化民[2*], 李晓东[2], 金 鑫[2], 刘 飚[2]

[1]北京邮电大学 网络空间安全学院 北京 中国 100876
[2]北京电子科技学院 北京 中国 100070

摘要 随着云计算的发展, 海量数据的处理正逐渐从用户本地转向云服务器, 然而数据本身可能携带大量用户隐私, 且一旦用户将数据上传至云服务器, 就失去了对数据的完全掌控能力, 该类数据一旦被非法获取, 用户身份、行为、偏好等各类隐私就可能被暴露。因此, 如何保证在不暴露原始数据的情况下让受委托的云服务器在密文下执行运算成为一个重要的研究课题。本文基于密码学和计算机视觉相关理论, 针对隐私数据安全处理的问题, 以模分量的同态性质为基础设计了两种加密方法, 分别为基于混淆模分解的同态加密方法和基于密模聚合的同态加密方法, 并给出了安全性分析。并将这两种方法应用于视觉盲计算领域中, 实现计算方在无需获取任何原始数据有效信息的密文条件下, 完成对数据的盲处理, 实现了数据的可用不可见。实验结果表明, 基于密模聚合模同态加密的运动目标盲提取方法, 在多数测试场景中能在不降低原始算法准确率的前提下, 在时间效率上明显优于基于混合高斯模型的运动目标盲提取和基于多服务器秘密共享的前景提取等方法; 基于混淆模分解同态加密的人脸盲检测方法, 能在不降低原始人脸检测算法识别的准确率前提下, 实现视频监控人脸的盲检测, 且检测速度大幅度快于基于随机子图的隐秘人脸检测方法和基于随机向量的隐秘人脸检测等算法。

关键词 同态加密; 视觉盲计算; 运动目标盲提取; 人脸盲检测
中图法分类号 TP3-05 DOI 号 10.19363/J.cnki.cn10-1380/tn.2023.09.03

Research and Implementation of Blind Video Processing Method Based on Modular Component Homomorphism

YU Haoyang[1,2], FENG Huamin[2*], LI Xiaodong[2], JIN Xin[2] , LIU Biao[2]

[1] School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2] Beijing Electronic Science and Technology Institute, Beijing 100070, China

Abstract With the development of cloud computing, the processing of massive data is gradually shifting from local users to cloud servers. However, the data itself may carry a lot of user privacy, and once users upload the data to the cloud servers, they will lose their full control over the data. Once such data is illegally obtained, all kinds of privacy such as user identity, behavior and preferences may be exposed, and the consequences will be unimaginable. Therefore, it has become an important research topic how to ensure that the entrusted cloud server can perform operations under ciphertext without exposing the original data. Based on the related theories of cryptography and computer vision, aiming at the problem of privacy data security, this paper designs two encryption methods based on the homomorphism of modular components, namely, homomorphism encryption method based on confusing modular decomposition and homomorphism encryption method based on dense modular aggregation, and gives a detailed security analysis. These two methods are applied to the field of visual blind computing, so that the computing side can complete the blind processing of the data without obtaining any ciphertext of the original data, and the data is available and invisible. The experimental results show that the blind extraction method of moving target based on dense mode aggregation homomorphism encryption is obviously superior to blind extraction of moving target based on mixed Gaussian model and foreground extraction based on multi-server secret sharing in most test scenarios without reducing the accuracy of the original algorithm. The blind face detection method based on homomorphic encryption of Confused Modulus Decomposition can realize blind face detection in video surveillance without reducing the recognition accuracy of the original face detection algorithm, and the detection speed is greatly faster than the hidden face detection method based on random subgraph and the hidden face detection algorithm based on random vector.

Key words homomorphic encryption; blind vision; blind object segmentation; blind face detection

通讯作者: 封化民, 博士, 教授, Email: fenghm@besti.edu.cn
本课题得到基金项目: 国家重点研发计划资助项目(No. 2018YFB0803600)、国家自然科学基金项目(No. 61872091)、北京电子科技学院一流学科建设项目(No. 3201024)、北京邮电大学博士生创新基金资助项目(No. CX2021124)。
收稿日期: 2022-01-12; 修改日期: 2022-04-12; 定稿日期: 2023-06-12

1 引言

计算机可视媒体领域借助云计算的发展, 正在全面改变人们的生产生活方式, 深刻影响人类社会历史发展进程[1]。信息技术广泛应用和网络空间发展的兴起, 虽然极大地促进经济社会繁荣进步, 但同时也带来新的安全风险和挑战[2]。图像视频智能云服务为人们带来极大的便利, 但用户一旦将图像视频数据上传到网上, 就失去对该数据的完全掌控能力, 人们开始担心云服务提供商无法妥善解决数据的安全性问题, 安全性问题逐渐成为网络空间安全领域不可忽视的隐患。一旦攻击者成功攻击云服务器, 用户的隐私数据将遭到泄露, 攻击者可能滥用这些携带大量用户隐私的图像视频数据并从事违法活动。例如, 2014 年 8 月 31 日, 多名好莱坞女艺人上传到 Apple 的云存储服务 iCloud 的 200 多张私人照片被盗并上传到贴纸网站 http://www.4chan.org。如果隐私泄露事件发生在政府、科研院所等信息更加敏感的部门或组织, 可能会对国家和社会造成无法挽回的损失。这些事件的发生, 使人们逐渐意识到: 图像视频等大数据的云存储与计算在给人们带来方便的同时, 也带来内容隐私泄露风险的网络空间安全的新挑战, 如何保证在不暴露数据隐私的情况下让云服务器安全地对数据进行处理成为当下热门的研究课题。

视觉盲计算技术能够在不接触图像、视频等视觉数据原始内容的情况下对其进行检测、识别、检索以及更复杂的处理, 视觉盲计算技术能够有效解决云端视频数据的安全性问题[3]。密码学中常用的分组密码虽然具有很高的加密效率和安全强度, 用户可以将加密后的密文数据存储在云端, 云端由于无法获得密钥而无法破解用户的明文信息, 然而, 用户如果想要对数据进行处理则必须将密文数据解密成明文, 与此同时, 云服务器端也无法在密文下实现对数据的检测、识别、检索以及更复杂的处理, 因此该问题成为云端视频数据的安全处理的难题, 而视觉盲计算的出现让该问题的解决成为可能。

视觉盲计算的一个典型的应用场景为: 客户端拥有敏感监控数据并想定位其中的人脸, 云服务器端拥有一个人脸检测算法可以为客户端提供面部检测服务, 但客户端不想让云服务器端得到敏感监控数据的有效信息, 云服务器端也不想将人脸检测算法直接泄露客户端。视觉盲计算的出现可以让云服务器端在不获取客户端原始敏感监控数据有效信息的情况下对其执行人脸检测算法, 得到正确的检测结果并将其返回给客户端, 同时不向客户端泄露算法参数。视觉盲计算技术使用的密码学方法主要包括同态加密、安全多方计算、函数加密等, 能够确保用户的隐私数据在一个安全、可信的条件下处理。视觉盲计算技术是计算机视觉与密码学等领域学科交叉的新方向, 在视频监控、多媒体数据共享、云计算、移动计算等领域有广泛的应用前景。

本文的主要贡献主要为三部分: (1)提出基于模分量的同态加密算法。该算法基于模投影、同余定理以及中国剩余定理(CRT), 利用该算法设计两种数据盲处理方法: 一种是基于混淆模分解的同态加密方法, 另一种是基于密模聚合的同态加密方法。基于混淆模分解的同态加密方法支持多种盲运算, 兼容多种视频处理算法; 基于密模聚合的同态加密方法可以实现单次盲加减运算, 该算法处理效率高, 适用于简单的视频处理算法。(2)设计并实现一种基于密模聚合模同态加密的运动目标盲提取方法。使用基于密模聚合的同态加密方法对 ViBE 运动目标提取算法进行修改, 使得云服务器在不获取原始视频信息的情况下, 完成对视频帧中运动目标的盲提取。(3)设计并实现一种基于混淆模分解的模同态人脸盲检测方法。该方法利用基于混淆模分解的同态加密方法将明文帧加密成密文数据, 将 V&J 人脸检测修改为盲化版人脸盲检测方法。在基于混淆模分解同态加密的人脸盲检测方法中, 客户端不向云服务器暴露视频隐私数据, 云服务器端的人脸检测算法参数也不会泄露给客户端, 整个计算过程在双盲的条件完成。

2 背景和相关工作

2.1 视觉盲计算

视觉盲计算(Blind Vision)最早由 Avidan 等人[4]于 2006 年提出, 是指在计算方在无需接触图像、视频等视觉媒体原始数据的条件下完成对其进行的检测、识别、提取、检索等处理的计算。Avidan 等人使用标准密码工具将目标检测 Viola & Jones 算法[5]盲化, 实现在不泄露客户端隐私图像和云端服务器目标检测分类器参数的条件下完成对物体位置的盲检测, 然而由于该方法利用的安全多方计算[6]复杂度过高, 导致其计算速度极其缓慢, 一张完整图像的检测时间通常需要若干小时, 远远不能达实时检测和可用性的要求。

2004 年, Viola 等[5]提出一种把安全多方计算方法应用到 Viola & Jones 目标检测算法的隐私保护的视频人脸检测方法。该方法利用不经意传输方法(oblivious transfer, OT[7])实现安全点积方法和安全阈

值比较方法, 并利用这两种方法成功改造安全分类器, 实现图像中人脸的安全检测, 但由于安全多方计算的是计算密集型算法, 如果将其应用在图像或视频等视觉媒体数据集上, 会出现计算效率低、处理速度慢的问题, 该问题也成为众多视觉盲计算方法的所面临的主要问题。

2009 年, Upmanyu 等人[8]在国际计算机视觉会议上提出一种基于秘密分享机制的监控系统视频安全处理方法, 该方法利用中国剩余定理(CRT)将视频监控采集端的视频帧分裂成多个随机子图像, 将该图像分别发送至对应个不同的服务器, 每个服务器对该视频采取相同的处理算法并将得到的对应个结果发送至可信终端, 该终端利用中国剩余定理将多个结果合并得到最终的结果, 整个过程任意单个提供计算的云服务器不会获得任何有意义的信息。

2013 年, Chu 等人[9]在国际多媒体会议上提出一种对运动目标的实时检测方法, 客户端将视频监控捕获的视频数据加密成密文数据, 并将密文结果发送给服务器端, 服务器端对加密后的数据进行目标检测计算, 将检测结果返回给客户端。2014 年, Chu 等人[10]提出一种基于扰乱电路的(Garbled Circuits)的多监控摄像机目标跟踪方法, 该方法利用扰乱电路将更多的计算安排在监控端, 从而减少个监控终端计算量。与传统非隐私保护的目标跟踪方法相比, 该方法在保证不降低识别准确率、不大幅提高计算时间的条件下实现目标的安全跟踪。

2015 年, Bost 等人[11]提出一种基于构建隐私保护分类器的加密机器学习分类方法。该方法对超平面(Hyperplane)决策、朴素贝叶斯(Naive Bayesian)、决策树(Decision Tree)进行构建, 借助 Adaboost 将分类器合并, 同时将该方法设计成盲计算模块库, 以合并这些盲计算模块库的方法构建更加复杂的隐私保护分类器, 成功将其应用于隐私保护的人脸安全检测等方面。

2017 年, Jin 等人[12]提出一种基于随机子图的隐私保护人脸检测方法, 该方法中 Alice 将一张图片分成 255 张随机子图, Bob 对这 255 张子图进行人脸检测最后将计算结果发送给 Alice 端, 由于 Bob 无法依据 255 张随机子图将原始图片复原, 所以该方法可以实现保护人脸隐私情况下的人脸检测。

2.2 同态加密

同态加密是一种特定的加密形式, 该加密允许对明文加密后生成的密文进行计算, 在密文上进行计算得到的结果经过解密与在明文上进行相对应的计算结果相同。同态加密经常被应用于涉及隐私安全的外包存储和外包计算, 运用同态加密, 服务需求方对数据执行同态加密并将加密结果发送到专业的云服务提供商, 因为云服务提供商所进行的操作都是在密文状态下进行的, 所以服务需求方的数据的安全性能得到很好的保障。同态加密也常常被应用在医疗保险、身份认证、党政信息化系统等安全性要求较高的行业, 例如, 服务需求方需要对医疗信息进行预测分析, 但需求方本身没有相应的处理算法而不得不将数据外包给预测分析服务提供商处理, 但医疗信息中所携带的大量隐私信息不能暴露给预测分析服务提供商, 如果预测分析服务提供商可以转而对加密数据进行操作, 则这一类隐私问题就能得到很好地解决。

同态加密的概念首先由 Rivest 等人[13]于 1978 年提出, 同态加密的提出是为了解决以下问题: 每次要对加密数据执行某些操作时, 必须先解密以获得纯文本, 然后再进行进一步的操作。同态加密主要分为 3 种, 分别为部分同态加密(Partially Homomorphic Encryption, PHE)、层次的同态加密(Somewhat Homomorphic Encryption, SHE)和全同态加密(Fully Homomorphic Encryption, FHE), 它们的主要区别方式为对其密文执行的数学运算的类型和频率。

部分同态加密通过仅允许对加密值执行选择的数学函数来帮助敏感数据保持机密。Rivest 等人[14]于 1978 年提出 RSA 公钥加密体制, 该加密体制能够实现同态乘法的运算, 但不能实现同态加法的运算。1985 年 Elgamal[15]提出一种基于椭圆曲线的加密算法, 该算法能够实现同态乘法运算, 但该算法不能实现同态加法运算。Hsieh 等人[16]于 1988 年提出一种加密同态加密方法, 能够实现加同态但不支持乘同态。1999 年, Paillier[17]提出一种支持任意次数加密运算的单同态算法并将其命名为 Paillier 加密。Kuribayashi 等人[18]于 2005 年提出一种同态加密体制, 该体制基于椭圆曲线上双线性对, 该加密体制能够实现任意次数的密文加法同态运算, 并且可以实现一次密文下的同态乘法运算。

层次的同态加密方案是支持有限次操作的同态加方法, 层次的同态加密最初由 Gentry[19]于 2009 年中提出, 该方案基于理想格的全同态加密算法, 能够满足有限次同态计算, 通过同态解密来实现对密文的复原, 达到全同态加密的效果。

同态加密与其他形式的同态加密不同, 能够实现在不加解密密文的条件下对其执行任意的计算, 解密后可以得到相应计算的明文结果。从 2009 年 Gentry[19]提出层次的同态加密算法至今, 很多全同

态加密算法被国内外学者们提出, 全同态加密算法的研究进入快速发展的阶段。全同态加密虽然仍处于开发阶段, 但其有望成为解决云计算隐私保护问题的重要关键手段, 借助 5G 时代下数据的超高速传输和强大的云计算处理算法, 有望实现安全、高效的云计算视觉信息盲处理, 用户隐私等安全性问题将得到有效解决。

在近 9 年的全同态加密发展过程中, 大致可以划分为三个阶段: 第一阶段是 Gentry[19]在 2009 年的突破性工作; 第二阶段是 Brakerski 和 Vaikuntanathan[20]首次利用容错学习(Learning with Errors)假设实现全同态加密; 第三阶段则是 Gentry 等人[21]首次利用近似特征向量的方法实现全同态加密, 该方案就是当前最为经典的 Gentry-Sahai-Waters 方案, 该方案在同态运算时不再依赖于计算公钥。

本文在文献[22]同态加密库中使用的方法的基础上, 创新性地设计了基于混淆模分解的同态加密方法, 并在该方法中增加了高效正负判断操作, 防止由于计算结果为负数造成的解密结果错误, 同时在算法中增加了浮点数运算的支持, 使得该算法从只支持整型数据运算变为同时支持整型数据和浮点数数据运算的同态加密方法, 因此该算法可以成功应用在机器学习等需要大量浮点数计算的计算场景中。此外, 本文创新性设计了基于密模聚合的同态加密方法, 并引入多数据并行的思想, 将加密操作和数据并行操作融合, 该方法为适应有快速盲加减运算需求的计算场景, 如电子拍卖、快速集合求交集、计算机视觉领域中的背景相减等。

2.3 运动目标提取技术

运动目标提取技术是一种在图像中将前景对象从静止的背景帧中提取出来的技术, 该技术被广泛应用于电影、电视、出版和摄影等领域。运动目标提取根据应用场景的不同, 可以分为非实时的运动目标提取和实时的运动目标提取。非实时的运动目标提取的一个重要应用场景为电影拍摄。实时运动目标提取的应用场景有视频通话、行为识别、轨迹追踪等。由于实时运动目标提取对算法的耗时要求非常高, 因此有很大的挑战。

常用的运动目标提取算法有帧差法[23]、基于光流场[24]的运动目标提取算法以及背景建模法[25]。帧差法最早由 Yang 等人[26]提出, 随后 Lipton 等人[23]在算法的实时性和鲁棒性进行改良, 提出三帧差法和多帧差法, 为了解决视频帧中噪声的问题, Seki 等[27]人提出一种以小区域为单位提取相似性的差分算法。基于光流场的运动目标提取算法最早由 Horn 等人[28]提出, 该方法将灰度和二维速度场结合实现光流约束算法。1997 年, Lucas 等人[29]对 Horn 的方法进行改良, 他们利用局部平滑性约束和最小化能量泛函来估计运动目标的光流场。最早的传统建模算法有中值滤波法[30]和均值滤波法[31], 两种滤波方法需要持续的保存背景数据, 提取准确率较低且只能适应多变的场景。混合高斯背景模型最早由 Grimson 等人提出, 该方法利用多个高斯模型去拟合背景像素状态, 然而该算法涉及的参数量较多, 导致计算量较大[32]。McKennat 等人[33]于 2000 年提出自适应背景建模方法, 该方法根据图像中像素值和相邻像素间的梯度信息来解决阴影问题, 进而提高提取准确率。2016 年, 宋志勤等[34]将像素值和纹理体征结合, 提出一种时空背景模型差分法, 该方法能处理单通道灰度图像, 内存占用低, 识别准确率高。

2.4 人脸检测技术

人脸检测是一种在多种应用中使用的计算机技术, 可以识别数字图像中的人脸。人脸检测是物体检测的一种, 在物体检测中, 任务是查找图像中属于给定类的所有对象的位置和大小。人脸检测算法专注于检测正面人脸, 它将待检测图像与数据库中存储的图像匹配, 数据库中任何面部特征的更改都会导致匹配失败。人脸检测技术常被应用在身份认证、安全防护、媒体娱乐以及图像匹配与检索中。已有的人脸检测算法有肤色模型、模板匹配模型、人工神经网络(artificial neural networks, ANN)模型、支撑向量机(support vector machine, SVM)模型、自适应增强(adaptive boosting, Adaboost)模型等, 但是隐私保护的机器视觉人脸检测算法却很少。

20 世纪 90 年代, 人脸检测技术开始取得较大进展 Yang 等人[35]于 1994 年提出基于多分辨率和指导搜索的人脸检测方法; Graf 等人[36]于 1995 年提出基于人脸面部特征的人脸检测方法。Samaria 等人[37]于 1994 年提出一种基于隐形马尔科夫模型的人脸检测方法。Osuna 等人[38]于 1997 年提出基于支持向量机的人脸检测方法。Paul Viola 等人[5]于 2001 年设计了一种 Viola&Jones 人脸检测方法, 该算法极大地缩短了人脸检测时间, 达到每秒 25 帧的检测速度。该算法的核心为 Haar-like 特征[39]、Adaboost[40]以及 Cascade 级联分类器。

随着深度学习技术的发展, 人脸检测技术发展也突飞猛进。Ross Girshick 等人[41]于 2013 年提出 R-CNN(Regions with Convolutional Neural Network Features), 成为深度学习领域的经典研究。He 等人[42]于 2014 年提出 SPP-net(Spatial Pyramid Pooling-net,

空间金字塔池化层), 该思想在人脸检测领域得到广泛应用。Ross Girshick 等人[43]在 2015 年提出一种快速 R-CNN 人脸检测方法, 解决了 R-CNN 和 SPP-net 多候选框造成的重复计算问题。2016 年, S. Ren 等人[44]设计出了 Faster R-CNN, 使区域提名、分类、回归共用卷积特征, 检测速度得到进一步加速。2016 年 Dai 等人[45]提出 R-FCN(Region-based Fully Convolutional Networks), 区域提名全卷积神经网络。2016 年 Redmon 等人[46]提出 YOLO(You Only Look Once)实时人脸检测技术, 能够到达 45-150FPS。可以看出基于深度学习的人脸检测技术发展迅猛, 检测速率和检测准确率都得到大幅提高。

3 协议设计

3.1 模分量同态理论

3.1.1 模投影

模投影定理与三维空间直角坐标系上点的投影相类似。空间任意选定一点 O, 过点 O 作 3 条互相垂直的坐标轴 O_x、O_y、O_z, 它们都以 O 为原点且具有相同的长度单位。以此坐标系为基准, 设定长度单位, 则在自然界中任何位置 M 在这该坐标系下的 3 个坐标轴上都有一个投影, 如果在 3 个坐标轴的投影分别为 x_1、y_1、z_1, 则该点记作 $\mathrm{M}(x_1,y_1,z_1)$。由此可见, 在三维空间中, 已知 3 个坐标轴上的投影可以确定该点的唯一位置, 只知道到 x_1、y_1、z_1 3 个数据中的任意两个或一个数据没有办法确定坐标点的确切位置。模投影定理与之相似之处在于, 将 3 个的两两互素的正整数 a、b、c 作为坐标轴, 组成坐标系 O_{abc}, 则任何一个正整数 N 在该坐标系下对 a、b、c 分别进行取模运算, 取模结果记作 $N(a_1,b_1,c_1)$。则在范围 $[0,a\times b\times c]$ 下, 已知 a_1,b_1,c_1 和坐标系 O_{abc}, 可以计算出 N 的唯一确定值。令 $a=7$、$b=11$、$c=13$, 则 31 在“7 轴”上的投影为 3, 在“11 轴”上的投影为 9, 在“13 轴”上的投影为 5, 这 3 个被模的数组成的集合也被称为一组模基。

3.1.2 同余定理

数学上, 如果两个整数同除以一个整数, 得到相同的余数, 那么这两个整数为同余关系(Modular arithmetic), 该理论经常被用于数据算法, 最先引用同余的概念与符号的人为德国数学家高斯。同余理论是初等数论的重要组成部分, 是研究整数问题的重要工具之一, 利用同余来论证某些整除性的问题是很简便的。

同余定义: 设:

m 是大于 1 的正整数,

a、b 是整数, 如果 $m\mid(a-b)$, 则称 a 与 b 关于模 m 同余, 记作 $a\equiv b(\mathrm{mod}\ m)$, 读作 a 与 b 对模 m 同余。

显然, 有如下事实

(1) 若 $a\equiv 0(\mathrm{mod}\ m)$, 则 $m\mid a$;

(2) $a\equiv b(\mathrm{mod}\ m)$ 等价于 a 与 b 分别用 m 去除, 余数相同。

同余性质: 如果 $a\equiv b(\mathrm{mod}\ m)$, $c\equiv d(\mathrm{mod}\ m)$, 那么有:

$$\begin{aligned} a\pm c &\equiv b\pm d(\mathrm{mod}\ m)\\ a\times c &\equiv b\times d(\mathrm{mod}\ m)\end{aligned} \tag{1}$$

3.1.3 模运算性质

(1) 基本概念:

给定一个正整数 p, 任意一个整数 n, 一定存在等式 $n=kp+r$; 其中 k、r 是整数, 且 $0\leqslant r<p$, 称 k 为 n 除 p 的商, r 为 n 除以 p 的余数。

取模运算即对于一个整数 a, $a\%p$ (或 $a\bmod p$), 表示 a 除以 p 的余数。

3.1.4 中国剩余定理

中国剩余定理, 又称中国余数定理、孙子定理, 是中国古代求解一次同余式组的方法, 是数论中的另一个重要定理。

对于一元线性同余方程组 S,

$$(S):\begin{cases} x\equiv a_1(\mathrm{mod}\ m_1)\\ x\equiv a_2(\mathrm{mod}\ m_2)\\ \vdots\\ x\equiv a_n(\mathrm{mod}\ m_n)\end{cases} \tag{2}$$

假设整数 $m_1,m_2,...,m_n$ 两两互质, 则对任意的整数: $a_1,a_2,...,a_n$, 方程组 (S) 有解, 通解可以通过以下方式构造得到:

设 $M=m_1\times m_2\times\cdots\times m_n=\prod_{i=1}^{n}m_i$ 是整数 $m_1,m_2,...,m_n$ 的乘积,

$M_i=M/m_i$, $\forall i\in\{1,2,\cdots,n\}$ 是除 m_i 以外的 $n-1$ 个整数的乘积,

$t_i=M_i^{-1}$ 为 M_i 模 m_i 的逆元, $M_it_i\equiv 1(\mathrm{mod}\ m_i)$, $\forall i\in\{1,2,\cdots,n\}$。

方程组 (S) 的通解形式为 $x=a_1t_1M_1+a_2t_2M_2+\cdots+a_nt_nM_n+kM=kM+\sum_{i=1}^{n}a_it_iM$, $k\in\mathbb{Z}$。

在模 M 的意义下, 方程组 (S) 有且只有一个解

$x=(\sum_{i=1}^{n} a_i t_i M) \bmod M$ 。

3.2 基于混淆模分解的同态加密方法

3.2.1 方法设计

基于混淆模分解的同态加密方法分为三个部分: 加密、盲计算、解密, 参数的详细说明如表 1 所示。

表 1 混淆模分解的同态加密方法参数说明

Table 1 Parameter description of homomorphic encryption method of confusion module decomposition

参数名称	含义
m_0	原始输入明文
a	将原始输入明文放大的倍数。
η	随机数噪声 η, 用于掩盖放大后的数据。
m'	放大加扰后的数据, $m'=am_0+\eta$
B	模基, 被模的数
n	模基中元素的个数
M	真实模分量
R'	和 m' 生成方式相同的初始冗余集合
m	冗余的个数
R	根据 m 挑选出的冗余集
S	冗余索引, 真实模分量的位置索引
$E(m_0)$	加密后得到的 $n\times m$ 的密文
d	真实模分量集对应的计算结果
pk	(a,B), 公钥包含放大倍数和模基
sk	(a,B,S), 私钥包含放大倍数、模基、冗余索引

在加密阶段, 客户端需要设定模基、冗余集以及公钥 $\text{pk}\leftarrow(a,B)$ 和私钥 $\text{sk}\leftarrow(a,B,S)$, 公钥中包括放大倍数 a 和模基 B, 私钥中包括放大倍数 a、模基 B 和冗余索引。模基是一组两两互素的正整数, 用于分裂数据, 冗余集可以将经过放大和加扰的原始数据分裂生成的模分量混淆, 使得云服务器端无法分辨真实模分量, 位置模板是密钥, 用于保存真实模分量的位置, 以便在客户端解密时能找到正确的计算后的模分量。设定完成后对原始数据进行放大和加扰操作, 生成一个整数, 然后将该整数依据模基生成模分量, 并将其依照密钥插入冗余集中, 生成密文数据, 将密文数据和模基发送给云服务器端。输入为明文 m_0、模基个数 n、冗余个数 m、冗余索引 S, 输出为密文 E, 模基 B。具体算法流程如下所示。

步骤 1: 生成答数, 将原始信息 m 放大 a 倍并增加随机数噪声 η, 得到经过放大并添加随机数的数据 m', 对于每一个明文数据, 其噪声 η 都不相同, 放大倍数 $a\gg\eta$;

$$m'=am_0+\eta \tag{3}$$

步骤 2: 设置模基 B, 模基也可称为投影基, 作用是在加密阶段客户端使用模基 B 将原始信息分裂。其中 b_1, b_2, …, b_n 为 B 中的元素, 称为模基元素, b_1, b_2, …, b_n 为正整数且两两互素;

$$\begin{gathered} B\leftarrow\{b_1,b_2,\cdots,b_n\} \\ \left(b_i,b_j\right)=1, i,j=1,2,\cdots,n, i\neq j \end{gathered} \tag{4}$$

步骤 3: 计算真实模分量集 M;

$$\begin{gathered} M\leftarrow\{m_1,m_2,\cdots,m_n\} \\ m_i=m' \bmod b_i \\ i=1,2,\cdots,n \end{gathered} \tag{5}$$

步骤 4: 计算初始冗余初始集 R';

$$R'\leftarrow\left\{r_1^{'},r_2^{'},\cdots,r_m^{'}\right\} \tag{6}$$

步骤 5: 计算冗余集 R;

$$\begin{gathered} R\leftarrow\{R_1,R_2,\cdots,R_n\} \\ R_i\leftarrow\left\{r_{i1},r_{i2},\cdots,r_{im}\right\}, i=1,2,\cdots,n \\ r_{ij}=\left(ar_j^{'}+\eta\right) \bmod b_i, j=1,2,\cdots,m \end{gathered} \tag{7}$$

步骤 6: 设置密钥 S;

$$\begin{gathered} S\leftarrow\{s_1,s_2,\cdots,s_n\} \\ 0\leqslant s_i\leqslant m,\ \ i=1,2,\cdots,n \end{gathered} \tag{8}$$

步骤 7: 据密钥 S 将真实模分量集 M 覆盖到冗余集 R 中, 得到大小为 $n\times m$ 的密文 $E(m_0)$;

$$\begin{gathered} E(m_0)\leftarrow E_1,E_2,\cdots,E_n \\ E_i\leftarrow\{e_{i1},e_{i2},\cdots,e_{im}\} \\ e_{ij}\leftarrow\begin{cases} r_{ij}, & j\neq s_i \\ m_i, & j=s_i \end{cases} \\ i=1,2,\cdots,n,\ \ j=1,\cdots,m \end{gathered} \tag{9}$$

在盲计算阶段, 云服务器可对密文进行加、减、乘以及幂运算, 得到计算结果。输入: 密文 $x=E(X)$、$y=E(Y)$、$z=E(Z)$, 输出: 计算结果 $\text{calculate}(x,y,z)$, 具体方法流程如下:

依据在模运算的基础上加法同态和乘法同态的性质及其扩展性质, 服务器端可对密文进行上述 calculate 运算(加、减、乘、幂)操作。假设明文为 X、Y、Z, 加密得到的密文为 $x=E(X)$、$y=E(Y)$、$z=E(Z)$, 则对其进行运算的到运算结果为 calculate (x,y,z)。

在解密阶段, 客户端可根据中国剩余定理算法, 依据模基和密钥将云服务端的计算结果还原成明文输入为服务器的计算结果, 模基 B, 密钥 S, 输出为: 还原后的计算结果 Decrypt(calculate(x,y,z)), 具体流程如下所示:

步骤 1: 客户端依据密钥 S 将真实模分量集 M(即密钥)对应的计算结果 d 取出;

$$d_i = \text{calculate}(X,Y,Z) \bmod b_i \quad i=1,2,\cdots,N \tag{10}$$

步骤 2: 利用中国剩余定理可解出原始数据的 f 运算结果 $f(X,Y,Z)$;

$$B_s = \prod_{i=1}^{N} \quad B_i = B_s / b_i \tag{11}$$

步骤 3: B_i^{-1} 是 B_i 在 $\mathbb{Z}_{b_i}$ 中的乘法逆元;

$$B_i B_i^{-1} = 1 \bmod b_i \tag{12}$$

步骤 4: 经过中国剩余定理处理后的结果为:

$$(\sum_{i=1}^{N} d_i B_i B_i^{-1}) \bmod B_s \tag{13}$$

步骤 5: 原始数据的 f 运算结果为:

$$Decrypt(x,y,z) = (\sum_{i=1}^{N} d_i B_i B_i^{-1}) \bmod B_s \tag{14}$$

步骤 6: 正负判断:

$$if\,(f(X,Y,Z) < 0.5\prod_{i=1}^{n} b_i): \quad f(X,Y,Z) = f(X,Y,Z) \tag{15}$$

$$else: f(X,Y,Z) = f(X,Y,Z) - \prod_{i=1}^{n} b_i$$

3.2.2 实例流程

1) 浮点数加法

本部分以变量 x=12.2, 变量 y=14.4, 客户端请求云服务器端盲计算 $x+y$ 为例, 详细介绍客户端进行的加密、解密操作以及云服务器进行的盲计算操作过程中数据的变化过程, 具体过程如图 1 所示。

在加密阶段, 放大的倍数 a=100, 对 x 增加扰动噪声为 3, 对 y 增加的扰动噪声为 2, 本示例设置模基中元素的个数为 3, 即 $B=\{b_1,b_2,b_3\}$, 其中 $b_1=233$, $b_2=239$, $b_3=241$, 冗余的个数设置为 3。对于明文 x, 生成的密文数据 $R_x=\{[58,8,26],[15,28,1],[8,19,18]\}$, 对于明文 y, 生成的密文数据 $R_y=\{[44,26,20],[14,8,9],[6,13,237]\}$, 客户端将 R_x 和 R_y 发送给云服务器端。

在盲计算阶段, 服务器对密文数据进行加法运算, 并依据模基 $B=\{b_1,b_2,b_3\}$ 进行取模运算, 最终得到密文下的计算结果为 $G=[102,34,46],[29,36,10],[14,32,14]$, 将计算结果发送给客户端。

在解密阶段, 客户端在收到云服务器在密文下的计算结果后, 依靠密钥 S 进行解密操作, 找到真实结果分量 $D=\{d_1,d_2,d_3\}$, 依据模基 B, 利用中国剩余定理对数据进行还原, 即可得到正确的经过放大的计算结果 2665。对 2665 去除噪声并缩小得到 26.60。

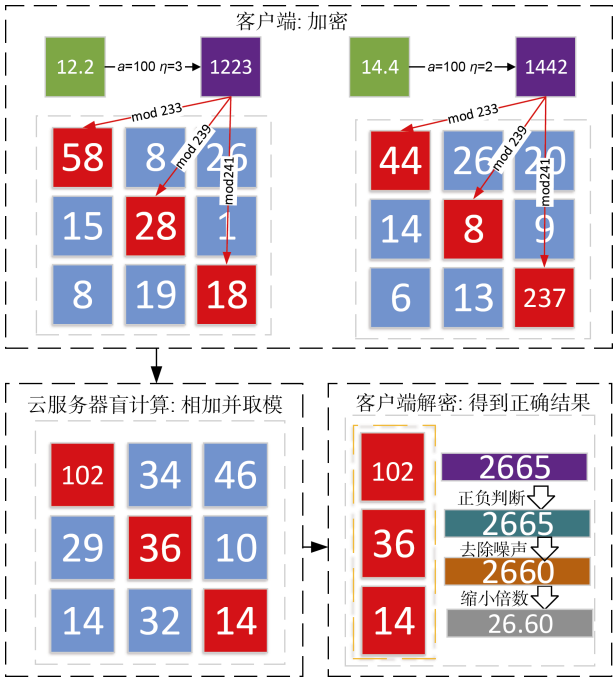


图 1 基于混淆模分解的同态加密方法盲加示例图

Figure 1 Example diagram of blind addition operation of homomorphic encryption method based on confusion modulus decomposition

2) 浮点数乘法

在机器学习等多项式算法中, 存在大量的乘法操作, 例如深度学习中的卷积操作, 卷积核的值和卷积之后的神经元参数往往都是浮点数。因此, 本部分以两个数据为例, 演示本方案的同态乘法运算。由于本方法是在整数域下进行运算的, 所以首先客户端使用放大和加扰的方式将浮点数数据放大成为整数数据, 再通过与上文中加同态相同的方法插入冗余并将密文发送给云服务器, 云服务器将执行盲计算的结果返回给客户端, 客户端得到返回的数据后, 首先进行正负判断, 再执行去噪和等比例缩小操作, 得到最终的计算结果, 如图 2 所示。

正负判断与噪声去除。正负判断的目的是在数据做减法等操作时可能会得到计算结果为负的情况, 而使用中国剩余定理解密出的解密都是正数, 因此需要对其进行正负判断, 因此预留一半的位置用来表示负数。与补码编码类似, 模分量同态加密下数的表示范围从 $[-0.5\prod_{i=1}^{n} b_i, 0.5\prod_{i=1}^{n} b_i]$ 转换成 $[0, \prod_{i=1}^{n} b_i]$, 即当 $\text{calculate}(X,Y,Z) \geq 0.5\prod_{i=1}^{n} b_i$ 时, 客户端可以判断该计算结果为负数, 需要对结果进行复原。在参数设置中本文设置 $\max(b_i) < \eta_{\max} << a$, 但为了方便展

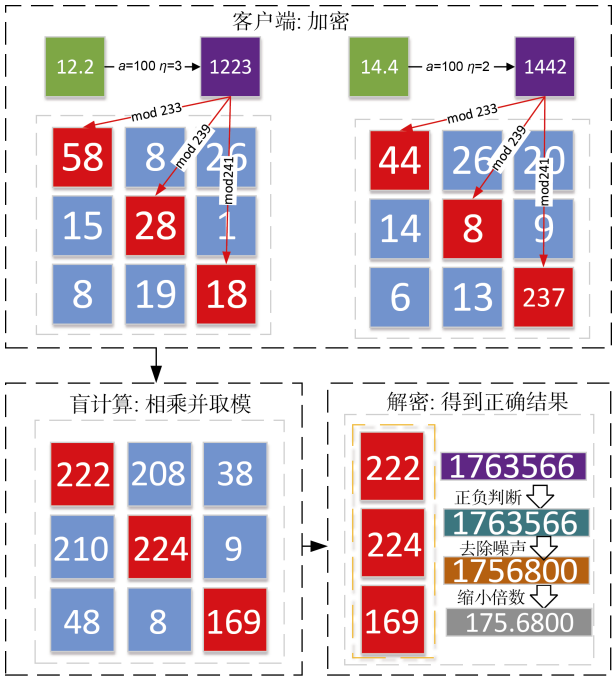


图 2　基于混淆模分解的同态加密方法盲乘示例图

Figure 2　Example diagram of blind multiple operation of homomorphic encryption method based on confusion modulus decomposition

示本节并没有严格遵循以上设置, 因此在解密时, 可以直接采用向下取整的方式舍去噪声, 如 $a=10^{10}$, η_1=1412, η_2=1232, $\eta_1\eta_2/a<1$, 因此可以在解密时可以对计算结果向下取整消去噪声。

3.3　基于密模聚合的同态加密方法

3.3.1　方法设计

本节提出一种适合单次加减盲运算的同态加密方法, 该方法能够在不降低安全性的前提下显著提高数据的处理效率。对于一个视频帧, 基于混淆模分解的同态加密方法是将单个像素加密成多个模分量再将加密数据发送到云服务器端进行运算, 而本方法是将视频图像的多个像素点作为模分量, 通过将这些模分量聚合成一个更大的数据, 该数据被称为最小答数。因此该方法可以被看作基于混淆模分解的同态加密方法的逆过程, 两者都依据模同态的性质来实现同态运算。由于模分量同态的特性, 服务器对最小答数的进行计算相当于对各个模分量(即原始数据)进行计算, 从而实现服务器对数据的盲处理。该方法适用于简单的视觉盲计算处理算法, 如前景提取中的背景相减算法, 图像差分算法, 边缘检测算法等。

方法主要分为以下三个部分: 加密、盲计算、解密, 详细的参数说明如表 2 所示。

表 2　基于密模聚合的同态加密方法参数说明

Table 2　Parameter description of homomorphic encryption method based on dense mode aggregation

参数名称	含义
n	明文像素的个数
M	明文像素值
B	模基, 被模的数

加密阶段, 客户端利用私钥将多个明文数据聚合成密文数据, 其中, 密钥为一组模基, 明文数据为像素值, 输入为 n 个明文像素 M, 秘钥 $\mathrm{pk}\leftarrow(B)$, 无公钥。输出为密文 $G(M)$, 具体流程如下所示(客户端加密)。

步骤 1: 生成聚合窗口, 由于本方法是将多个像素点合成为一个更大的加密数据, 因此将一组像素点组成的集作为原始数据 M 为一组待计算的像素点。

$$\begin{aligned} M &\leftarrow \{m_1,m_2,\cdots,m_n\} \\ 0 &\leqslant m_i \leqslant 255 (i=1,2,\cdots,n) \end{aligned} \tag{16}$$

步骤 2: 设置密钥 B, 根据模分量同态的性质可知, B 内元素 b_i 应大于 m_i 且为两两互素的正整数。

$$B\leftarrow\{b_1,b_2,\cdots,b_n\},(b_i,b_j)=1 \tag{17}$$

步骤 3: B_i^{-1} 是 B_i 在 $\mathbb{Z}_{b_i}$ 中的乘法逆元

$$B_iB_i^{-1}=1\bmod b_i \tag{18}$$

步骤 4: 依据中国剩余定理(CRT)生成最小答数 $G(M)$ 即为密文

$$\begin{aligned} G(M)&=(\textstyle\sum_{i=1}^{n} m_iB_iB_i^{-1}) \bmod B_s \\ B_s&=\textstyle\prod_{i=1}^{n} b_i \\ B_i&=B_s/b_i \end{aligned} \tag{19}$$

盲计算阶段: 云服务器在收到密文数据后, 对该密文数据进行单词盲加减运算, 输入为两个密文, 其中 $x=G(X)$、$y=G(Y)$, 输出为盲计算结果 calculate (x,y)。由于密文本身的大小足够掩盖原始信息, 因此在加密过程中不需要放大运算, 密文之间的计算都是在齐次下进行, 从而在盲计算阶段不需要考虑是否需要补齐的问题。由于该方法只使用于简单的加减运算, 服务器可以对密文进行有限次的加减运算操作。假设明文分别为 X、Y, 经过加密后得到的密文为 $x=G(X)$、$y=G(Y)$。其进行的运算得到的结果为, 其结果为一个新的答数。由于不需要进行取模运算, 服务器将计算结果直接发送给客户端。

解密阶段, 客户端在收到云服务器发送的计算

结果, 利用在加密阶段设置好的密钥对密文进行解密并进行正负判断, 具体解密流程如下所示(基于密模聚合的同态加密方法: 客户端解密)。

输入: calculate(x,y) 、密钥 B

输出: Decrypt(calculate(x,y)) $=N_i$, $i=1,2,\cdots,n$

解密: 客户端在收到计算结果 calculate(x,y) 后, 依据模基 B 对 $f(x,y)$ 执行解密操作, 并执行正负判断从而得到正确结果。

3.3.2 实例流程

本部分以两组四像素为模分量为例, 用 $X=\{94,36,21,45\}$, $Y=\{92,78,45,36\}$, 客户端要求云服务器端计算两组像素的差。数据经过客户端加密, 再经由服务器计算, 最后由客户端解密的变化过程示意图如 3 所示。

在加密阶段, 第一个聚合窗口 $X=\{94,36,21,45\}$, 第二个聚合窗口 $Y=\{92,78,45,36\}$, 密钥 $B=\{523,579,613,691\}$ X 与 Y 分别与密钥 B 通过中国剩余定理(CRT)模同态运算得出答数(即密文) $x=79748036409$, $y=110667041643$, 客户端将两个密文发送给云服务器端。

在盲计算阶段: 云服务端由于没有密钥 B 而无法将密文复原成明文, 服务器进行减法运算, 得到计算结果 $f(x,y)=-31249005234$, 并将计算结果反馈给客户端。

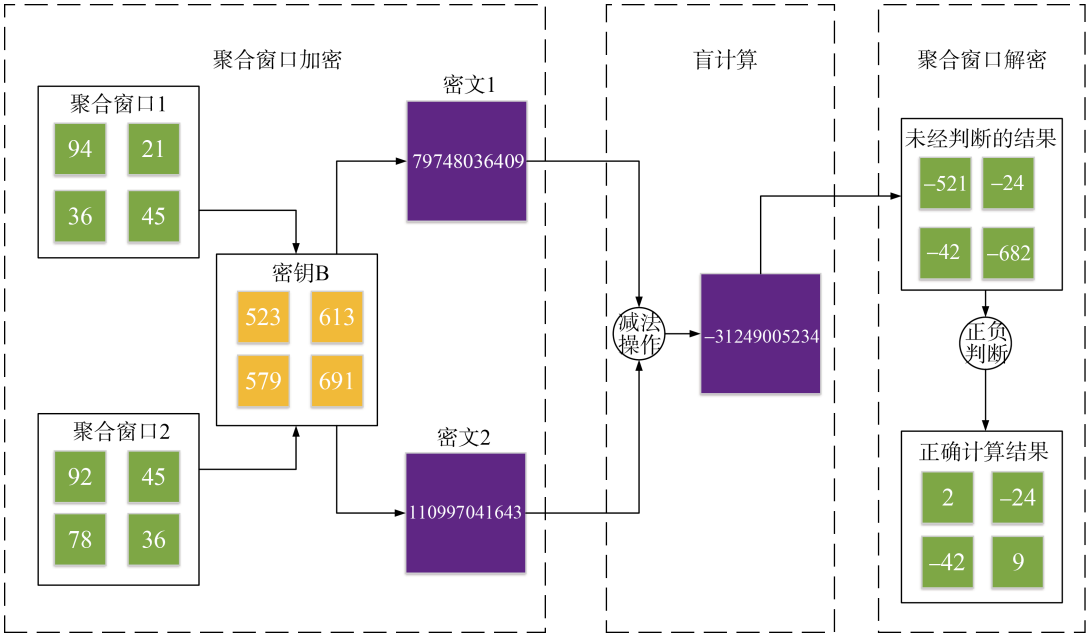


图 3 基于密模聚合的同态加密方法示例图

Figure 3 Example diagram of homomorphic encryption method based on dense module aggregation

在解密阶段: 用户单利用密钥(即模基 B)执行解密操作, 得到未经判断正负计算结果 $\{-521,-42,-24,-682\}$, 之后依据结果与模基之间的关系算法判断该结果是否为正, 得到最终计算结果 $N=\{2,-42,-24,9\}$, 该结果即为两个聚合窗口中对应位置做减法的值。

3.4 效率测试

3.4.1 实验环境

本文实验环境为一台普通 PC 机, 操作系统为 Windows10, 内存为 16G, 处理器为 2.6 GHz Intel(R)Core(TM) i7-6700HQ, IDE 为 Visual Studio 2015, 程序编程语言为 C++。

3.4.2 测试结果

本节测试本文提出的两种方法的加密效率, 其中方法 1 为基于混淆模分解的同态加密方法, 模基数为 20, 冗余数为 64, 方法 2 为基于密模聚合的模同态加密方法, 聚合窗口大小为 4, 测试结果如表 3 所示, 方法 3 为 IBM HElib 同态加密库[47], 方法 4

表 3 加密时间对比

Table 3 Encryption time comparison

图像尺寸	方法 1/s	方法 2/s	方法 3/s	方法 4/s
720*576	29.03	1.01	712.14	3690.39
350*320	5.95	0.21	125.89	732.12
100*00	0.82	0.04	31.20	159.28
20*20	0.03	0.002	7.23	32.09

为 Microsoft SEAL 同态加密库[48]。

3.5 安全性分析

私钥中包含的扩大倍数、模基组和真假位置模板是保障同态加密方法安全性的重要因素。若要恢复出秘密信息, 必须获取所有正确的模基和真假位置索引。公钥(包含扩大倍数和模基组, 不包括真假位置模板)在服务端, 无法还原原始的隐秘数据, 而客户端则使用私钥对这些秘密信息进行加密和解密。

为深入了解这种加密原理, 如图 4 所示, 以一张图片的加密过程为例。A 为原始图片的灰度值范围是[0, 255], BCDEF 为通过模(mod 43, 61, 140, 162, 178)运算后, 可以得到子图, 然而, 这样做仅将原始图片的取值范围压缩, 而不能有效地打乱原始图片的像素。为了增强安全性, 需要对每个像素加入一个不同的随机数以混淆原始像素。

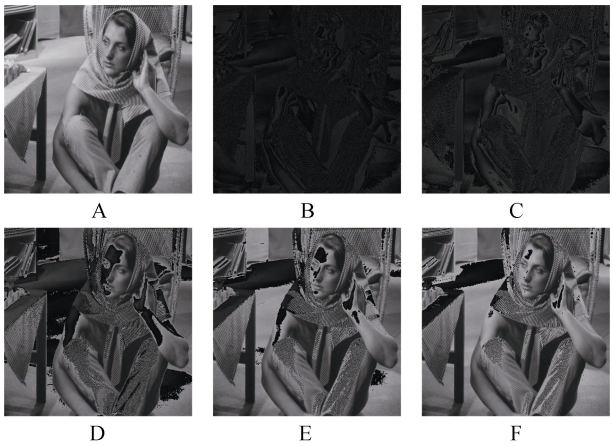


图 4 原始数据直接做取模运算的加密效果

Figure 4 Encryption effect of modulo operation directly on original data without amplification and scrambling

但是, 这种方式在解密后复原原始数据时会面临挑战。为解决这个问题, 我们先将原始数据扩大一定倍数, 再加上一个相对于扩大倍数极小的随机数, 解密时再除以相应的倍数, 如图 5 所示, A 为原始图片, BCDEF 为乘以一个大数, 再模(mod 43, 61, 140, 162, 178)运算后, 可以得到的子图。小随机数在解密阶段会被视作微小误差并被忽略。但如果乘法次数过多, 或者乘数过大, 误差就可能变得显著。

为了深入确认加密的安全性, 我们会在接下来的部分进行分析和证明, 当参数选择得当, 解密这些密文信息将面临巨大困难, 统计学攻击也无法轻易成功。

如果 η 遵循均匀分布 $U(0,\eta_{\max})$, 每个像素值做取模运算后的值设为 x, $x\in[0,b_i)$,令 $l=\lfloor \eta_{\max}/b_i \rfloor$, 则 $\eta'=\eta\%b_i$ 的分布为:

$$\Pr(\eta'=x)=\begin{cases} l+1/\eta_{\max}, x<\eta_{\max}\%b_i \\ l/\eta_{\max}, x\geqslant \eta_{\max}\%b_i \end{cases} \tag{20}$$

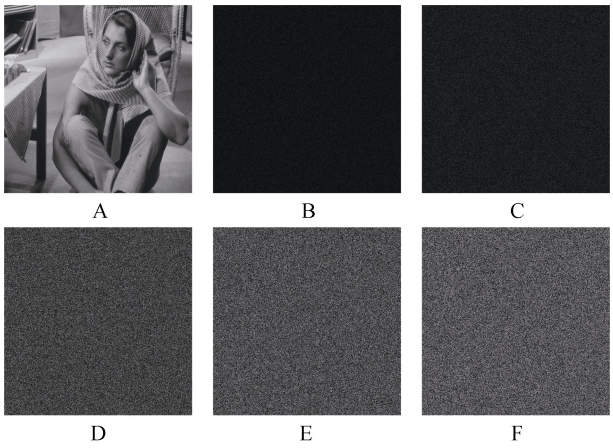


图 5 原始数据放大加扰后的加密效果

Figure 5 Encryption effect of original data after amplification and scrambling

理想情况下, $\eta_{\max}$ 为 b_i 的公倍数, 则 η' 完全符合均匀分布, 因此加密后的该像素也是均匀分布的, 原始帧中的数据不会被暴露。而如果 $\eta_{\max}$ 不是 b_i 的公倍数, 在经历放大加随机数并取模投影后的像素值 x, $\Pr\left(x\in[0,\eta_{\max}\%b_i)\right)>\Pr\left(x\in[\eta_{\max}\%b_i,b_i)\right)$, 即 x 稍微更倾向于处于某一特定区域。

因此, 本文对参数的作如下设置: $\max(b_i)<\eta_{\max}<<a$, 此外, $\eta_{\max}$ 需要尽可能的取最大值。这是因为, $\eta_{\max}$ 越大, l 就越大, $1/\eta_{\max}$ 的值就对应越小, 图像的像素分布就越均匀。

更重要的是, 本文对模投影增加了混淆操作, 即使用假的随机数来混淆真实的投影, 对于 N 个模基, M 个冗余的情况下, 鉴于服务端拥有有私钥的任何信息, 因此, 在不得到真实投影 S 的情况下, 通过穷举攻击, 想要得到 N 个正确的模投影的概率为 $1/M^N$。因此, 越大的 M 合 N 的值, 将导致攻击越困难。故, 当 M 和 N 设定为足够大时, 被穷举攻击成功的概率就越小。此外基于密模聚合的同态加密方法的安全性结合 4.1.6 分析。

4 视觉盲计算应用

4.1 基于密模聚合同态加密的运动目标盲提取方法

4.1.1 方法设计

本方法的目标是让拥有 ViBE 算法的服务器在

不得到客户端任何有效原始视频信息的条件下完成视频监控对象的分割。该方案身份对象分为两种: 客户端和云服务器端。客户端将监控摄像头捕获的视频帧进行图像预处理, 随后利用基于密模聚合的同态加密方法确定聚合窗口, 根据每个聚合窗口中像素点的个数选定相对应的模基并将其作为密钥, 利用基于聚合窗口的模同态加密算法将每个组中的像素点聚合成密文并将密文发送给云服务器端。云服务器端首先根据第一张密文帧的信息建立密文下的背景模型, 然后开始接收后续视频帧, 对后续视频帧和背景模型采取背景差值算法操作, 并将计算结果发送回客户端。客户端将云服务器发来的计算结果解密并将解密结果返回给云服务器。云服务器对解密信息进行阈值判断得到分割结果, 最后将分割结果发送给客户端。总框架如图 6 所示。

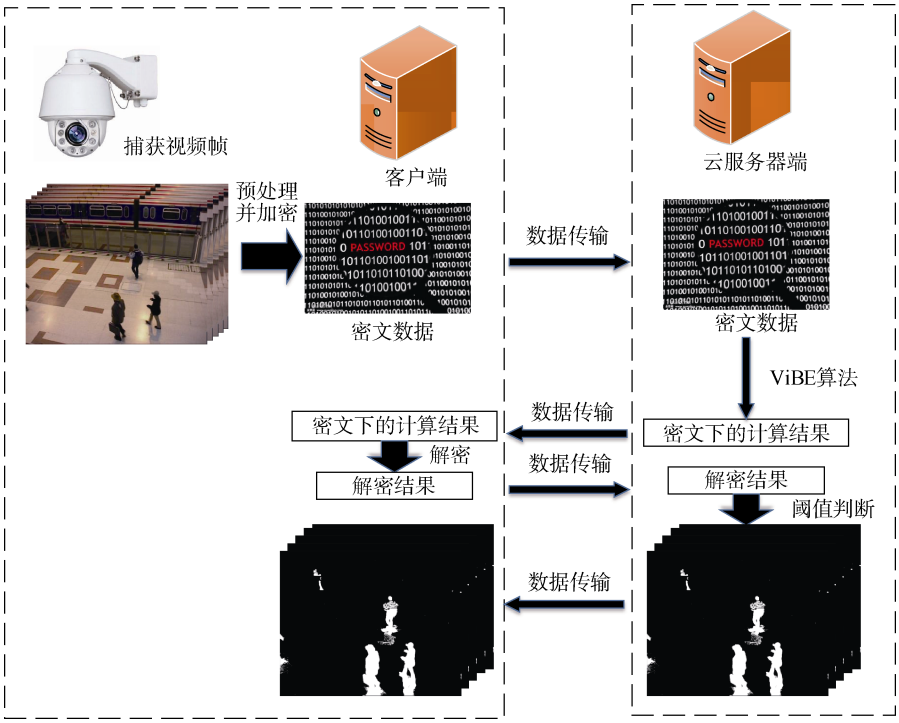


图 6　运动目标盲提取框架图

Figure 6　Frame diagram of blind extraction of moving objects

输入为客户端输入视频帧和云服务器端拥有运动目标提取算法, 输出云服务器端在密文下计算出运动目标提取结果, 算法流程如下所示(基于密模聚合的同态加密方法):

步骤 1: 客户端加密阶段, 客户端图像帧预处理, 对图像帧采取灰度图像转换、图像补齐操作;

$$Gray(x,y)=\left[R(x,y)\times 30+G(x,y)\times 59+B(x,y)\times 11\right]/100 \tag{21}$$

$Gray(x,y)$ 即为坐标为 $G=\{Gray(x,y)\,|\,x=1,2,\cdots,m,y=1,2,\cdots,n\}$ 的原始像素经过转化生成的一维灰度值, 其组成的集合即为灰度图:

$$G=\{Gray(x,y)\,|\,x=1,2,\cdots,m,y=1,2,\cdots,n\} \tag{22}$$

图像帧补齐操作为当 $m/n \bmod 2=1$ 时:

$$\begin{aligned} m&=m+1,Gray(m+1,y)=255\\ n&=n+1,Gray(x,n+1)=255 \end{aligned} \tag{23}$$

步骤 2: 客户端依据密钥 B 中元素的个数确定聚合窗口大小, 在本方案中, 聚合窗口大小为 $key=\{k_1,k_2,k_3,k_4|(k_i,k_j)=1,i,j=1,2,3,4,i\neq j,510\leqslant k_i\}$, 输入密钥为 key:

$$\begin{aligned} key=\{k_1,k_2,k_3,k_4|(k_i,k_j)=1\}\\ i,j=1,2,3,4,i\neq j,510\leqslant k_i \end{aligned} \tag{24}$$

步骤 3: 客户端数据加密并传输: 在加密过程中, 依照基于窗口聚合的模同态加密方法, 图像像素点被看成模分量, 选择的密钥 $(0.5m,0.5n)$ 作为模基, 生成的答数作为密文被服务器所处理。

对于一个已经转换为灰度图像的视频帧 $(0.5m,0.5n)$, 最后将尺寸为 $(0.5m,0.5n)$ 二维密文数组发送到云服务器端。

$$\begin{aligned} M\leftarrow\{m(x,y)\,|\,x=1,2,\cdots,0.5m,y=1,2,\cdots,0.5n\}\\ m(x,y)=enc(key,G_{x,y}) \end{aligned} \tag{25}$$

得到 $G_{x,y} \leftarrow \{G(2x-1,2y-1), G(2x,2y-1), G(2x-1,2y), G(2x,2y)\}$ 。

步骤 4: 云服务器端执行 ViBE 运动目标提取算法:

初始化模型: 对于密文位置 (x,y), 其八连通区域内取 8 个值用于初始化模型, 该位置的模型数据为:

$$V(x,y) = A_1, A_2, \cdots, A_8 \tag{26}$$

差值计算并传输: 使用第一帧密文建立样本模型后, 接收第二帧图像。对于密文位置 (x,y), 将 (x,y) 位置的新密文值 $m(x,y)$ 与 $V(x,y)$ 中的全部元素进行差值运算, 即可得到差值结果 $T(x,y)$, 之后将每个位置的差值结果组成的集合. 发送给客户端服务器并将该点的密文值存入模型数据 $V(x,y)$ 中。

$$T = \{T(x,y) \mid x = 1,2,\cdots,0.5m; y = 1,2,\cdots,0.5n\} \tag{27}$$

$$T(x,y) = \{A_1 - m(x,y), A_2 - m(x,y), \cdots, A_8 - m(x,y)\} \tag{28}$$

步骤 5: 客户端解密, 客户端依据密钥 *key* 结合基于密模聚合的同态加密方法对 T 进行解密操作, 得到明文结果 R, 并将结果返回给云服务端用于阈值判断;

$$R = \{R(x,y) \mid x = 1,2,\cdots,m; y = 1,2,\cdots,n\} = Dec(T, key) \tag{29}$$

$$R(x,y) = \{r_{(x,y)1}, r_{(x,y)2}, \cdots, r_{(x,y)8}\} \tag{30}$$

步骤 6: 云服务器端执行阈值判断并返回提取结果, 对 $R(x,y)$ 中的每个元素与阈值进行比较, 统计小于阈值的个数, 设为 #, 当 # 大于规定值 #*min* 时, 判定该像素点为背景点, 否则为前景点, 对每个点执行相同的判断即可得到整张图的运动目标提取结果, 依概率更新模型并将该运动目标提取的结果返回给客户端。

参数设置: 聚合窗口的大小的设置应合理, 据模分量聚合的模同态加密方法可知, 几个模分量可以依据中国剩余定理聚合成密文, 理论上每个聚合窗口中的元素个数为 $n(2 \leqslant n)$, 但是选定的 n 每增加一个, 答数数据的大小都会指数被变大, 以 $n=2$ 为例, 聚合窗口中的像素元素分别为 $\{98,127\}$, 密钥为 $\{676,677\}$, 得到的密文结果为 438146, 当 $n=4$ 时, 聚合窗口中的像素元素为 $\{98,127,39,169\}$, 密钥变为 $\{676,677,701,705\}$, 密文结果变为 93026902534, 扩大近 212319.41 倍, 因此, 选取的 n 不应过大, 本文选取 n 的范围为 $[2,8]$。在本方案中, 选定即一张尺寸为 256×256 的视频图像帧, 其转换为密文矩阵的尺寸为 128×128 。

密钥的设定范围: 根据任意两个像素进行图像差分算法时, 由于灰度图像像素范围是 $[0,255]$, 所以任意两个像素之间差为 $[-255,255]$, 因此选定的密钥的值应大于 510, 密钥中子密钥的个数与聚合窗口中模分量的个数 n 相等, 且两两互素。在本文中选择聚合窗口对应密钥为取值范围是 $[511,1000]$ 的两两互素的数。

初始化模型: 对应云端服务器接收加密信息后, 云端服务器对背景模型进行初始化操作, ViBE 算法建立背景模型只需要一帧图像, 即使用单帧视频序列初始化背景模型。将视频的第一帧作为背景模型的同时, 算法也将该帧中每一个像素点周围随机取多个像素点, 填充该像素点的样本集, 这样样本集中就包含了像素点的时空分布信息。对于密文下的运算, 选择它的相邻位置的密文得到它的样本模型。对于密文位置 (x,y), 其八连通区域内取 8 个值 $V(x,y) = A_1, A_2, \cdots, A_8$, $V(x,y)$ 即为密文位置 (x,y) 的样本模型。在每个密文点都建立点背景模型即可得到整体背景模型 V 。为了使背景模型能够适应背景的不断变化, 比如光照, 背景物体的变更等, 每个像素点的样本模型都有一定的概率进行更新。

差值策略: 本部分将 ViBE 的交集策略改变为差值策略, 由于本方法利用的聚合窗口将的多个像素合成一个大答数, 该数值较大, 因此如果按照交集策略(即两个数值相减为 0), 则无法实现匹配。改变成差值策略后, 云服务计算出密文下的差值结果, 该结果经过明文解密, 返回给云服务器, 如果差值在阈值范围之内则视为匹配成功, 如果差值超出阈值则视为匹配失败。

4.1.2 实验环境

本文实验环境为一台普通 PC 机, 操作系统为 Windows10, 内存为 16G, 处理器为 2.6 GHz Intel(R)Core(TM) i7-6700HQ, IDE 为 Visual Studio 2015, 程序编程语言为 C++。

4.1.3 运动目标提取结果

测试设定聚合窗口大小为 4, 密钥为 $\{676,677,701,705\}$, 明文帧 a 的尺寸为 720×576, 明文帧 b 的尺寸为 320×240, 明文帧 c 的尺寸为 320×240 。视频帧和对应云服务器计算的运动目标提取结果如图 7 所示。

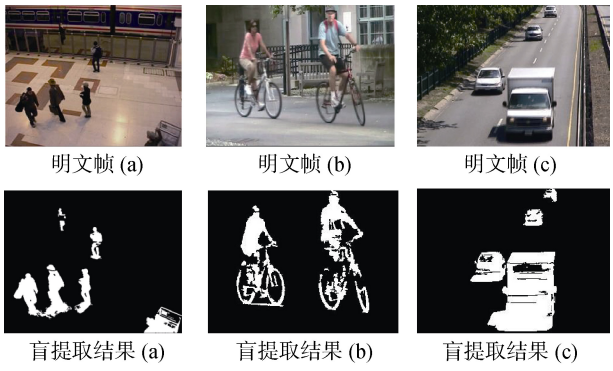


图 7　运动目标盲提取结果图

Figure 7　Blind extraction result map of moving target

4.1.4　运动目标提取准确率对比

本论文测试运动目标提取准确率使用的数据集为 CDW-2014, 该测试数据集包含 12 个视频类别(基线, 动态背景, 相机抖动, 间歇物体, 运动, 阴影, 热, 恶劣天气, 低帧率, 夜视, PTZ, 湍流), 每个类别有 4~6 个视频序列, 每个单独的视频文件(.zip 或.7z)都可以单独下载, 方便进行对比测试。本文挑选 5 个视频序列作为本文的测试集, 将本文提出的基于聚合窗口的模同态运动目标提取方法、Jin 等人[49]提出的基于混沌的混合高斯模型分割法、Jin 等人[50]提出的多服务器秘密共享的前景提取方法(PPViBE)、Barnich[51]提出的基于随机技术的目标提取方法及 Chu 等人[9]提出的运动目标的实时检测方法进行准确率对比测试, 以上方法均基于隐私保护的运动目标盲提取方法, 测试结果如表 4 所示。

为了进一步测试本文提出的方法与其他运动目标提取方法的准确率对比, 为了排除随机视频帧对实验结果的影响, 本文对整段视频的运动目标提取准确率进行测试, 测试结果如表 5 所示。

表 4　单帧运动目标提取算法对比

Table 4　Comparison of single frame moving target extraction algorithms

视频/视频帧	[49]/%	[50]/%	[51]/%	[9]/%	本文/%
copyMachine/1338	94.92	97.95	97.95	95.23	97.95
Corridor/666	95.84	95.74	95.74	95.53	95.74
Park/552	83.74	85.44	85.44	82.95	85.44
pepleInShade/47	93.85	91.12	91.12	93.36	91.12
winteDriveway/9	76.21	78.22	78.22	75.16	78.22

表 5　整段视频运动目标提取算法对比表

Table 5　Comparison table of moving object extraction algorithms for the whole video

视频/视频帧	[49]/%	[50]/%	[51]/%	[9]/%	本文/%
copyMachine	90.63	92.83	92.83	90.75	92.83
Corridor	90.35	90.77	90.77	89.56	90.77
Park	82.46	83.05	83.05	83.65	83.05
peopleInShade	90.43	88.93	88.93	91.56	88.93
winteDrriveway	74.23	75.02	75.02	72.03	75.02

4.1.5　运动目标提取对比

本节将本文提出的基于密模聚合同态加密的运动目标盲提取方法与基于监督优化的提取方法(明文)[52]、多服务器秘密共享的前景提取方法(PPViBE)以及基于混合高斯模型的运动目标盲提取方法进行对比, 在最后一列展示标准正确结果(Groundtruth), 选用的测试集为 CDW-2014 中的 7 个测试数据: Backdoor, Busstation, Cubicle, Highway, Office, Pedestrians, Sofa。测试结果如图 8 所示。

4.1.6　安全性分析

1) 抵抗统计攻击能力

统计为使用统计手段获取密文的分布特性, 从而缩小密文搜索量, 提高攻击效果的密码攻击。直方图是一个重要的统计特征, 是一个离散的函数, 它表示图像中每个灰度级与该灰度级出现次数的对应函数, 用来描述图像中像素的灰度级分布。加密图像的直方图应该是平坦的, 否则一些信息可能会被泄漏, 导致可以通过分析密码统计特性来进行密码攻击, 即统计攻击。本文将原始数据统计图与使用基于密模聚合同态加密的运动目标盲提取方法加密后的密文数据统计信息进行对比, 对比结果如图 9 所示, 测试结果表明, 本方法加密后的数据具有较强的抵抗统计攻击能力。

2) 抵抗攻击能力分析

本文选取的聚合窗口的大小为 2×2, 即将每四个像素作为模分量聚合成一个答数, 该答数即为密

图 8　对象提取结果对比图

Figure 8　Object extraction result comparison chart

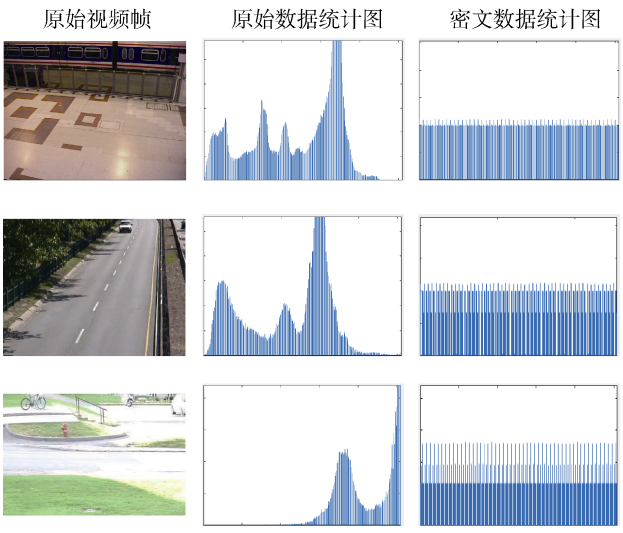


图 9　统计分析结果对比图

Figure 9　Comparison chart of statistical results

文, 无论聚合窗口的尺寸大小为多少, 其加密后的结果均为一个单值密文, 攻击者无法确定从密文中获取聚合窗口大小的信息。对于密钥攻击, 本文设置四个两两互素取值范围为$\left[511,2^{64}\right]$的整数, 如果攻击者在已知聚合窗口尺寸, 则同时攻击成功四个密钥的概率为$1/2^{256}$, 而且在通常情况下, 攻击者无法破解聚合窗口的大小以及密钥的范围, 所以使得该攻击成为不可能, 所以本文提出的基于密模聚合同态加密的运动目标盲提取方法能满足安全性需求。

4.2　基于混淆模分解同态加密的人脸检测方法

4.2.1　方法设计

为了解决人脸检测中隐私泄露的问题, 本文设计并提出一种盲化版本的人脸检测方法, 在经典的 V&J 人脸检测算法基础上结合第三节提出的基于混

淆模分解同态加密的人脸检测方法, 实现了云服务器端对客户端视频帧数据中人脸的盲检测, 如图 10 所示。

基于混淆模分解同态加密的人脸盲检测方法利用混淆模分解的乘同态性质实现图像的盲卷积运算, 最终完成整个人脸盲检测流程。客户端确定模基和冗余的数量, 生成两者乘积个数的模分子图, 每张子图中的数据是被加密的。客户端将模分子图发送到云服务器端, 云服务器端对每一个模分子图进行卷积运算, 并引入随机数保护 V&J 人脸检测分类器权值。客户端在收到云服务器发送的经过卷积运算的结果后, 利用基于混淆模分解的模同态算法将卷积结果复原, 为复原结果生成多个冗余并将冗余发送给服务器端进行阈值判断, 并将阈值判断结果返回给客户端。客户端将根据真实结果的位置将真正的阈值判断结果发送给客户端。通过以上多个若分类器的阈值积累, 云服务器端将最终的人脸位置框发送给客户端, 完成人脸盲识别。该方法的优势在于: 客户端中包含大量人脸隐私的视频帧信息不会暴露在服务器中, 服务器中有价值的分类器权值信息也不会暴露给客户端。基于混淆模分解的模同态人脸盲检测方法的步骤描述如下所示。

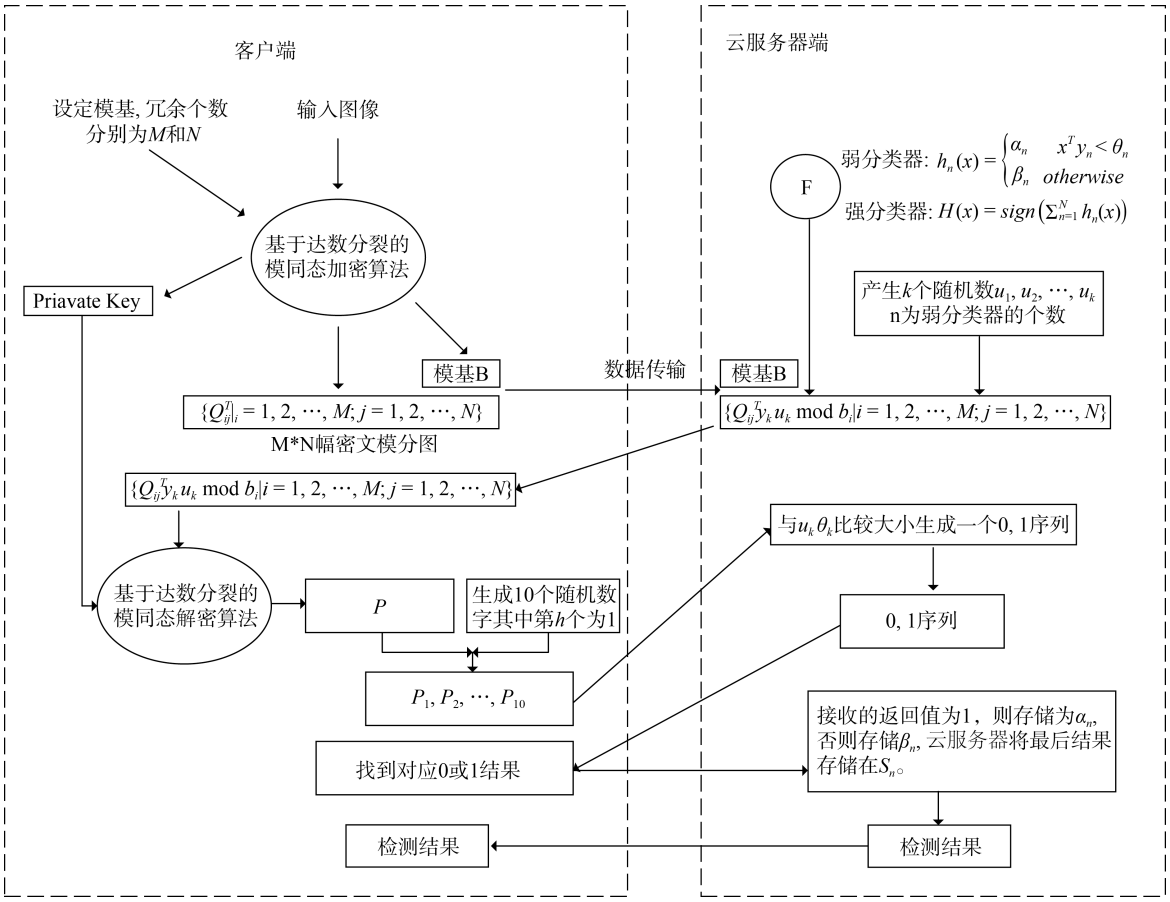


图 10 人脸盲检测方法流程框架图

Figure 10 Flow chart of face blindness detection method

输入:

①客户端输入视频帧 X

②云服务器端拥有强分类器: $H(x)=sign\left(\sum_{n=1}^{N}h_n(x)\right)$

其中: $h_n(x)=\begin{cases}\alpha_n & x^T y_n<\theta_n\\ \beta_n & otherwise\end{cases}$

输出:

①客户端只知道 $H(x)$ 的检测结果, 不知道 $H(x)$ 的任何参数。

②云服务器端只知道 $H(x)$ 的检测结果, 不能计算出图像 X 的完整信息

步骤 1: 客户端设定模基中元素数 M, 冗余数 N;

步骤 2: 客户端利用基于混淆模分解的模同态生成算法将原始图像 X 加密成 $M\times N$ 幅密文模分图 $\left\{Q_{ij}^T \mid i=1,2,\cdots,M; j=1,2,\cdots,N\right\}$ 并得到私钥 Private Key 和公钥 Public Key;

客户端将密文模分图 $\left\{Q_{ij}^{T} \mid i=1,2,\cdots,M;\ j=1,2,\ \cdots,N\right\}$ 和公钥 Public Key 发送给云服务器端;

步骤 3: 云服务器端依据加密模分图的大小, 计算出 H 个检测窗口。

步骤 4: 对于 $h=1,\cdots,H$ 检测窗口, 客户端和云服务器端进行以下子步骤:

(1) 对于一个检测窗口, 每一幅模分图都要经过 $k=1,2,\cdots,K$ 个弱分类器的分类, 客户端和云服务器端进行以下子步骤:

(a) 对于一幅模分图像, 该图像可以表示为 $\left\{Q_{ij}^{T} \mid i=1,2,\cdots,M; j=1,2,\cdots,N\right\}$, 强分类器中弱分类器的特征权值为 y_k, 云服务器端生成 k 个数值全部为且不大于 20 的随机数 $u_1,u_2,\cdots,u_k$。对于每一个弱分类器 y_k, 云服务器端计算当前窗口下的特征值结果为 $\left\{Q_{ij}^{T} y_k u_k \bmod b_i \mid i=1,2,\cdots,M; j=1,2,\cdots,N\right\}$。云服务器将 $M\times N$ 幅模分子图的特征值发送给客户端;

(b) 客户端计算图像 X 特征值 P;

(c) 客户端生成 10 个随机数据, 其中第 h 项为 1, 利用特征值 P 与随机数乘积生成冗余特征值 $P_1,P_2,\cdots,P_{10}$。

(d) 客户端将 $P_1,P_2,\cdots,P_{10}$ 发送给云服务器端, 云服务器端比较 $P_1,P_2,\cdots,P_{10}$ 和弱分类器的阈值 $u_k\theta_k$ 的大小, 大于 $u_k\theta_k$ 则保存为 1, 反之为 0。得长度为 10 的向量 $\boldsymbol{C}$。

(e) 云服务器端将 $\boldsymbol{C}$ 发送给客户端, 客户端将第 h 项的值返回给云服务器端, 若云服务器端接收的返回值为 1, 则记录结果为, 否则记录结果 β_n, 云服务器端将全部的记录结果存储在 S_n。

(2) 云服务器端比较 S_n 和强分类器的阈值的大小, 若 S_n 大于该阈值, 则该检测窗口被认定为检测结果包含人脸, 若 S_n 小于该阈值, 则该检测窗口被认定为不含人脸。

(3) 待全部窗口都检测完毕, 云服务器将全部包含人脸的窗口位置返回给客户端, 完成整个人脸盲检测。

与基于 OT 协议的人脸安全检测算法相比, 基于混淆模分解同态加密的人脸盲检测方法有如下优势: 首先, 基于混淆模分解同态加密的人脸盲检测方法中客户端将原始视频帧加密成多个模分子图, 能够有效保护客户端视频帧的隐私信息, 并且依旧可以利用 V&J 算法中积分图来提高 Haar 特征的计算速度。而基于 OT 协议的人脸安全检测算法不支持使用积分图加速。其次, 点积安全运算协议的基础是 OT 协议, 该协议拥有大量的加解密操作, 其计算复杂度巨大。对于视频帧图像来说, 对单像素进行 1 次 OT 安全操作需要进行 256 次 RSA 加密和 1 次对称加密。基于混淆模分解的同态加密方法不需要进行任何 OT 操作, 取而代之的是引入随机数来保护每个弱分类器中的特征向量隐私, 从而使得其不需要多次的加解密。最后, 在特征值和阈值的比较操作中, 本方法引入多个随机数构成一维随机数组, 该数组中只有一项数值为 1, 客户端将特征值分别与该数组中的每一个元素相乘并发送给云服务器, 云服务器收到的是一组混淆特征值却无法判断哪一特征值为真。随着随机数个数的增加, 云服务猜到真实特征值的概率就月底, 整个双盲检测方案的安全性就越高。从而可以看出基于混淆模分解同态加密的人脸盲检测方法计算速度快、效率高, 能够更好地保护客户端和云服务器端视频帧数据和分类器参数的隐私, 使得人脸盲检测走向实际应用成为了可能。

4.2.2 安全性分析

基于混淆模分解同态加密的人脸盲检测方法是保障客户端视频帧数据和云服务器端分类器参数即两方的数据不被泄露, 因此从客户端和云服务器端两方来分析整个人脸盲检测方法的安全性。

1) 从客户端到云服务器端

(1) 步骤 2 中客户端将原始视频帧加密成 $M\times N$ 张模分子图密文数据, 如果云服务器想依据 $M\times N$ 张模分子图复原出明文数据需要将密钥即真实模分量的位置模板破解, 则破解出一个模分量真实位置的概率为 $1/N$, 全部破解出的概率为 $(1/N)^M$, 本文设置的 M 为 20, N 设置为 64, 则模分量全部被攻破的概率为 $1/2^{120}$, 假设可以制造一部可以在 1s 内破解 DES 密码的机器, 那么使用这台机器破解本方法的密钥需要大约 0.582 亿万年的时间, 因此攻击者无法用穷举攻击等暴力破解方法获取密钥并将原始数据破解。

(2) 步骤 4, (c)(d)(e)中客户端将特征数据 P 发送给云服务器端, 客户端引入多个随机数(本文以 10 为例)并将该特征值分别与该随机数组中的元素相乘, 云服务器收到 10 个其无法判断真假的特征值 $P_1,P_2,\cdots,P_{10}$。云服务器端将 $P_1,P_2,\cdots,P_{10}$ 分别与阈值进行比较得到一个长度为 10 的 0,1 向量 $\boldsymbol{C}$ 并将该向量返回给客户端, 客户端将真实特征值 P 对应的 0

或 1 发回给云服务器端。云服务器端猜到正确的 P 值的概率与 $\boldsymbol{C}$ 中 0 或 1 的个数有关。$\boldsymbol{C}$ 中真值 P 对应 0 或者 1 的个数等于 q，则云服务器端能正确猜出模分子图权值的概率是 $(1/q)^{C_N^{1280}}$，N 代表弱分类器的个数，本文实验所用 N 的大小超过 2000，由此可以看出破解正确权值的概率极低。所以云服务器端分辨正确模分子图权值，从而不能恢复原图像。

2) 从云服务器端到客户端

(1) 步骤 4 中第(1)步(a)中，若一个弱分类器的特征权值为 y_n，则云服务器端生成 n 个随机数 $b_1,b_2,...,b_N$ 与该特征权值乘积得到是 y_nb_n，因此客户端不能计算出 y_n 的值。

(2) 步骤 4 中第(1)步(b)中，云服务器端把 $\boldsymbol{C}$ 发送给客户端，客户端无法确定准确的阈值 θ_n 大小，V&J 人脸检测算法对训练出来的阈值 θ_n 精确度要求极高，估算 θ_n 的范围无法获得准确的检测结果。

从安全性分析上看，客户端和云服务器端双方的隐私信息都得到有效的保护。基于混淆模分解的模同态人脸盲检测方法从理论上达到了保护客户端和云服务器端隐私的效果。

4.2.3 实验环境与数据集

实验环境为一台普通 PC 机，操作系统为 Windows10，内存为 16G，处理器为 2.6 GHz Intel(R)Core(TM) i7-6700HQ, IDE 为 Visual Studio 2015，程序编程语言为 C++。本文使用的分类器为 OpenCV 自带的 haarcascade_frontalface_alt.xml，测试集为 FDDB 数据集和 Caltech10k Web Faces 数据集。

4.2.4 实验结果

FDDB 数据集包含 2845 张图片共 5171 张人脸，是人脸检测经典测试集，基于混淆模分解的模同态人脸盲检测方法人脸盲检测结果如图 11 所示。Caltech10k Web Faces 数据集包含通过在 Google 图片搜索中键入常用的给定名称从网上收集的人的图像。数据集包含 10524 张不同分辨率的包含人脸的图片。对该测试集的人脸盲检测结果如图 12 所示。

4.2.5 准确率对比

FDDB 数据集和 Caltech10k Web Faces 数据集中各随机选择 300 张图片共 723 张人脸作为准确率对比测试集，使用基于混淆模分解同态加密的人脸检测方法对该测试集进行测试，并以原始 V&J 人脸检测算法作为对照检验本方法是否对原版算法准确率造成影响，具体检验结果如表 6 所示。

从表中实验结果可以看出，两种算法的检测结果相同，待检测人脸的个数为 627 个，正确检测人脸的个数为 611 个，检测错的人脸个数为 16 个，准确率为 84.51%，漏检率等于 15.49%。

图 11 FDDB 数据集下的盲检测结果图

Figure 11 Blind detection result graph under FDDB data set

图 12 Caltech10k 人脸盲检测结果图

Figure 12 Face Blindness Detection Results of Caltech10k

表 6 人脸盲检测方法与原始检测算法准确率对比

Table 6 Comparison of accuracy between face blindness detection method and original detection algorithm

检测算法	检测人脸数	正确人脸数	错检数	准确率(%)	漏检率(%)
V&J	627	611	16	84.51	15.49
本方法	627	611	16	84.51	15.49

4.2.6 效率测试

图像测试集是 FDDB 人脸数据集中的 80 幅图片，选取不同分辨率下的图像各 10 张，一共 80 张待检测图像组成测试数据集。使用本文提出的人脸盲检测方法检测该测试数据集，计算不同分辨率下检测所消耗的时间。基于混淆模分解同态加密的人脸检测方法加密时间和检测时间如表 7 所示。

时间对比实验所用到的测试集为 FDDB 数据集中随机选取的 100 张尺寸为 100×100 的图像。分别使用 V&J 人脸检测方法、基于 OT 协议的人脸安全

检测方法(明文)、基于随机子图的隐秘人脸检测方法(简称 RSI 方法[12])、基于随机矩阵的隐秘人脸检测方法(简称 RM 方法[53])对比检测 100 张人脸图片所消耗的时间, 并计算平均检测一张图片消耗的时间, 实验结果如表 8 所示。

表 7　人脸盲检测方法与原始检测方法准确率对比

Table 7　Comparison of accuracy between face blindness detection method and original detection algorithm

尺寸	加密时间/s	人脸盲检测时间/s
90×90	0.65	42.16
100×100	0.82	76.09
110×110	1.00	93.80
120×120	1.24	102.23
144×144	1.61	146.26
196×196	2.93	386.72
260×196	3.80	549.95
300×300	6.90	1214.5

表 8　人脸盲检测方法准确率对比

Table 8　Comparison of accuracy between face blindness detection method and original detection algorithm

方法名称	检测时间/s
V&J 人脸检测方法	0.38
基于随机矩阵的隐秘人脸检测方法(RM 方法)	288.73
基于 OT 协议的人脸安全检测方法(OT 方法)	>72000
基于随机子图的隐秘人脸检测方法(RSI 方法)	25.56
本方法	23.05

5　结论

本文提出一种基于模分量的同态加密算法, 以该算法为基础设计基于密模聚合的同态加密方法和基于混淆模分解的同态加密方法。依据 ViBE 算法和基于密模聚合同态加密的设计运动物体盲提取方法, 依据 V&J 人脸检测算法和基于混淆模分解的同态加密方法设计人脸盲检测方法。未来可以将本文提出的基于模分量的同态加密算法应用在更多需要隐私保护的领域, 如文档盲处理、人脸盲识别等。

参考文献

[1] Shen C X, Zhang H G, Feng D G, et al. Overview of Information Security[J]. *Science in China* (*Series E* (*Information Sciences*)), 2007, 37(2): 129-150.
(沈昌祥, 张焕国, 冯登国, 等. 信息安全综述[J]. *中国科学(E 辑: 信息科学)*, 2007, 37(2): 129-150.)

[2] Ruibin Z. Global cyberspace game intensifies [J]. *Financial and Economic circles*, 2017 (13): 89-91.
(赵睿斌. 全球网络空间博弈加剧[J]. *财经界*, 2017(13): 89-91.)

[3] Li X D, Jin X, Zhou B, et al. Recent Advances on Blind Vision[J]. *Science & Technology Review*, 2018, 36(17): 68-74.
(李晓东, 金鑫, 周彬, 等. 视觉盲计算技术研究进展[J]. *科技导报*, 2018, 36(17): 68-74.)

[4] Avidan S, Butman M. Blind Vision[M]. *Computer Vision – ECCV 2006*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006: 1-13.

[5] Viola P, Jones M. Robust Real-Time Face Detection[C]. *Proceedings Eighth IEEE International Conference on Computer Vision*, 2002: 747.

[6] Atallah M J, Du W L. Secure Multi-Party Computational Geometry[M]. *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001: 165-179.

[7] Ishai Y, Kilian J, Nissim K, et al. Extending Oblivious Transfers Efficiently[C]. *Annual International Cryptology Conference*, 2003: 145-161.

[8] Upmanyu M, Namboodiri A M, Srinathan K, et al. Efficient Privacy Preserving Video Surveillance[C]. *2009 IEEE 12th International Conference on Computer Vision*, 2010: 1639-1646.

[9] Chu K Y, Kuo Y H, Hsu W H. Real-Time Privacy-Preserving Moving Object Detection in the Cloud[C]. *The 21st ACM international conference on Multimedia*, 2013: 597-600.

[10] Chu C T, Jung J, Liu Z C, et al. STrack: Secure Tracking in Community Surveillance[C]. *The 22nd ACM international conference on Multimedia*, 2014: 837-840.

[11] Bost R, Popa R A, Tu S, et al. Machine Learning Classification over Encrypted Data[C]. *Proceedings 2015 Network and Distributed System Security Symposium*, 2015: 4324.

[12] Jin X, Yuan P, Li X D, et al. Efficient Privacy Preserving Viola-Jones Type Object Detection via Random Base Image Representation[C]. *2017 IEEE International Conference on Multimedia and Expo*, 2017: 673-678.

[13] Rivest R L, Adleman L, Dertouzos M L. On data banks and privacy homomorphisms[J]. *Foundations of secure computation*, 1978, 4(11): 169-180.

[14] Rivest R L, Shamir A, Adleman L. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems[J]. *Communications of the ACM*, 1978, 21(2): 120-126.

[15] Elgamal T. A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms[J]. *IEEE Transactions on Information Theory*, 1985, 31(4): 469-472.

[16] Hsieh P G, Ou C Y. Shape of Ground Surface Settlement Profiles Caused by Excavation[J]. *Canadian Geotechnical Journal*, 1998, 35(6): 1004-1017.

[17] Paillier P. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes[M]. Advances in Cryptology — EUROCRYPT '99. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007: 223-238.

[18] Kuribayashi M, Tanaka H. Fingerprinting Protocol for Images Based on Additive Homomorphic Property[J]. *IEEE Transactions on Image Processing*, 2005, 14(12): 2129-2139.

[19] Gentry C. Fully Homomorphic Encryption Using Ideal Lattices[C]. *The forty-first annual ACM symposium on Theory of computing*,

2009: 169-178.

[20] Brakerski Z, Vaikuntanathan V. Efficient Fully Homomorphic Encryption from (Standard) LWE[C]. *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, 2011: 97-106.

[21] Gentry C, Sahai A, Waters B. Homomorphic Encryption from Learning with Errors: Conceptually-Simpler, Asymptotically-Faster, Attribute-Based[C]. *Annual Cryptology Conference*, 2013: 75-92.

[22] Jin X, Zhang H Y, Li X D, et al. Confused-Modulo-Projection-Based Somewhat Homomorphic Encryption—Cryptosystem, Library, and Applications on Secure Smart Cities[J]. *IEEE Internet of Things Journal*, 2021, 8(8): 6324-6336.

[23] Lipton A J, Fujiyoshi H, Patil R S. Moving Target Classification and Tracking from Real-Time Video[C]. *Proceedings Fourth IEEE Workshop on Applications of Computer Vision*. *WACV'98* (*Cat. No.98EX201*), 2002: 8-14.

[24] Jain A K, Zhong Y, Lakshmanan S. Object Matching Using Deformable Templates[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, 18(3): 267-278.

[25] Behrad A, Shahrokni A, Motamedi S A, et al. A robust vision-based moving target detection and tracking system[C]. *University of otago, 2001*.

[26] Yang W, Zhang T W. A New Method for the Detection of Moving Targets in Complex Scenes[J]. *Journal of Computer Research and Development*, 1998, 35(8): 724-728.
(杨威, 张田文. 复杂景物环境下运动目标检测的新方法[J]. *计算机研究与发展*, 1998, 35(8): 724-728.)

[27] Seki M, Fujiwara H, Sumi K. A Robust Background Subtraction Method for Changing Background[C]. *Proceedings Fifth IEEE Workshop on Applications of Computer Vision*, 2002: 207-213.

[28] Horn B K P, Schunck B G. Determining optical flow[C]. *Techniques and Applications of Image Understanding. International Society for Optics and Photonics*, 1981, 281: 319-331.

[29] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[J]. 1981, 81(3):674-679.

[30] McFarlane N J B, Schofield C P. Segmentation and Tracking of Piglets in Images[J]. *Machine Vision and Applications*, 1995, 8(3): 187-193.

[31] Lee B , Hedley M . Background estimation for video surveillance[C]. *IVCNZ02*, 2002: 315-320.

[32] Stauffer C, Grimson W E L. Adaptive Background Mixture Models for Real-Time Tracking[C]. *1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*Cat. No PR00149*), 2002: 246-252.

[33] McKenna S J, Jabri S, Duric Z, et al. Tracking Groups of People[J]. *Computer Vision and Image Understanding*, 2000, 80(1): 42-56.

[34] Song Z Q, Lu J Z, Nie S L. Improved Spatiotemporal Background Subtraction Method for Target Detection[J]. *Opto-Electronic Engineering*, 2016, 43(2): 27-32, 39.
(宋志勤, 路锦正, 聂诗良. 改进的时空背景差分目标检测[J]. *光电工程*, 2016, 43(2): 27-32, 39.)

[35] Yang G Z, Huang T S. Human Face Detection in a Complex Background[J]. *Pattern Recognition*, 1994, 27(1): 53-63.

[36] Graf H P, Chen T, Petajan E, et al. Locating Faces and Facial Parts[J]. *Proceedings of the International Workshop on Automatic Face & Gesture Recognition*, 1995, 41-46.

[37] Samaria F, Young S. HMM-Based Architecture for Face Identification[J]. *Image and Vision Computing*, 1994, 12(8): 537-543.

[38] Osuna E, Freund R, Girosit F. Training Support Vector Machines: An Application to Face Detection[C]. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2002: 130-136.

[39] Papageorgiou C P, Oren M, Poggio T. A General Framework for Object Detection[C]. *Sixth International Conference on Computer Vision* (*IEEE Cat. No.98CH36271*), 2002: 555-562.

[40] Freund Y, Schapire R, Abe N. A short introduction to boosting[J]. *Journal-Japanese Society For Artificial Intelligence*, 1999, 14 (771-780): 1612.

[41] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 580-587.

[42] He K M, Zhang X Y, Ren S Q, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.

[43] Girshick R. Fast R-CNN[C]. *2015 IEEE International Conference on Computer Vision*, 2016: 1440-1448.

[44] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.

[45] Dai J F, Li Y, He K M, et al. R-FCN: Object Detection via Region-Based Fully Convolutional Networks[EB/OL]. 2016: arXiv: 1605.06409. https://arxiv.org/abs/1605.06409

[46] Redmon J, Divvala S, Girshick R, et al. You only Look Once: Unified, Real-Time Object Detection[C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779-788.

[47] Halevi S, Shoup V. Faster Homomorphic Linear Transformations in HElib[C]. *Annual International Cryptology Conference*, 2018: 93-120.

[48] Microsoft Research, Redmond, WA. Microsoft SEAL (release 4.0). *https://github.com/Microsoft/SEAL*, 2022.

[49] Jin X, Guo K, Song C G, et al. Private Video Foreground Extraction through Chaotic Mapping Based Encryption in the Cloud[C]. *International Conference on Multimedia Modeling*, 2016: 562-573.

[50] Jin X, Wu Y M, Li X D, et al. PPViBe: Privacy Preserving Background Extractor via Secret Sharing in Multiple Cloud Servers[C]. *2016 8th International Conference on Wireless Communications & Signal Processing*, 2016: 1-5.

[51] Barnich O, Van Droogenbroeck M. ViBE: A Powerful Random Technique to Estimate the Background in Video Sequences[C]. *2009 IEEE International Conference on Acoustics*, *Speech and Signal Processing*, 2009: 945-948.

[52] Chen Y D, Hao C Y, Liu A X, et al. Multilevel Model for Video

Object Segmentation Based on Supervision Optimization[J]. *IEEE Transactions on Multimedia*, 2019, 21(8): 1934-1945.

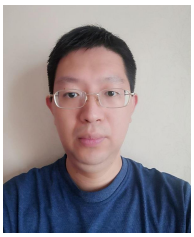
[53] Yuan P. *Privacy preserving and application of image convolution algorithm*[D]. Xi'an: Xidian University, 2017.
(袁鹏. 图像卷积算法的隐私保护和应用研究[D]. 西安: 西安电子科技大学, 2017.)

于浩洋 于 2020 年在北京电子科技学院计算机技术专业获得专业硕士学位。现在北京邮电大学网络空间安全专业攻读博士学位。研究领域为隐私计算、计算机视觉。Email: yuhaoyang@bupt.edu.cn

封化民 于 2004 年在新加坡国立大学获得博士学位。北京电子科技学院教授, 博士研究生导师。研究领域为网络安全、多媒体信息处理。Email: fenghm@besti.edu.cn

李晓东 河南三门峡人, 北京电子科技学院副教授。主要研究领域为隐私计算, 网络空间按安全。Email: lxd@besti.edu.cn

金鑫 博士, 副教授, 研究方向为计算美学、计算机视觉、人工智能安全, 北京电子科技学院可视计算与安全实验室(victory-lab)主任, 北京通用人工智能研究院(BIGAI)访问学者, 中国计算机学会高级会员。Email: jinxin@besti.edu.cn



刘飚 1980 年生, 博士。主要研究方向为信息安全和机器学习。Email: liubiao521@aliyun.com