前　言

社会的信息化给人们的工作和生活带来极大变革,各种业务平台、学习平台、娱乐平台、购物网站及信息资源服务成为人们生活中不可缺少的部分.但同时,个人信息被广泛使用,各行各业、各种机构积累的行业数据和社会数据越来越多,其中不乏大量的隐私信息,或可挖掘出大量隐私信息.如果不能保证这些数据的安全性、隐私性和应用的合理性,必将成为社会发展的巨大隐患.为了保障数据的安全与隐私,同时合理有效地对数据进行利用,发挥数据的价值,密码技术具有支撑性作用.面对大数据、物联网、区块链和人工智能等应用环境,密码理论与数据隐私保护方法需要新的机制与新的思想,需要有新的技术突破.

为推动我国学者在密码学与数据隐私保护领域的研究,促进各类数据的安全与协同利用,及时报道我国学者在该领域的最新研究成果,《计算机研究与发展》和我们共同策划组织了“密码学与数据隐私保护研究”专题,本期专题主要聚焦密码学的基础性研究、密码技术在数据隐私保护中的应用研究,以及大数据、物联网、区块链和人工智能等特定领域的数据隐私保护研究和应用.本专题通过公开征文共收到85篇投稿,并特邀了3篇综述稿件,分别在多个方面阐述了密码学与数据隐私保护研究领域具有重要意义的研究成果.本专题严格按照该刊审稿要求进行,特约编委先后邀请了近百位相关领域的专家参与评审,每篇论文邀请至少3～4位专家进行评审,历经初审、复审、终审等阶段,整个流程历经一个半月,最终本专题共精选录用文章25篇(含3篇特邀稿件).这25篇文章分别涵盖了密码算法与协议、隐私保护、信息隐藏与隐蔽传输及其他相关研究内容,在一定程度上反映了当前国内各单位在密码学与数据隐私保护研究领域的主要研究方向.由于刊物单期容量所限,本专题分别刊登在2020年第10期和第11期,信息隐藏与隐蔽传输及其他2组文章将在第11期刊登.

1　综　　述

在当今信息化社会中,大量数据被收集、传输和利用,为人们的生活、工作和学习带来极大便利.但信息安全与隐私泄露问题也日益严重,给社会带来了严重威胁.保护信息的安全与隐私需要前瞻性地考虑不同时期的计算能力及信息存储和处理的各个阶段和各个层面.本专题收录了4篇综述性文章(其中3篇为特邀稿件,1篇为普通投稿),涉及量子计算与量子密码、边缘计算、安全威胁情报共享以及机器学习的安全与隐私保护等多个方面.

量子计算与量子密码是基于量子效应的计算技术和密码技术.1984年第一个量子密钥分发协议BB84出现,开启了量子密码学的研究;1994年可在多项式时间内分解大整数的Shor算法出现,量子计算在原理上对传统的基于数学困难问题的密码学体制造成威胁.经过近半个世纪的研究与发展,量子计算与量子密码不但在理论上形成了自身的框架体系,在技术上也取得了突破性进展.“量子计算与量子密码的原理及研究进展综述”一文,从量子力学的数学框架、基本概念和原理、量子计算基本思想、量子密码研究进展及主要思想等方面对量子计算和量子密码的研究进行了梳理和总结.

目前,云计算作为大批量数据处理的主流模式被广泛应用,本地用户将大规模的数据和开销巨大的计算任务外包给云服务器完成.然而,单一的云服务器极易成为敌手的重点攻击目标;在多用户多任务场景中云服务器由于远离用户端易出现反馈延迟较大的问题.因此,边缘计算应运而生.

边缘计算在利于解决外包系统实时性问题的同时,也带来了安全与隐私保护的挑战.“边缘计算隐私保护研究进展”一文,从边缘计算特有的网络模型与安全模型、各种应用场景下的隐私保护安全计算及相关的各种密码学技术等方面,对边缘计算隐私保护领域的国内外最新研究成果进行了系统的阐述、总结与科学归类,探讨了边缘计算隐私保护当前面临的挑战、未来潜在的研究方向及其解决思路.

网络空间新生威胁以其复杂多变的攻击方式对网络安全造成严重危害,传统网络安全防御手段越来越不能适应这种新的网络环境.威胁情报的交换与共享可以使威胁情报价值最大化,降低情报搜集成本和改善信息孤岛问题,进而提高参与共享各方的威胁检测与应急响应能力.“网络安全威胁情报共享与交换研究综述”一文,介绍了网络安全威胁情报的概念和主流的威胁情报共享规范,分析了近10年来国内外有关威胁情报共享与交换的研究成果,归纳和总结了威胁情报共享与交换的现状与发展趋势.并对威胁情报共享与交换的研究方向和发展趋势进行了分析和展望.

机器学习在很多领域已经开始有了比较成熟的应用,如在广告推荐、自动驾驶、人脸识别等各个领域,机器学习都扮演着重要的角色.然而,因为机器学习的精准模型需要大量的训练数据作为支撑,蓬勃发展的机器学习技术使数据安全与隐私面临更加严峻的挑战.“机器学习的安全问题及隐私保护”一文总结了机器学习在训练和预测阶段常见的安全及隐私威胁,如投毒攻击、对抗攻击、隐私攻击等;介绍了常见的安全防御方法,同时重点总结了同态加密、安全多方计算、差分隐私等隐私保护的方法,并将典型的方案进行比较.在分析了现有研究的不足与挑战后,展望了机器学习隐私保护的未来发展趋势和研究方向.

2 密码算法与协议

在基于格的后量子密码领域,不使用复杂纠错码的基于模LWE/LWR问题设计的高效密钥封装方案主要有2类.“基于模格的密钥封装方案系统比较分析与优化”一文首次成体系地比较了直接基于LWE/LWR构造的密钥封装方案和基于密钥共识机制结合模LWE/LWR问题设计的密钥封装方案的异同,并在采用相同的压缩函数和相同的参数设置情形下,通过理论分析和实际测试得到了有关2类方案的优劣的结论.

可搜索加密是保护外包数据隐私的重要密码原语之一,动态可搜索加密允许用户对外包数据执行更新操作,因而得到更多重视.然而,目前大多数动态可搜索加密方案泄露信息较多、只支持单用户模式,在实际应用过程中存在安全和可用性问题.“一种增强的高效多用户前向安全动态对称可搜索加密方案”一文提出了一个支持多用户的前向安全动态可搜索加密方案,文章通过引入一个半可信代理服务器支持多用户操作,同时设计索引支持前向安全性质,实现了安全性和可用性的有机统一.

全同态加密可以对密文进行有效计算,是实现云计算、大数据以及机器学习中数据隐私安全的一项重要密码技术.然而,公钥规模和密文规模大以及密文运算计算复杂性高依然是全同态加密方案实用性的一个主要瓶颈.如果同态加密方案满足循环安全性,即可以对方案的私钥进行安全的加密,则可以使得运算密钥的规模独立于运算电路的深度,从而可以使得方案性能得到有效提升.因此,如何构造满足循环安全性的高效同态加密方案是值得研究的一个问题.拒绝采样技术常用于基于格的数字签名算法中以优化算法性能.“循环安全的同态加密方案研究”一文,首次将拒绝采样技术用于构造循环安全的同态加密方案,在增加部分采样算法的代价下,将系统参数从超多项式级降低到多项式级,大大约减了方案公钥和密文规模,从而有效改善了密文运算的计算复杂性.

随着云计算与5G通信的快速发展与广泛应用,云移动用户数迅速增长,云数据的隐私性保护

越来越受大众关注.早期提出的带关键字搜索的公钥加密方案和公共通道带关键字搜索的公钥加密方案允许系统中的任何用户向服务器发送加密文件供接收者检索,起到一定的隐私保护作用,但存在关键词隐私性保障不足、方案计算效率较低等问题.为此,“无配对公钥认证可搜索加密方案”一文提出一种非双线性对运算的公共通道的公钥认证可搜索加密方案,该方案的计算效率相对于双线性对方案高,能够满足抵抗离线模式下内部攻击者的关键词猜测攻击和在线模式下外部攻击者的关键词猜测攻击以及多关键词密文不可区分安全性,且具有认证功能,有很强的应用价值.

近年来无线通信技术的发展极大促进了移动设备的普及,用户可以使用移动设备随时随地访问到网络服务.由于网络空间的虚拟性,数字签名作为一种具有消息完整性认证,可鉴别性和不可否认性的技术应运而生.但是,移动设备自身存在易丢失或被劫持等安全隐患,导致对签名密钥(数字签名的信任根)的保护相对较弱.“移动互联网环境下轻量级 SM2 两方协同签名”一文针对 GM/T 0003—2012《SM2 椭圆曲线公钥密码术》标准中的 SM2 数字签名算法,设计了一种适用于客户端/服务器非平衡构架的 SM2 两方协同签名协议.该方案将签名密钥的控制权分配到客户端和服务器,以轻量级交互的方式产生 SM2 签名,从而提高了签名密钥在移动终端的安全性.

祖冲之密码算法(ZUC),是一个面向字设计的同步序列密码算法,已在 2011 年被正式批准成为 3GPP 的 LTE 国际标准密码算法,是一种应用较为广泛的轻量级密码算法.S 盒作为 ZUC 算法中的唯一非线性部件,其安全强度对整个算法的安全性起着至关重要的作用.ZUC 算法中所使用的 S 盒是固定不变的,而“一种基于混沌系统的 ZUC 动态 S 盒构造及应用方案”一文提出了复合混沌系统的思想,选取混沌特性良好的 Tent 和 Henon 两个混沌映射进行叠加,同时将图像置乱的思想引入到 S 盒的设计中来,利用 Arnold 映射对复合混沌系统所产生的 S 盒进行二次置乱,大幅度提高了 S 盒的非线性度.此外,通过对初始值和控制参数的控制和改变,可以动态地生成不同的 S 盒.NIST 随机性测试和软硬件性能测试的结果表明,本文所产生的 S 盒安全性更高,不仅增强了 ZUC 算法所产生密钥流序列的随机性,还兼具较好的软硬件可实现性.

随着量子计算的发展和量子计算机制造技术的进步,量子 Shor 算法逐渐对现有的公钥密码体制构成威胁,抗量子密码体制的设计迫在眉睫.2016 年美国国家标准技术研究所(NIST)抗量子密码算法标准征集所提交的算法中,基于格的密码体制被广泛认为最有希望成为标准算法.但是,由于缺乏后量子公钥基础设施(PKI)支持,基于格的密钥交换协议在当下还不能作为后量子密钥交换的完整解决方案.另外,基于格的密钥交换协议的研究主要是无认证密钥交换协议(KE)和密钥封装(KEM)算法.“后量子前向安全的可组合认证密钥交换方案”一文提出了一种签密方案和 DHKE-like 密钥交换的可组合 AKE 方案,使得 KE 方案满足认证性的同时,也具有前向安全性.作者考虑了经典密码算法到后量子密码算法过渡时期对 AKE 方案的可扩展性需求,将后量子无认证密钥交换与现有的公钥密码系统整合,利用现有公钥密码技术实现了实体的相互认证.这样,方案能够更广泛地适用现有的网络应用环境.

属性基签名可以有效保护物联网中用户的身份隐私,而且可以实现数据认证,但是属性基签名中签名生成和验证过程的计算开销过大,给物联网中资源受限的用户造成很大的计算负担.“工业物联网中服务器辅助且可验证的属性基签名方案”一文提出了一种服务器辅助且可验证的属性基签名方案,该方案将签名阶段和验证阶段的大部分计算开销委托给服务器,减小了签名者和验证者的计算开销.此外,方案对服务器产生的部分签名进行有效性验证,保证了服务器辅助签名产生阶段的安全性,而且抵抗了签名者和服务器的共谋攻击,保证了服务器辅助验证阶段的安全性.

3 隐私保护

随着人类社会进入大数据时代,新增数据量呈现爆炸式增长,有效利用这些数据能够创造很大

的实用价值.因此,如何在保护数据隐私的前提下进行有效分析,成为当前面临的一大问题.k-均值聚类作为一种常用的数据分析手段,在信息检索、机器学习等领域已经得到广泛应用.然而,常用k-均值聚类算法的安全实现主要基于比特分解,计算开销很大.针对该问题,"安全的常数轮多用户k-均值聚类计算协议"一文设计了常数轮交互的多用户k-均值聚类安全计算协议.该协议基于ABY混合协议框架,设计了针对同态密文的最小元素标记协议和除法协议,通过常数轮交互实现了同态密文、算术分享份额、Yao分享份额之间的相互转换,并利用Yao混乱电路技术实现了对同态密文的最小元素标记以及除法运算.文章无需使用昂贵的比特分解技术,同时在半诚实模型下给出了安全性证明.

声纹识别通过分析声音的物理特性进行身份认证,是一种重要的生物识别技术.如何对声纹模板进行安全保护,是声纹识别认证技术的一个关键问题.为了保护用户声纹模板的安全,"基于随机映射技术的声纹识别模板保护"一文提出了基于身份向量(i-Vector)和线性判别分析技术(LDA)的声纹模板保护方案,通过改进随机映射算法,对声纹特征进行随机化处理,允许用户在随机域注册并完成声纹识别.该方案的识别精度不受人数的影响,对身份向量进行随机化的正交矩阵处理时,声纹识别系统的识别精度基本不变,能够有效保证语音数据的安全.该方案可通过选择不同的随机正交矩阵,使用户数据具有多样性和随机性,提高声纹模板抵抗已知密文攻击的能力.

海量的物联网设备会带来频谱资源紧缺,因此亟需对授权用户的频谱资源实施共享来解决这一矛盾.然而,出于自私性和顾虑位置隐私泄露,一些授权用户不愿共享其空闲频谱,这将严重制约频谱共享在物联网中的有效实施."抗位置隐私泄露的物联网频谱共享激励机制"一文采用Geohash编码前缀和二进制编码后缀相结合的k匿名区域位置编码方式,设计编码优化的Casper模型(GB-Casper).该模型以授权用户所需的最小匿名区域面积$A_{\min}$控制Geohash编码长度,利用二进制编码进行k匿名区域的细粒度划分,通过字符串比较运算判断生成的k匿名区域中是否包含$k-1$个用户,以此减少二进制编码位数来逐渐扩大扫描区域,得到满足位置隐私保护的k匿名区域代替授权用户真实位置.引入频谱贡献度,连同位置隐私保护水平量化到博弈模型中,形成抗位置隐私泄露的物联网频谱共享激励机制.仿真结果表明,该文提出的方案可以快速构建k匿名区域,在防止位置隐私泄露的条件下,能有效激励授权用户积极参与频谱共享.

近年来,相关的数据安全和隐私保护法律法规对保护数据隐私提出了更高要求.然而.数据共享又是挖掘数据潜在价值、提高社会治理能力的重要途径.因此,支持隐私保护数据共享的保密集合求交协议受到越来越多重视.然而,大多数保密集合求交协议只支持计算交集,无法在不泄露交集的前提下高效计算交集的某些函数.针对该问题,"面向集合计算的隐私保护统计协议"一文设计了一组协议,能够高效地计算交集的统计量,如交集大小、交集权值和以及权值方差等.该构造基于茫然传输技术,可利用高效的茫然传输拓展技术优化计算效率.与此同时,文章借助Hash分桶技术对通信效率进行了优化.

随着区块链技术的发展,很多基于区块链技术的应用采用密码学技术保证参与者的隐私,但是隐私保护程度的加强会导致对区块链中的交易数据审计困难甚至无法审计."ACT:可审计的机密交易方案"一文中提出了可审计的机密交易方案(ACT),该方案利用签名对审计方进行身份验证,确保审计方的合法身份;在审计过程中加入同态加密,保证审计方只能知道一段时间内网络用户中所有交易总额,同时不泄露单个用户的交易金额.同时,为了同时保证交易数据的隐私性和提供数据的正确性,审计方案中采用了零知识证明技术.通过安全性证明,该方案满足可审计性,审计可靠性和交易金额隐私性.

近年来,联邦学习已经成为一种新兴的协作式机器学习方法.在联邦学习中,分布式用户可以

仅通过共享梯度来训练各种模型,但是共享梯度也可能导致用户隐私信息的泄露.“基于秘密分享和梯度选择的高效安全联邦学习”一文中将秘密共享与 Top-K 梯度选择相结合,设计了高效且安全的联邦学习协议,来保证用户隐私和数据安全,减少通信开销,并提高模型训练效率.此外,还提出了一种高效的方法来构造消息验证码,以验证服务器返回的聚合结果的有效性.

4 信息隐藏与隐蔽传输

隐写术是一种利用图像、视频、音频等多媒体载体实现隐蔽传输的技术,如何尽量减少对载体影响的同时嵌入尽可能多的信息是隐写算法的一个研究重点.为了解决载体图像中异常点带来的不利影响,“基于多尺度滤波器的空域图像隐写增强算法”一文提出了一种适用于空域隐写算法的隐写增强算法,通过不同大小的滤波器来提取出不同大小的图像纹理细节,并通过调整这些滤波器的权重,使得在增强纹理复杂区域的同时尽可能减少对平坦区域噪声的增强.实验结果表明,与传统自适应空域隐写算法相比,该方案在多种嵌入率下都能获得抗隐写分析检测的安全性提升,适用于现有的空域隐写算法,并且能够提升它们的抗隐写分析检测能力.

在计算开销和实时性敏感的物联网环境中,利用移动边缘计算可以实现基于图像隐写的隐蔽通信,不可察觉性、安全性和容量是衡量图像隐写算法优劣的重要指标.为了解决这些指标的多目标优化问题,“基于边缘计算的进化多目标优化图像隐写算法”一文提出了一种基于人工免疫原理的多目标优化图像隐写算法 MO-GA,通过建立带约束的多目标优化模型,并基于人工免疫原理将秘密嵌入到图像中难以统计建模的纹理、噪声区域,使得目标性能得到优化.该算法通过高通滤波器组对图像进行预处理,在多个方向上寻找图像中难以建模的噪声、纹理区域用于嵌入秘密.通过实验分析和对比表明,该算法能够很好地保持图像质量,具有较好的抵抗隐写分析的能力.

兼顾含密 JPEG 图像的文件增量和视觉失真,“基于失真-扩展代价的 JPEG 图像可逆数据隐藏”一文提出一种基于失真-扩展代价的 JPEG 图像可逆数据隐藏算法,采用直方图平移实现秘密数据的可逆嵌入,研究解决如何根据嵌入容量自适应选择嵌入频率和图像块,以最小化含密 JPEG 图像的视觉失真和文件增量.作者分析了通过模拟计算数据嵌入不同频率的单位文件增量确定频率嵌入顺序、根据图像块零 AC 系数个数和平滑度确定图像块嵌入顺序的合理性,数据嵌入时优先选取较小单位文件增量的频率和较平滑图像块;并分别定义了单位文件增量、单位失真-增长比作为算法文件扩展、视觉质量与文件扩展关系的定量评价指标.与现有同类算法相比,作者所提出的算法可以在有效降低含密 JPEG 图像文件增量的同时保持较高的视觉质量.

端信息跳扩混合技术是一种在端到端的网络数据传输中伪随机改变端信息,并利用端信息扩展序列实现高速同步认证的主动防御技术.“基于端信息跳扩混合的文件隐蔽传输策略”一文将端信息跳扩混合技术引入文件隐蔽传输,研究了端信息跳扩混合网络环境下的文件隐蔽传输策略,提出组播时间校正方案,解决了通信过程中的同步问题;提出基于时间传输和基于传输大小传输的 2 种适用于端信息跳扩混合网络环境文件传输方案,并在文件传输过程中增加数据迁移技术,实现文件的隐蔽传输和完整性传输;设计实现端信息跳扩混合文件隐蔽传输原型系统并进行了有效性、安全性测试,实验结果表明该文件隐蔽传输策略能够有效满足文件传输完整性和隐蔽性要求.

5 其 他

浏览器指纹是 2010 年提出的一种技术,指服务器使用访问浏览器的特征标识、canvas 特征值以及部分硬件信息和系统信息,并通过特定的指纹生成算法为该用户使用的浏览器生成唯一的字符串标识.传统的浏览器指纹技术在追踪用户方面问题颇多,无论系统升级、浏览器更新还是指纹

篡改程序伪造导致的指纹特征值改变,都会使浏览器指纹发生变化."基于双向循环神经网络的安卓浏览器指纹识别方法"一文基于双向 RNN 的指纹识别模型,提出了一种针对安卓设备浏览器指纹识别的监督学习框架 RNNBF,在数据方面构建基于指纹的数据增强技术生成增强数据集,在模型方面采用注意力机制令 RNNBF 专注于具有不变性的指纹特征.实验结果表明,RNNBF 模型在动态链接指纹方面的效果优于单层 LSTM 模型以及随机森林模型.

传统的工业控制系统正在向网络化转变,越来越多的以太网协议被应用到工业控制系统中,但由于工业控制系统最初设计为封闭的系统,缺乏加密认证等安全手段,因此面临越来越多的安全隐患."工业以太网 EtherCAT 协议形式化安全评估及改进"一文首先提出一种基于 Petri 网理论和 Dolev-Yao 攻击方法的模型检测方法,然后利用该方法对 EtheCAT 协议进行了安全分析,发现了篡改、重放和欺骗 3 类中间人攻击的漏洞.最后,通过在原协议中加入密钥分发中心来辅助身份认证,利用 Hash 实现消息完整性检测,从而弥补协议的漏洞.

图像和文本相结合的多模态网络谣言更具迷惑性和煽动性,因此对国家安全和社会稳定的危害性更为严重.当前网络谣言检测工作虽然充分考虑了谣言中配文的文本内容,但却忽略了图像内容以及图像中的内嵌文本."MSRD:多模态网络谣言检测方法"一文提出了一种基于深度神经网络的针对图像、图像内嵌文本以及配文文本内容的多模态的网络谣言检测方法 MSRD.该方法使用 VGG-19 网络提取图像内容特征,使用 DenseNet 提取图像内嵌文本内容,使用 LSTM 网络提取文本内容特征,与图像特征串接后,通过完全连接层获取图像与文本共享表示的均值与方差向量,借助从高斯分布中采样的随机变量以形成重新参数化的多模态特征并作为谣言检测器的输入进行谣言检测.实验表明该方法在 Twitter 和微博两大数据集上达到了 68.5%和 79.4%的准确率.

承蒙各位投稿人、审稿专家和编辑部等方面的全力支持,本专题得以顺利出版,密码学与数据隐私保护研究领域,特别是密码技术在隐私保护中的应用,发展十分迅速,涉及范围十分广泛,这给审稿人及特邀编辑的审稿、选稿工作带来了巨大挑战.由于投稿数量大、主题广泛、专题容量有限等原因,本专题只能选择部分有代表性的研究工作予以发表,部分优秀稿件无法列入其中,深表遗憾,敬请谅解.

我们要特别感谢《计算机研究与发展》编委会和编辑部,从专题的立项到征稿启事的发布,从审稿专家的邀请到评审意见的汇总,以及最后的定稿、编辑和出版工作,都凝聚了他们辛勤的汗水.本专题的出版期望能给广大相关领域研究人员带来启发和帮助.特别要说明的是,由于我们水平所限,在审稿、选稿过程中难免出现不尽人意之处,敬请各位作者和读者包容谅解,同时也请各位同行不吝批评指正.最后,再次衷心感谢各位作者、审稿专家、编辑部和特邀编委的辛勤工作.

曹珍富(华东师范大学)
徐秋亮(山东大学)
张玉清(中国科学院大学)
董晓蕾(华东师范大学)

2020 年 8 月