文章编号:1671-4229(2021)04-0046-10

简洁非交互零知识证明综述

朱旭东[1,2],张心轩[1,2],邓 燚[1,2*]
(1.中国科学院 信息工程研究所/信息安全国家重点实验室,北京 100093;
2.中国科学院大学 网络空间安全学院,北京 100049)

摘 要:零知识证明允许证明者向一个验证者证明一个断言的正确性而不泄漏其它任何知识。在理论层面,这一概念的提出对密码学和计算机科学都产生了深远影响。在应用层面,零知识证明以及它所衍生出来的可验证计算为近年来迅猛发展的以区块链为代表的金融科技提供数据隐私保护和去中心化的技术。近年来高效简洁非交互零知识证明(zero-knowledge Succinct Non-interactive Arguments of Knowledge, zk-SNARKs)方面的研究取得了令人瞩目的进展,其更是凭借简洁高效的特点被广泛应用于电子金融领域之中。文章主要梳理了zk-SNARKs的发展历程,并依据它们依赖的不同模型进行分类与总结。

关键词:零知识证明;SNARK;密码协议
中图分类号:TP 309.7 **文献标志码**:A

A survey on succinct non-interactive zero-knowledge proofs

ZHU Xu-dong[1,2], *ZHANG Xin-xuan*[1,2], *DENG Yi*[1,2*]
(1. State Key Laboratory of Information Security/Institute of Information Engineering, Chinese Academy of Sciances, Beijing 100093, China;
2. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Zero knowledge proof allows the prover to prove the correctness of a statement to a verifier without revealing any other knowledge. At the theoretical level, this concept has a profound impact on cryptography and computer science. While at the application level, zero knowledge proof and its derived verifiable computation provide data privacy protection and decentralization technology for the rapid development of financial technology represented by blockchain in recent years. In the last decade, the research on zero-knowledge succinct non-interactive arguments of knowledge (zk-SNARKs) has made remarkable progress, and it is widely used in the field of e-finance by virtue of its simplicity and efficiency. This paper surveys the development process of zk-SNARKs, and summarizes them according to the different models they rely on.

Key words: zero-knowledge proofs; SNARK; cryptographic protocol

1 研究背景

Goldwasser等[1]给出了零知识证明(Zero-Knowledge Proofs, ZKP)的概念,零知识证明的出现在密码学上有着重要的意义,它给人们的现实生活也带来了巨大的冲击。零知识证明是一个强大的密码学工具,能让证明者向某个验证者证明某个断言是正确的而不透露其它任何信息,它需要满足3个性质:①完整性:如果证明者做出正确的证明,那么验证者一定会通过验证;②可靠性:

基金项目:国家自然科学基金资助项目(61932019, 61772521, 61772522);中科院前沿学科重点研究资助项目(QYZDB-SSW-SYS035)

作者简介:朱旭东(1997—),男,博士研究生. E-mail:zhuxudong@iie.ac.cn
*通信作者. E-mail:deng@iie.ac.cn

引文格式:朱旭东,张心轩,邓燚. 简洁非交互零知识证明综述[J]. 广州大学学报(自然科学版),2021,20(4):46-55.

如果证明者伪造一个虚假的证明,那么验证者除了可忽略的概率外会拒绝;③零知识性:验证者只能知道断言是正确的而不知道任何其他知识。

随着对零知识协议研究的深入,以及对其应用要求的增加,大家并不满足于需要双方进行交互的零知识证明系统。Blum 等[2]在公共参考串模型下(Common Reference String model,CRS model)将零知识的概念延伸至非交互零知识证明的概念(Non-interactive Zero-knowledge proofs, NIZK proofs),在这个系统里,证明者可以只发一条消息给验证者,然后验证者验证即可。NIZK 在包括数字签名、选择密文攻击(Chosen-ciphertext Attack, CCA)和安全的公钥加密等非交互密码学方案中都十分有用。

在大数据时代的今天,互联网上的数据量日益增大,零知识证明作为一个被广泛应用的工具,大家对其通信复杂度和效率的要求也越发严格。于是简洁非交互零知识的知识论证(Zero-knowledge Succinct Non-interactive Arguments of Knowledge, zk-SNARKs)概念也被提出并被广泛研究,事实上 zk-SNARKs 是如今区块链等密码学应用领域上重要的一环。

早期的 zk-SNARKs 都是在 CRS 模型下讨论的,虽然目前已经可以在这种模型下构造十分高效的 zk-SNARKs 方案,但该模型需要依赖受信任的初始设置(Setup)。从现实的角度来看,实现可信的初始设置是很困难的,因为实现过程中往往会产生有害信息(toxic waste),万一被泄漏,敌手则可以利用这些有害信息生成能够通过验证的伪造证明。此外,这些 zk-SNARK 的初始设置是特定电路的一次性受信任的初始设置,即由该初始设置产生的 CRS 只能针对特定电路,而不能针对任意的电路。为了避开这些问题出现了以下一系列的解决方法和新的研究路线。

1.1　通用可更新(Universal updatable)的 zk-SNARKs

在这个模型下 CRS 可以被任意更新参与方非交互的更新,这个更新是可验证的,并且只要在所有参与方中至少有一个是诚实做的,那么模拟的陷门就不会被泄露,从而达到了取代可信第三方的目的。另外,这种可更新的 CRS 一般都是通用的(Universal),也就是说它们的生成不依赖于具体电路,而是直接被用于针对任意电路的零知识证明之中。

1.2　透明的(Transparent)zk-SNARKs

这个模型主要研究 CRS 生成时无需可信第三方的 SNARK 协议。绝大多数透明的 zk-SNARKs 协议,其 CRS 是无结构的均匀随机串,有时候,也称这种使用无结构的均匀随机串的 CRS 模型为均匀随机串模型(Uniform Random String model, URS)。透明的 zk-SNARKs 的安全性通常是在随机预言机模型(Random Oracle Model)下证明的。

1.3　纯公钥模型(Bare Public Key model,BPK)下的 zk-SNARKs

在该模型下的 zk-SNARKs 协议分为 2 个阶段,第一个阶段不同于 CRS 模型下由可信第三方来生成 CRS,该模型下是由验证者生成 CRS;第二个阶段再由证明者与验证者在这个 CRS 下执行零知识证明协议(当然为了保证零知识性质,目前的方案都需要证明者先对这个 CRS 进行一些验证),因此,该模型不需要可信第三方,要比 CRS 模型更弱。

1.4　模拟多方安全计算(MPC-in-the-head)的 zk-SNARKs

在该模型下,证明者在一些(至少 3 个)虚拟的服务器之间模拟一个 MPC 协议,然后将他们所看见的副本和每个服务器的中间状态承诺起来。此后,验证者可以让证明者打开这些承诺的一个子集来高效验证是否一致。由 MPC 的私密性可以保证证明者打开给验证者看的内容不会泄露额外信息,由 MPC 的正确性也可以保证证明者无法欺骗验证者。

2　公共参考串模型下的 zk-SNARKs

零知识证明[1],特别是非交互零知识证明[2]在密码学理论和实际密码学应用中都扮演着重要的角色。而在诸如文献[3-7]等大量的研究下,基于 pairing 的简洁非交互零知识的论证(Zero-knowledge Succinct Non-interactive Arguments of Knowledge, zk-SNARKs)出现在人们的视野中,并被不断发展。相比于非交互的零知识证明,zk-SNARKs 最大的特点就是简洁性,事实上它甚至可以让证明者在证明长度仅为常数个群元素的情况下证明任意大小电路的可满足性。此外,为了满足不断提高的应用需求,如今许多的 zk-SNARKs 方案在计算效率方面也被设计得十分高效。

关于为什么 CRS 模型会被提出的背景,Goldreich 等[8]提出在标准模型下,对于足够困难的语言想要构造一轮的协议来抵抗带辅助输入的证明者的攻击是不可能的,因此,需要考虑一个更强的模型——CRS 模型。

另外,知识的论证(Argument of Knowledge/ Knowledge Soundness)是对合理性(Soundness)的强化定义,在

该定义下,一个高效证明者争对某个断言生成的证明通过验证,不仅能说明该断言对应的证据的存在性,并且这个证据还能被高效地抽取出来。许多证明系统都满足这个性质,这个性质对于简洁论证系统的应用也有重要的影响。

概率可检验证明(Probabilistically Checkable Proofs, PCP)定理[9-12]为NP语言提供了一个新的、对证明系统产生深远影响的特征,它说明了NP断言拥有能在正常证明尺寸和对数多项式时间内被验证的概率可检验证明。Kilian[3]将这种新的NP特征应用于密码场景,证明可以对NP语言用PCP去构造交互论证。而通信量大小是零知识证明的一个重要表现参数,Klilian给出了第一个亚线性通信规模和验证时间的零知识证明,该方案需要发送的比特数少于需要被证明的断言,这样的论证系统被称为简洁的(Succinct)论证系统。

需要注意的是Klilian的协议需要4轮交互,而对于应用而言非交互显然更加有意义。因此,朝着这个方向尝试,Micali[4]通过让证明者在这些论证中使用密码学函数计算验证者的挑战,而在RO模型下构造出了一个公开可验证的简洁非交互论证方案。正如Kilian[13]所述,这种Fiat-Shamir启发式[14]可以把公开掷币(Public Coin)的零知识论证变成亚线性规模的NIZK方案。此外,还有一些早期工作,如文献[15-18]研究了怎么利用可抽取的抗碰撞哈希(Extractable Collision-resistant Hashing)[19-21]来去除Kilian的基于PCP的协议的交互,或者利用有一个可抽取同态性质的全同态加密(Fully-homomorphic Encryption with An Extractable Homomorphism Property)[22]去构造基于MIP(Multi-prover Interactive Proofs)的SNARKs。

经典的构造简洁论证的方法是首先给出一个信息论构造,然后在对证明者有某种约束的模型下利用一些概率检查(Probabilistic Checking)去验证计算(例如前面提到过的PCP模型[3-4,18-21,23],或者其它的概率检查的模型[22,24-29]),最后再利用密码学工具将信息论构造编译成论证系统(此时对于证明者,除了要求其算法为高效的,并无其它约束)。然而如今的构造都偏离了这个方法而基于新的方法,新的方法有很多吸引人的好处,例如,公开可验证与证明只包含常数个群元素等。

Groth等[30-33]提出了基于pairing的NIZK证明,这也导致了标准假设下的第一个线性通信规模的证明诞生。Groth[5]结合这些技术和他此前提出的交互零知识论证的想法[34],在指数知识假设(Knowledge of Exponent Assumptions, KEA)[35-37]和双线性群下给出了第一个证明只包含常数个群元素的NIZK论证。Groth的常数规模的NIZK论证基于一系列多项式等式,并采用pairing的方式去高效地验证这些等式。Lipmaa[38]在同样的群和假设下使用了一个基于进展集合(progression sets)的新构造进一步压缩CRS的长度。文献[5, 38]都用到了不可证伪的假设,事实上Gentry等[39]给出了一个负面的结论——不可证伪的假设(包括知识假设和一些强的模型)是证明SNARK的知识的论证所必须的。

不同于基于之前所提到的概率检查的方法,Gennaro等[6]首先通过电路的编码和拉格朗日插值将电路可满足性问题规约到了更容易处理的代数可满足问题上,从而跳过了经典方法中的概率检查的步骤而直接构造具体的密码学工具来验证代数可满足问题,这种方式是对信息论材料和密码学材料的进一步区分。他们给出了2种新的代数可满足问题,一个是与证明布尔电路可满足性相对应的QSP(Quadratic Span Programs),另一个是与证明算术电路可满足性相对应的QAP(Quadratic Arithmetic Programs),并给出了一个CRS长度与断言和证据大小呈比例的基于pairing的zk-SNARK,这样构造出的SNARK拥有更加高效的验证时间和更短的证明长度。Lipmma[40]结合纠错码提出了更加高效的QSP。Danezis等[41]将quadratic span programs重新定义为了square span programs,并给出了一个针对布尔电路的包含4个群元素的zk-SNARK。Bitansky等[42]给出了基于对域元素做线性编码的SNARKs的抽象模型——LIPs(Linear-interactive Proofs),这个模型抓住了证明者被限制为只能使用线性操作计算消息的特征,此外,该文章给出了一个一般的转化——从2轮LIP结合基于pairing的技术转化为公开可验证的SNARK或者结合同态加密技术得到指定验证者的SNARK。在LIP的框架下,Groth[7]提出了迄今为止在验证时间和证明长度(证明只包含3个群元素)上表现最优的zk-SNARKs,也因此已被广泛用于电子货币等互联网领域中。

这些传统的zk-SNARKs方案虽然已经在通信量和效率上有了不错的结果,但是其还有2个比较大的缺陷:①其CRS生成都需要依赖可信第三方,因而在中心化系统中难以实现;②其大部分只能针对特定线路,即对每个不同线路的证明都需要重新生成一次CRS,这使得其在应用上受到限制。

当然,也存在CRS模型下的针对通用电路的zk-SNARKs。例如Libra[43]也需要CRS,但是其CRS是通用的,与具体电路逻辑无关而只与电路的输入长度有关。该方案也是第一个拥有最优证明者时间(证明时间

为 $O(C)$ 且与电路类型无关)和简洁证明长度/验证时间的零知识证明系统。特别地,如果令 C 是电路大小,d 是电路深度,那么 Libra 的证明时间为 $O(C)$,证明长度和验证时间都是 $O(dlogC)$。该方案的核心技术有 2 个,一个是线性证明者时间的 GKR 协议(Goldwasser, Kalai 和 Rothblum 的交互证明协议);另一个是高效的将 GKR 协议转化为零知识协议的方法。

3　通用可更新的 zk-SNARKs

为了取代 CRS 模型,Groth 等[44]定义了可更新模型(Updatable Model),这个模型是完全的 CRS 模型和完全的 BPK 模型的折衷模型,在这个模型下,CRS 可以被任意更新参与方非交互地更新,这个更新是可验证的,并且只要在所有参与方中至少有一个是诚实的,那么模拟的陷门(Trapdoor)就不会被泄露,从而达到了取代可信第三方的目的。另外,一般这种可更新的 CRS 都是通用的(universal),也就是说它们的生成不依赖于具体电路,因此,可以被用于对任意电路的零知识证明。

算术电路的可满足性问题可以被归约到一系列有限域上的二次和仿射约束问题。二次约束是通用的,并且很容易在 pairing 场景下通过 Hadamard 乘积论证(Hadamard product argument)来对其进行证明。关于二次约束的证明,大部分 zk-SNARKs 构造的基础核心来源于文献[6]。而仿射约束是依赖电路的,因此,如何在通用 CRS 下高效地证明仿射关系是一个充满挑战的问题。

Groth 等[44]给出的构造尽管只需要常数个群元素的证明长度和常数的验证时间,但需要用到花销很大的子空间论证(Subspace Argument),它要求 CRS 和预处理的长度分别为电路大小的平方和立方。另外,更新 CRS 要求平方多个群指数运算,而验证更新的正确性则需要线性多个 pairing 计算。而 Sonic[45]是第一个高效的、通用的、可更新的 SNARK,它的证明长度为常数个群元素而且 CRS 规模仅与电路大小呈线性关系。Sonic 的技术目标是去对一个代表 NP-hard 语言的约束系统的可满足性做零知识论证,而这个约束系统与 Bulletproofs[46]中的约束系统类似,也表达为二元多项式等式。为了给多项式做承诺,该文章还用到了 Kate 等[47]提出的多项式承诺方案的一个变体,并证明这个承诺方案在代数群模型(Algebraic Group Model, AGM)下是安全的。对于正确赋值的证明,Sonic 也给出了一个新的置换论证和一个 Grand-product 论证。Sonic 的验证需要检查一个稀疏二元多项式的赋值,该文章提出了一个简洁地检查这个赋值的方法。总的来说,Sonic 给出了 2 种不同的方式去证明仿射约束,一个方式是完全简洁的但并不是十分高效,另一个方式通过引入一个不受信任的第三方,在批验证场景下(Batch Verification Context)可以做得十分高效。由于 Sonic 的整个协议是公开掷币的,而其用到的子部件是 Kate 的承诺方案的一个变形,该承诺方案是可更新的且其安全性可以在代数群模型下被证明,因此,在 RO 模型、AGM 下,Sonic 是通用可更新的 zk-SNARK。

Marlin[48]是 Sonic 的改进版,其拥有与电路规模呈线性的 CRS 长度和常数个群元素的证明长度且都比 Sonic 要短,Marlin 的证明时间比 Sonic 快 10 倍,而验证者时间比 Sonic 快 3 倍。Marlin 可以在标准假设下实现,也可以在代数群模型(Sonic 也用到了这个模型)下实现以追求最大的效率。Marlin 的技术核心是一个高效的争对 R1CS(Rank-1 Constraint Satisfiability)的代数全息证明(Algebraic Holographic Proof, AHP),它能在线性的证明长度和常数的问询复杂度(Query Complexity)下实现。事实上,构造一个“全息证明(holographic proofs)”[49]与在线下阶段对一个线路做预处理的能力息息相关。这意味着验证者可以不以电路的描述作为输入而是对其的编码做少量的问询。该文章中对电路描述的编码包括一些低次多项式,与此同时,证明本身也是一些低次数的多项式,这可以视为要求证明者是代数的,而这就构成了 AHP。该文章扩展了 Kate 的多项式承诺方案[47],设计了一个高效的 AHP,并指出利用公开掷币的 AHP 和多项式承诺就可以得到简洁的交互论证系统,最后利用 Fiat-Shamir 转化,可以得到通用可更新的 zk-SNARKs。

Plonk[50], 也是 Sonic 的改进,其相比于 Sonic 而言,Plonk 提高了证明时间(群指数运算少 7.5 ~ 20 倍)。与 Sonic[45]类似,Plonk 也依赖于 Bayer 和 Groth 的置换论证[51]。然而 Plonk 关注的是子群上的赋值而非单项式的系数,因此,同时简化了这个置换论证和算术步骤。Plonk 的构造思路是首先将电路的约束分为 2 种——门约束和置换约束,用一元多项式表达门约束,然后和之前的方案类似同样用了 Kate 多项式承诺[47]的变形来打开多项式在一个点的值进行证明,而用置换论证来证明置换约束(也要用到多项式承诺)。由于整个过程是公开掷币的,而 Kate 的承诺方案的变形是可更新的且与 Sonic 类似可以在代数群模型下证明安全性,使得在 RO

模型、AGM 下,Plonk 是通用可更新的。

Lunar[52]与前面这些方案相比则有着更好的证明长度和证明者验证时间,而代价是其在验证密钥的长度和验证时间上有更高的限制。其技术贡献主要有如下 2 点:①提出了一种新的代数型 IOPs(Interactive Oracle Proofs)变体——多项式全息 IOPs(Polynomial Holographic IOPs, PHPs);②提出了一个新的编译器,它利用这个 PHPs 结合针对被承诺多项式的 CP-SNARKs (Commit-and-Prove zkSNARKs)来构造通用的 zk-SNARKs。其它的贡献包括对这些 CP-SNARKs 在 pairing 场景下实现、对 R1CS 和其新变体的 PHPs 构造的实现。另外,前面所说的编译器稍做改变可以变成第一个从 IOPs 到 Commit-and-Prove 的通用 zk-SNARKs 的编译器。

4 透明的 zk-SNARKs

随着诸如区块链等去中心化的系统的发展与应用,类似文献[7]的传统 CRS 模型下的 SNARK 协议的需要可信第三方的特性成为了其在去中心化系统应用中的障碍。为了避免可信第三方的需求,一方面,正如前面所描述的,各种可更新 CRS 的 zk-SNARK 系统作为一种去中心化的解决方法被提出并得到广泛应用;另一方面,正如本节中将要介绍的,各种透明的 zk-SNARK 方案被构造提出并发展以实现去中心化。

与通用可更新的 zk-SNARKs 的 CRS 可更新并且只要更新用户中有一个用户是诚实可信的 CRS 就是安全的不同,透明的 zk-SNARKs 的 CRS 的生成不需要任何可信第三方。尽管透明的零知识协议的概念很早就被提出了,但是可实用的、透明的 zk-SNARKs 直到最近几年才出现。2018 年 Ben-Sasson 等[53]提出了 STARK 方案。在 STARK 中,首先需要将待证明的线路或者约束条件转化成代数中间表示(algebraic intermediate representation, AIR)。简单来说,AIR 将一个线路的执行表示成线路的中间状态(包括输入状态与输出状态)的转换,线路则可以表示成这些相邻状态的转换/约束。之后使用代数的布局与布线(Algebraic Placement and Routing,APR)技术优化线路的约束,并将线路的可满足性转化成 2 组 Reed-Solomon 有效性检测问题(Reed-Solomon Proximity Testing (RPT) Problem)。之后使用代数链接(Algebraic Linking)将上述 2 组 RPT 问题聚合成 2 个 RPT 问题,最后使用快速 Reed-Solomon 有效性 IOP 协议(FRI)实现证明。大致来说,对于结构较好(拥有较多重复子线路)的线路而言,STARK 可以实现 $O(n\log^2 n)$的证明者时间与 $O(\log^2 n)$的验证者时间与通信量,其中 n 是线路的大小。

同一年,Wahby 等[54]提出了 Hyrax 方案,该方案基于优化过的 Giraffe 交互证明协议,结合 Cramer-Damgård 技术将其转化成零知识论证协议。为了优化该协议的效率,Hyrax 给出了 2 个关键的改进技术,首先是通过将验证算法、多重承诺方案与 Schnorr 式的证明相结合优化 IP 的结构,其次是提出一种新的证据承诺方案减少验证者计算时间。改进后的协议拥有 $n+d\cdot g\log g$ 的证明者时间,$d\cdot\log g+\sqrt{w}$的验证者时间与 $d\cdot\log g+\sqrt{w}$的证明大小,其中,n 是线路的大小,d 是线路的深度,g 是线路的宽度,w 是证据的长度。与 STARK 一样,Hyrax 的实际效率也很大程度上依赖于线路的结构。

不同于上述 2 种方案其表现很大程度上依赖于线路结构,Ben-Sasson 等[55]提出的 Aurora 是一种针对 R1CS 实例的透明的 zk-SNARK,其表现不容易受到 R1CS 实例结构的影响。具体而言,对于一个 R1CS 实例 $((A,B,C,v),w)$ s. t. $(Az)\circ(Bz)=Cz$ for $z:=(1,v,w)$,其中,$\circ$代表逐项相乘。Aurora 将其转化成 2 种问题:给定 $x,y,z\in\mathbb{F}^m$检查是否有 $x\circ y=z$; 给定 $x\in\mathbb{F}^m,y\in\mathbb{F}^n,M\in\mathbb{F}^{m\times n}$是否有 $x=My$。类似于 STARK,Aurora 将上述问题中的向量编码成多项式/RS 码字的形式,此时第一个问题$f_x(a)\cdot f_y(a)-f_z(a)=0$ for $a\in H\subset\mathbb{F}$可以使用 FRI 协议高效证明。对于第二个问题:For $H_1,H_2\subset\mathbb{F}$, $a\in H_1$, $f_x(a)=\sum_{b\in H_2}M_{a,b}\cdot f_y(b)$, Aurora 提出了一种新的单变量 Sumcheck 协议来实现上述等式的高效检查。Aurora 实现了 $O(n\log n)$的证明者时间与 $O(\log^2 n)$的通信量,$O(n)$的验证者时间,并且尽管通信量与 STARK 在渐进意义下是相同的,但是在实际操作中要比 STARK 短得多。

2019 年由 Chiesa 等[56]提出的 Fractal 也是一种针对 R1CS 断言的 SNARK。该方案通过对 R1CS 断言构造一种新的全息证明,并给出一种从全息 IOP 到透明的带预处理的 SNARK 的转化方式,从而实现了对 R1CS 类的透明的带预处理的 SNARK 构造。这个构造实现了 $O(n\log n)$时间的预处理时间,$O(n\log n)$的证明者时间,$O(|x|+\log^2 n)$的验证者时间与 $O(\log^2 m)$的证明长度。

同年,Setty[57]提出了另一种针对 R1CS 断言的 SNARK——Spartan。该方案将一个 R1CS 实例 $x=(F,A,B,C,io,m,n)$编码成 3 次的 $\log m$ 变量多项式 G 使得该实例是可满足的,当且仅当 $\sum_{x\in\{0,1\}^{\log m}}\mathcal{G}(x)=0$。之后将上述等式转换成线性多变量多项式的 Sum-Check,

从而实现 NIZK 的构造。在预处理的情形下,这个 NIZK 是一个 SNARK,并根据使用的多项式承诺不同,其效率表现也不同。其可以在证明者时间 $O(n)$ 或者 $O(n\log n)$ 的情形下,分别实现 $O(\log^2 n)$ 或者 $O(\sqrt{n})$ 的验证者时间和证明长度。另外,文章中也提出了用 SPARK 等技术来提升方案的表现,使方案在实际运行中取得良好的表现。

在 2016 年,Bootle 等[58]为 Pedesen 承诺设计了一种新的内积求值的递推论证技术,并使用这个技术为代数线路的可满足性设计了一种对数的通信复杂度的交互式零知识论证协议。在其基础上,Bünz 等[46]于 2018 年提出了一种新的透明的短证明 SNARK 协议,名为 Bulletproof。Bulletproof 优化了先前的内积求值的递推论证技术,实现了证明者验证者时间为 $O(n)$,证明大小为 $O(\log n)$ 的透明的 SNARK 协议。

某种程度上,Bulletproof 蕴含着一个无需可信第三方的具备内积求值功能的多项式承诺方案,事实上许多 CRS 可更新的 SNARK 协议,如 sonic 和 plonk 等,都可以视作是通过多项式承诺与多项式 IOP 的结合。其之所以是可更新的而不是透明的,原因在于其使用的多项式承诺不是透明的。最近 Bünz 等[59]在未知阶群上使用 Diophantine Argument of Knowledge (DARK)技术构造了新的多项式承诺。未知阶群可以使用 RSA 群或者类群,当使用类群时,多项式承诺不需要任何可信方。DARK 多项式承诺①拥有常数的参数与承诺大小,当承诺 μ 个变量的 d 次多项式时,其多项式上点的打开需要 $O(d^{\mu}\mu\log d)$ EXP 的证明时间,$O(\mu\log d)$ EXP 的验证时间与 $2\mu\log d$ $\mathbb{G}_U$ 的证明长度,其中 EXP 指的是类群上的一次 λ-bit 指数操作(所需的时间),$\mathbb{G}_U$ 指的是未知阶群上的元素(长度)。Bünz 等利用 DARK 承诺方案改进了 Sonic 方案,得到了 Supersonic 方案,后者是透明的 SNARK,证明者时间为 $O(n\log n)$EXP,验证者时间为 $3\log(n)$EXP,证明长度为 $2\log n$ $\mathbb{G}_U$。当证明 $n=2^{20}$ 规模的线路时,Supersonic 仅需要 10.1 KB 的证明大小。

5　纯公钥模型下的 zk-SNARKs

不同于通用可更新和透明的 zk-SNARKs 构造思想,还有一些值得考虑的可实现去中心化的比 CRS 模型更弱的模型,例如在由文献[61]中提出的参数篡改(Parameter Subversion)模型下,考虑生成公共参数的一方有可能被敌手控制进而生成有利于敌手的参数,并最终导致协议丧失可靠性或零知识性。构造该模式下安全的 zk-SNARKs 协议将避免传统协议下对于可信第三方的依赖。另外,为了在更强的零知识概念下(如并发零知识、可重置零知识)进一步降低协议的轮复杂性,Canetti 等[62]提出了一个比 CRS 模型更弱的纯公钥(BPK, Bare Public Key)模型。与参数篡改模型类似,BPK 模型也不需要可信第三方来生成 CRS,而是验证者在与证明者执行协议之前在公共文件夹里放上自己的公钥,显然该模型下也存在参数篡改所带来的威胁。事实上 BPK 模型下的验证者就是参数篡改模型下的参数篡改敌手,正如文献[63]所证,这 2 个模型几乎可以被认为是等价的。文献[61]第一个研究了在非交互零知识证明(NIZK, Non-interactive Zero Knowledge Proofs)的可信参数被篡改时所保持的安全性定义:防篡改可靠性与防篡改零知识性,并揭示了 2 者无法同时满足;此外,他们证明了同时满足防篡改零知识性和可靠性的协议是存在的,并构造了一个满足这些性质的 NIZK 协议,但该 NIZK 协议只有理论意义,实用价值不高。文献[64]基于提出新的知识假设,给出了比文献[61]更强的抗参数篡改的安全性定义,在文献[7]的方案下通过在 CRS 中添加部分信息并让证明者执行一个高效的 CRS 验证算法,以确保 CRS 的良好结构,使得其同时具备防篡改零知识性和知识可靠性(Knowledge Soundness)。文献[65]同样提出了新的知识假设以及新的模拟技术,在不扩大 CRS 尺寸的情况下,遵循文献[61]的安全性定义并展示了包括文献[7]提出的方案在内的大多数基于 QAP 的 zk-SNARK 方案都可以同时实现防篡改零知识性和知识可靠性。文献[66]通过改进文献[64]的方案,总结了构造抗参数篡改 zk-SNARK 的一般方法,并将文献[65]中的模拟技术应用于文献[64]的方案之中。

文献[63]认为 CRS 模型下的防篡改零知识性等价于 BPK 模型下的没有辅助输入的非黑盒的零知识。文献[67]提出了个体化规约技术并利用这个技术构造了 2 轮延迟输入的零知识协议等应用,事实上也提供了一个构造 BPK 模型下的 NIZK 的构造框架。这个框架目前只能被用来构造 BPK 模型下的 NIZK 方案而不是

① 需要特别说明的是,在 Crypto 2021 的文献[60]中,Block 等指出在 DARK 多项式承诺方案的安全性证明中存在着一处错误。目前尚不知这个错误能否在方案不变的情况下被修复。

BPK 模型下的 zk-SNARK。尽管如文献[7]中的十分高效的 zk-SNARKs 已经有办法被改造成抗参数篡改的版本,但是这些改造都需要引入不可证伪的知识假设并且它们的零知识性只能满足防篡改零知识,换句话说,在 BPK 模型下只能实现没有辅助输入的非黑盒零知识。除此之外,文献[68]还研究了如何在零知识方面保证防篡改零知识性的同时在可靠性方面保证模拟可抽取(Simulation-Extractable)。

6 模拟多方安全计算的 zk-SNARKs

姚期智先生在 1982 年提出了著名的百万富翁问题,进而诞生了安全多方计算(Secure Muti-part Computation, MPC)这一概念[69]。MPC 技术能使彼此互不信任的多方协同计算任何函数并输出计算结果,并保证任何一方均无法得到除应得的计算结果之外的其他任何信息。2007 年, Ishai 等[70]首次提出 MPC-in-the-head 的概念,这个概念是指证明者先在"头脑中"执行 MPC 协议然后再对"头脑中"各个参与方的看到的副本进行承诺并发送给验证者,随后验证者 V 再与 P 交互,来判断是否相信 P 的断言。基于这个框架做的 zk-SNARKs 的优点是不需要可信的设置、证明时间高效而且抗量子攻击。基于这个范式, Giacomelli 等[71]发表了 ZKBoo 方案,该方案相比方案[70]取消了交互,且验证时间和在此之前的 zk-SNARKs 差不多的情况下,证明时间几乎快了 1 000 倍,缺点是其通信复杂度与电路大小呈线性关系。ZKBoo 主要是通过用一个电路分解的方法来取代 MPC 协议从而扩展了方案[70]的想法,这意味着该证明者不需要完全遵循 MPC 协议的性质从而获得更高的效率。Chase 等[72]对方案[71]进行了改进,在不增加计算成本的前提下将证明长度减少了一半以上。Ames 等[73]将高效的安全多方计算和 Ishai 等[70]提出的将 MPC 协议转化为零知识交互 PCPs(ZKIPCP)的一般转换的变形相结合,从而将通信复杂度几乎降低到了验证电路大小的根号级别,其证明长度大约比方案[72]短 4 倍。

7 总 结

零知识证明理论在密码学中处于极其重要的基础地位。从应用上来说,零知识证明所具有的强大的隐私保护特性使得它被广泛应用于身份认证、群签名、安全多方计算、密码货币、电子投票等领域之中。从零知识证明到非交互零知识证明再到 zk-SNARKs,目前的零知识证明在功能性、简洁性和效率上都已经做得很好但仍有很多提升空间。尽管目前文献[7]提出的 CRS 模型下的方案在证明长度和验证效率上达到了最优,但在是现实上依然存在中心化的问题。而通用可更新的 zk-SNARKs 的研究目前比较热门,出现了很多各有优劣的新方案,未来主要是研究如何在效率和通信量上做到更优。透明的 zk-SNARKs 相比较存在可信第三方的 SNARKs,其通信量与验证者时间仍有较大差距,这也是其实用化的最大障碍,因此,其研究重点在于如何进一步减少验证者计算量与通信量。而 BPK 模型下的 zk-SNARK 主要依靠添加额外 CRS 验证算法来解决 CRS 模型下的中心化问题,但目前这个验证算法的的运行时间并不理想,未来如何提出更高效的解决方式是一个待研究的课题。模拟多方安全计算的 zk-SNARKs 相比于其它 SNARKs 的缺点主要在于通信量,因此其研究重点在于如何在保证效率的同时设计通信量更小的方案。

参考文献:

[1] Goldwasser S, Micali S, Rackoff C. The knowledge complexity of interactive proof systems[J]. SIAM Journal on Computing, 1989, 18(1): 186-208.

[2] Blum M, Feldman P, Micali S. Non-interactive zero-knowledge and its applications[C]//The 1988 Symposium on the Theory of Computing. New York: Association for Computing Machinery, 1988: 103-112.

[3] Kilian J. A note on efficient zero-knowledge proofs and arguments[C]//The 24th Annual ACM Symposium on the Theory of Computing. New York: Association for Computing Machinery, 1992: 723-732.

[4] Micali S. Computationally sound proofs[J]. SIAM Journal on Computing, 2000, 30(4): 1253-1298.

[5] Groth J. Short Pairing-based non-interactive zero-knowledge arguments[C]//Advances in Cryptology -ASIACRYPT 2010. Berlin: Springer, 2010: 321-340.

[6] Gennaro R, Gentry C, Parno B, et al. Quadratic span programs and succinct NIZKs without PCPs[C]//Advances in Cryptology-EUROCRYPT 2013. Berlin: Springer, 2013: 626-645.

[7] Groth, J. On the size of pairing-based non-interactive arguments[C]//Advances in Cryptology-EUROCRYPT 2016. Berlin:

Springer, 2016: 305-326.

[8] Goldreich O, Oren Y. Definitions and properties of zero-knowledge proof systems[J]. Journal of Cryptology, 1994, 7: 1-32.

[9] Babai L, Fortnow L, Levin L, et al. Checking computations in polylogarithmic time[C]//The 23th Annual ACM Symposium on the Theory of Computing. New York: Association for Computing Machinery, 1991: 21-32.

[10] Feige U, Goldwasser S, Lovasz L, et al. Interactive proofs and the hardness of approximating cliques[J]. Journal of the ACM, 1996, 43(2): 268-292.

[11] Arora S, Safra S. Probabilistic checking of proofs: A new characterization of NP[J]. Journal of the ACM, 1998, 45(1): 70-122.

[12] Arora S, Lund C, Motwani R, et al. Proof verification and the hardness of approximation problems[J]. Journal of the ACM, 1998, 45(3):501-555.

[13] Kilian J. Improved efficient arguments (preliminary version)[C]//Advances in Cryptology-CRYPTO'95. Berlin: Springer, 1995: 311-324.

[14] Fiat A, Shamir A. How to prove yourself: Practical solutions to identification and signature problems[C]//Advances in Cryptology-CRYPTO'86. Berlin: Springer, 1987: 186-194.

[15] Aiello W, Bhatt S, Ostrovsky R, et al. Fast verification of any remote procedure call: Short witness-indistinguishable one-round proofs for NP[C]//Proceedings of the 27th International Colloquium on Automata, Languages and Programming. Berlin: Springer, 2000: 463-474.

[16] Dwork C, Langberg M, Naor M, et al. Succinct NP proofs and spooky interactions[EB/OL]. (2005-01-30)[2021-08-12]. https://www.researchgate.net/publication/228774797_Succinct_Proofs_for_NP_and_Spooky_Interactions.

[17] Mie T. Polylogarithmic two-round argument systems[J]. Journal of Mathematical Cryptology, 2008, 2(4):343-363.

[18] Crescenzo G, Lipmaa H. Succinct NP proofs from an extractability assumption[C]//Proceedings of the 4th Conference on Computability in Europe. Berlin: Springer, 2008: 175-185.

[19] Bitansky N, Canetti R, Chiesa A, et al. From extractable collision resistance to succinct non-interactive arguments of knowledge, and back again[C]//Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. New York, Association for Computing Machinery, 2012: 326-349.

[20] Damgard I, Faust S, Hazay C. Secure two-party computation with low communication[C]//Proceedings of the 9th Theory of Cryptography Conference. Berlin, Springer, 2012: 54-74.

[21] Goldwasser S, Lin H, Rubinstein A. Delegation of computation without rejection problem from designated verifier CS-proofs[J]. Cryptology and Information Security Series,2011,2011:456.

[22] Bitansky N, Chiesa A. Succinct arguments from multi-prover interactive proofs and their efficiency benefits[C]//Advances in Cryptology-CRYPTO 2012. Berlin: Springer, 2012: 255-272.

[23] Barak B, Goldreich O. Universal arguments and their applications[J]. SIAM Journal on Computing, 2008, 38(5): 1661-1694.

[24] Ishai Y, Kushilevitz E, Ostrovsky R. Efficient arguments without short PCPs[C]//Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity. Piscataway: IEEE, 2007: 278-291.

[25] Kalai Y, Raz R. Interactive PCP[C]//Proceedings of the 35th International Colloquium on Automata, Languages and Programming. Berlin, Springer, 2008: 536-547.

[26] Setty S, Blumberg A, Walfish M. Toward practical and unconditional verification of remote computations[C]//Proceedings of the 13th USENIX Conference on Hot Topics in Operating Systems. Berkeley: USENIX Association, 2011: 29-29.

[27] Setty S, McPherson R, Blumberg A, et al. Making argument systems for outsourced computation practical (sometimes)[C]//Network Distributed System Security Symposium (NDSS). San Diego: The Internet Society, 2012: 180-199.

[28] Setty S, Vu V, Panpalia N, et al. Taking proof-based verified computation a few steps closer to practicality[C]//Proceedings of the 21st USENIX Security Symposium. Berkeley: USENIX Association, 2012: 253-268.

[29] Setty S, Braun B, Vu V, et al. Resolving the conflict between generality and plausibility in verified computation[J]. Cryptology and Information Security Series,2012,2012:622.

[30] Groth J, Ostrovsky R, Sahai A. New techniques for noninteractive zero-knowledge[J]. Journal of the ACM, 2012, 59(3):

1-35.

[31] Groth J, Ostrovsky R, Sahai A. Non-interactive zaps and new techniques for NIZK[C]//Advances in Cryptology-CRYPTO 2006. Berlin: Springer, 2006: 97-111.

[32] Groth J. Simulation-sound NIZK proofs for a practical language and constant size group signatures[C]//Advances in Cryptology-ASIACRYPT 2010. Berlin: Springer, 2006: 444-459.

[33] Groth J, Sahai A. Efficient noninteractive proof systems for bilinear groups[J]. SIAM Journal on Computing, 2012, 41(5): 1193-1232.

[34] Groth J. Linear algebra with sub-linear zero-knowledge arguments[C]//Advances in Cryptology-CRYPTO 2009. Berlin: Springer, 2009: 192-208.

[35] Damgard I. Towards practical public key systems secure against chosen ciphertext attacks[C]//Advances in Cryptology-CRYPTO'92. Berlin: Springer, 1992: 445-456.

[36] Hada S, Tanaka T. On the existence of 3-round zero-knowledge protocols[C]//Advances in Cryptology-CRYPTO'98. Berlin: Springer, 1998: 408-423.

[37] Bellare M, Palacio A. The knowledge-of-exponent assumptions and 3-round zero-knowledge protocols[C]//Advances in Cryptology-CRYPTO 2004. Berlin: Springer, 2004: 273-289.

[38] Lipmaa H. Progression-free sets and sublinear pairing-based non-interactive zero-knowledge arguments[C]//Proceedings of the 9th Theory of Cryptography Conference. Berlin: Springer, 2012: 169-189.

[39] Gentry C, Wichs D. Separating succinct non-interactive arguments from all falsifiable assumptions[C]//The 43rd Annual ACM Symposium on Theory of Computing. New York: Association for Computing Machinery, 2011: 99-108.

[40] Lipmaa H. Succinct non-interactive zero knowledge arguments from span programs and linear error-correcting codes[C]//Advances in Cryptology-ASIACRYPT 2013. Berlin: Springer, 2013: 41-60.

[41] Danezis G, Fournet C, Groth J, et al. Square span programs with applications to succinct NIZK arguments[C]//Advances in Cryptology-ASIACRYPT 2014. Berlin: Springer, 2014: 532-550.

[42] Bitansky N, Chiesa A, Ishai Y, et al. Succinct non-interactive arguments via linear interactive proofs[C]//Theory of Cryptography, TCC 2013. Berlin: Springer, 2013: 315-333.

[43] Xie T, Zhang J, Zhang Y. Libra: Succinct zero-knowledge proofs with optimal prover computation[C]//Advances in Cryptology-CRYPTO 2019. Cham: Springer, 2019: 733-764.

[44] Groth J, Kohlweiss M, Maller M, et al. Updatable and universal common reference strings with applications to zk-SNARKs[C]//Advances in Cryptology-CRYPTO 2018. Cham: Springer, 2018: 698-728.

[45] Maller M, Bowe S, Kohlweiss M, et al. Sonic: Zero-knowledge SNARKs from linear-size universal and updatable structured reference strings[C]//2019 ACM SIGSAC Conference on Computer and Communications Security. New York: Association for Computing Machinery, 2019: 2111-2128.

[46] Bünz B, Bootle J, Boneh D, et al. Bulletproofs: Short proofs for confidential transactions and more[C]//2018 IEEE Symposium on Security and Privacy. Piscataway: IEEE Computer Society, 2018: 315-334.

[47] Kate A, Zaverucha G, Goldberg I. Constant-size commitments to polynomials and their applications[C]//Advances in Cryptology-ASIACRYPT 2010. Berlin: Springer, 2010: 177-194.

[48] Chiesa A, Hu Y, Maller M, et al. Ward. Marlin: Preprocessing zkSNARKs with universal and updatable SRS[C]//Advances in Cryptology-EUROCRYPT 2020. Cham: Springer, 2020: 738-768.

[49] Babai L, Fortnow L, Levin L, et al. Checking computations in polylogarithmic time[C]//The 23rd Annual ACM Symposium on Theory of Computing. New York: Association for Computing Machinery, 1991: 21-32.

[50] Gabizon A, Williamson Z, Ciobotaru O. PLONK: Permutations over lagrange-bases for oecumenical noninteractive arguments of knowledge[J]. Cryptology and Information Security Series,2019,2019:953.

[51] Bayer S, Groth J. Efficient zero-knowledge argument for correctness of a shuffle[C]//Advances in Cryptology-EUROCRYPT 2012. Berlin: Springer, 2012: 263-280.

[52] Campanelli M, Faonio A, Fiore D, et al. Lunar: A toolbox for more efficient universal and updatable zkSNARKs and commit-and-prove extensions[C]//Advances in Cryptology-ASIACRYPT 2021. Berlin: Springer, 2021: 3-33.

[53] Ben-Sasson E, Bentov I, Horesh Y, et al. Scalable zero knowledge with no trusted setup[C]//Advances in Cryptology-

CRYPTO 2019. Cham: Springer, 2019: 701-732.
[54] Wahby R, Tzialla I, Shelat A, et al. Doubly-efficient zkSNARKs without trusted setup[C]//2018 IEEE Symposium on Security and Privacy. Piscataway: IEEE Computer Society, 2018: 926-943.
[55] Ben-Sasson E, Chiesa A, Riabzev M, et al. Ward. Aurora: Transparent succinct arguments for R1CS[C]//Advances in Cryptology-EUROCRYPT 2019. Cham: Springer, 2019: 103-128.
[56] Chiesa A, Ojha D, Spooner N. Fractal: Post-quantum and transparent recursive proofs from holography[C]//Advances in Cryptology-EUROCRYPT 2020. Cham: Springer, 2020: 769-793.
[57] Setty S. Spartan: Efficient and general-purpose zkSNARKs without trusted setup[C]//Advances in Cryptology-CRYPTO 2020. Cham: Springer, 2020: 704-737.
[58] Bootle J, Cerulli A, Chaidos P, et al. Efficient zero-knowledge arguments for arithmetic circuits in the discrete log setting [C]//Advances in Cryptology-EUROCRYPT 2016. Berlin: Springer, 2016: 327-357.
[59] Bünz B, Fisch B, Szepieniec A. Transparent SNARKs from DARK compilers[C]//Advances in Cryptology-EUROCRYPT 2020. Cham: Springer, 2020: 677-706.
[60] Block A, Holmgren J, Rosen A, et al. Time- and space-efficient arguments from groups of unknown order[C]//Advances in Cryptology-CRYPTO 2020. Cham: Springer, 2021: 123-152.
[61] Bellare M, Fuchsbauer G, Scafuro A. NIZKs with an untrusted CRS: Security in the face of parameter subversion[C]//Advances in Cryptology-ASIACRYPT 2016. Berlin: Springer, 2016: 777-804.
[62] Canetti R, Goldreich O, Goldwasser S, et al. Resettable zero-knowledge[C]//The 32nd Annual ACM Symposium on the Theory of Computing. New York: Association for Computing Machinery, 2000: 235-244.
[63] Abdolmaleki B, Lipmaa H, Siim J, et al. On QA-NIZK in the BPK model[C]//Public-Key Cryptography-PKC 2020. Cham: Springer, 2020: 590-620.
[64] Abdolmaleki B, Baghery K, Lipmaa H, et al. A subversion-resistant SNARK[C]//Advances in Cryptology-ASIACRYPT 2017. Cham: Springer, 2017: 3-33.
[65] Fuchsbauer G. Subversion-zero-knowledge SNARKs[C]//Public-Key Cryptography-PKC 2018. Cham: Springer, 2018: 315-347.
[66] Abdolmaleki B, Lipmaa H, Siim J, et al. On subversion-resistant SNARKs[J]. Journal of Cryptology, 2021, 34(3): 1-42.
[67] Deng Y. Individual simulations[C]//Advances in Cryptology-ASIACRYPT 2020. Cham: Springer, 2020: 805-836.
[68] Baghery K. Subversion-resistant simulation (knowledge) sound NIZKs[C]//IMA International Conference on Cryptography and Coding-IMACC 2019. Cham: Springer, 2019: 42-63.
[69] Yao A. Protocols for secure computations[C]//The 23rd Annual Symposium on Foundations of Computer Science. Piscataway: IEEE, 1982: 160-162.
[70] Ishai Y, Kushilevitz E, Ostrovsky R, et al. Zero-knowledge proofs from secure multiparty computation[J]. SIAM Journal on Computing, 2009, 39(3): 1121-1152.
[71] Giacomelli I, Madsen J, Orlandi C. Zkboo: Faster zero-knowledge for boolean circuits[C]//In Proceedings of the 25th USENIX Security Symposium. Berkeley: USENIX Association, 2016: 1069-1083.
[72] Chase M, Derler D, Goldfeder S, et al. Post-quantum zero-knowledge and signatures from symmetric-key primitives[C]//2017 ACM SIGSAC Conference on Computer and Communications Security. New York: Association for Computing Machinery, 2017: 1825-1842.
[73] Ames S, Hazay C, Ishai Y, Venkitasubramaniam M. Ligero: lightweight sublinear arguments without a trusted setup[C]//2017 ACM SIGSAC Conference on Computer and Communications Security. New York: Association for Computing Machinery, 2017: 2087-2104.

【责任编辑: 孙向荣】