摘要

最近报道的监视和安全漏洞事件增加了用户的隐私权，对当前模型提出质疑，在该模型中，第三方收集并控制大量个人数据。 比特币已经证明在财务领域中，可以进行可信的，可审核的计算。在公众的陪同下使用分散的同行网络分类帐。 在本文中，我们描述了去中心化的个人数据确保用户拥有并控制他们的管理系统数据。 我们实现了将区块链转变为不需要信任的自动访问控制管理器第三方。 与比特币不同，我们系统中的交易不是严格的财务–它们用于执行说明，例如存储，查询和共享数据。 最后，我们讨论可能区块链的未来扩展可以利用它们针对社会中可信计算问题的全面解决方案。

Keywords blockchain; privacy; bitcoin; personal data

1. INTRODUCTION

我们世界中的数据量正在迅速增加。根据最近的一份报告[22]，估计有20％过去两年中收集了世界各地的数据。最大的在线社交网络Facebook收集了300个自成立以来已达到PB的个人数据[1] –一百倍于国会图书馆收集的金额200年[13]。在大数据时代，数据不断收集和分析，从而带来创新和经济增长。公司和组织使用他们收集的数据来个性化服务，优化公司决策过程，预测未来趋势等等。今天的数据是是我们经济中的宝贵资产[21]。尽管我们都从数据驱动型社会中受益，公众越来越关注用户隐私。集中大量的公共和私人组织个人和敏感信息。个人很少或无法控制关于它们的数据及其存储方式用来。近年来，公共媒体屡屡报道

与隐私有关的有争议事件。在更好的之中已知的例子是有关政府监督的故事[2]和Facebook的大规模科学实验显然没有明确告知参加者[10]。

**相关工作**。从立法的角度出发，已经进行了各种尝试来解决这些隐私问题。（[4]，[20]），以及从技术角度来看。 OpenPDS是最近开发的框架，它提供了用于PDS自主部署的模型，该模型包括一种用于返回数据上的计算从而返回的机制。回答而不是原始数据本身[6]。在整个行业中领先的公司选择实施自己的专有基于OAuth协议[19]的身份验证软件他们充当集中信任的权威。从安全角度来看，研究人员开发了各种针对隐私问题的技术侧重于个人数据。数据匿名化方法试图保护个人身份信息。 k-匿名，一个共同的属性匿名数据集要求每个记录与至少k-1个其他记录没有区别[24]。 k匿名性的相关扩展包括l-多样性，这样可以确保敏感数据由多种形式表示足够的可能值集[15]；和t紧密度着眼于敏感数据的分布[14]。最近的研究已经展示了如何使用这些数据来匿名化数据集可以将技术取消匿名[18]，[5]，即使是很小的数据点或高维数据的数量。其他隐私保护方法包括差分隐私，这种技术会干扰数据或在计算过程中增加噪声共享数据之前进行处理[7]，以及加密方案允许通过加密运行计算和查询数据。具体来说，完全同态加密（FHE）[9]方案允许任何计算运行在加密数据上但目前效率太低，无法在实践中广泛使用。近年来，一类新的责任制出现了。第一个这样的系统是比特币，它允许用户可以使用可公开验证的开放式分类帐来安全地转移货币（比特币），而无需中央监管机构区块链）。从那时起，其他项目（统称为到比特币2.0 [8]）展示了这些区块链如何可以满足需要可信计算的其他功能，并且可审核性。

**我们的贡献**。 1）我们结合区块链和区块链存储来构建个人数据管理专注于隐私的平台。 2）我们通过我们的插图平台以及对未来改进的讨论技术，区块链如何成为重要资源在可信计算中。

**组织**。 第二部分讨论了我们的隐私问题解决本文； 第三节概述了平台，而第四节详细介绍了技术实施； 第五节讨论了将来的扩展 第六节中找到了区块链及其结论。

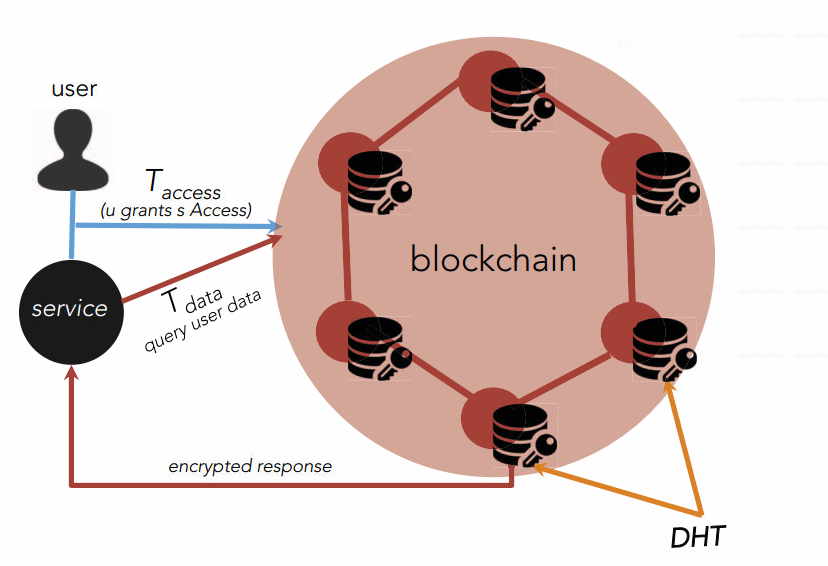
1. **THE PRIVACY PROBLEM**

在整篇文章中，我们都会解决隐私问题用户在使用第三方服务时面临的问题。我们特别专注于移动平台，在该平台上服务可以部署应用程序供用户安装。这些应用程序不断收集用户没有特定信息的高分辨率个人数据知识或控制。在我们的分析中，我们假设服务是诚实而又好奇的（即它们遵循协议）。请注意，同一系统可用于其他数据隐私问题，例如患者共享其医疗数据用于科学研究，同时具有监控方法它具有立即退出的功能。有鉴于此，我们的系统可以防止以下常见的隐私问题：数据所有权。我们的框架致力于确保用户拥有并控制其个人数据。这样，系统将用户识别为数据和服务的所有者作为具有委派权限的访客。数据透明度和可审核性。每个用户都有关于收集哪些数据的完全透明她以及如何访问它们。细粒度的访问控制。一个主要关注点移动应用程序是要求用户授予一组注册时的权限设置。这些权限被授予无限期地更改协议的唯一方法是选择退出。相反，在我们的框架中，在任何给定时间用户可以更改权限集并撤消对先前收集的数据。该机制的一种应用将是改进移动设备中的现有权限对话框应用程序。虽然用户界面可能会保留同样，访问控制策略将安全地存储在区块链，仅允许用户更改它们。

1. **PROPOSED SOLUTION**

我们从系统概述开始。如图所示在图1中，构成我们系统的三个实体是有兴趣下载和使用的手机用户应用；服务，此类应用程序的提供者出于运营和业务相关原因要求处理个人数据（例如，定向广告，个性化服务）；和节点，负责维护区块链的实体和分布式私有键值数据存储，以换取激励。请注意，虽然系统中的用户通常会保留（伪）匿名，我们可以将服务配置文件存储在区块链上并验证其身份。系统本身的设计如下。区块链接受两种新的交易类型：Taccess，用于访问控制管理；和Tdata，用于数据存储和检索。这些网络操作可以轻松地集成到服务可以使用的移动软件开发套件（SDK）在他们的发展过程中。为了说明，请考虑以下示例：用户安装使用我们平台保护她隐私的应用程序。当用户首次注册时，一个新的共享（用户，生成服务身份并将其与关联的权限一起发送到Taccess交易中的区块链。手机上收集的数据（例如位置等传感器数据）使用共享的加密密钥加密并发送到Tdata交易中的区块链，随后将其路由到区块链外的键值存储，同时仅保留指向公共分类帐上的数据的指针（该指针是数据的SHA-256哈希值）。服务和用户现在都可以使用以下方法查询数据关联了指针（键）的Tdata事务。的区块链然后验证数字签名属于用户或服务。对于服务，其权限访问数据也会被检查。最后，用户可以随时通过以下方式更改授予服务的权限发出具有一组新权限的Taccess事务，包括撤消对以前存储的数据的访问。发展基于Web的（或移动的）仪表板，可进行概述数据和更改权限的能力相当微不足道，类似于开发集中式钱包，例如比特币的Coinbase1。区块链外键值存储是一个实现Kademilia [16]，一个分布式哈希表（或DHT​​），添加了使用LevelDB2和与区块链的接口实现持久性DHT由节点网络维护（可能

与区块链网络脱节）读/写事务。数据在节点并进行复制以确保高可用性。它是有启发性的指出替代区块链解决方案可以考虑用于存储。例如，集中式云可能用于存储数据。虽然这需要一些对第三方的信任程度，它在某些方面具有优势可扩展性和易于部署的特性。





IV. THE NETWORK PROTOCOL

现在我们详细描述所使用的底层协议在系统中。我们利用标准的密码学构建我们平台中的块：定义的对称加密方案由三元组（Genc，Eenc，Denc）–生成器，加密以及解密算法；数字签名三元组（Gsig，Ssig，Vsig）描述的方案（DSS）–生成器，签名和验证算法，分别使用带有secp256k1曲线的ECDSA实现[12]；和由SHA-256实例化的加密哈希函数H[11]实施。

1. *Building Blocks*

现在，我们简要介绍相关的构建基贯穿本文的其余部分。我们假设熟悉比特币[17]和区块链。

1）身份：区块链利用伪身份机制。本质上是公共密钥，每个用户都可以生成她想要增加许多这样的伪身份隐私。现在我们引入复合身份，这是扩展在我们的系统中使用了该模型。复合身份是两个或多个参与方的共享身份，其中一些各方（至少一个）拥有身份（所有者），其余限制访问它（来宾）。协议1说明了单个所有者（用户）和单个来宾的实现（服务）。如图所示，身份由签名组成所有者和来宾的密钥对，以及对称的用于加密（和解密）数据的密钥，以便数据受系统其他所有玩家的保护。正式地，复合身份在外部（如网络所见）由2元组观察：

2）区块链内存：我们让L为区块链内存空间，表示为hastable L：{0，1} 256→

{0，1} N，其中N >> 256，可以存储足够大的文档。我们假设该内存在比特币和比特币使用的相同对抗模型下是防篡改的其他区块链。为了直观地解释为什么如此受信任数据存储可在任何区块链上实现（包括比特币），请考虑以下简化方法，尽管效率低下，实现：区块链是带有时间戳的序列交易，其中每个交易包括可变数量的输出地址（每个地址是一个160数字）。大号然后可以如下实现–前两个输出在事务中编码256位内存地址指针，以及一些辅助元数据。其余的输出构造序列化的文档。查找L [k]时，仅返回最近的事务，这允许更新

和插入操作之外的删除操作。

3）策略：用户u授予服务的一组权限，用POLICYu，s表示。例如，如果您安装了手机

需要访问用户的位置和联系人的应用程序，然后政策u，s = {位置，联系人}。这对请注意，任何类型的数据都可以通过这种方式安全地存储，假设服务不会破坏协议和标签数据不正确。可以部分防止这种情况发生的保障措施引入移动SDK，但无论如何，用户可以轻松检测到欺骗的服务，因为所有更改都是对她可见。

4）辅助功能：Parse（x）反序列化发送到事务的消息，其中包含参数；协议2中所示的CheckP olicy（pksig k，xp）验证了发起者具有适当的权限。

B.区块链协议

这里我们提供了核心协议的详细描述在区块链上执行。协议3由节点执行在网络中收到Taccess交易时，以及类似地，针对Tdata事务执行协议4。如本文前面所述，Taccess事务

允许用户更改授予某项权限的权限集服务，通过发送POLICYu集。发送空集撤销先前授予的所有访问权限。发送Taccess首次具有新复合身份的交易是解释为用户注册服务。同样，Tdata事务控制读/写操作。借助CheckP政策，只有用户（始终）或服务（如果允许）可以访问数据。请注意，在第9行和协议4的16中，我们使用了简写形式来访问DHT就像普通的哈希表一样。实际上，这些说明导致链外网络消息（已读或已读）写入）发送到DHT。

C.隐私和安全性分析

我们依赖于区块链无篡改的假设，该假设需要足够大的不受信任的网络同行。另外，我们假设用户管理她的密钥以安全的方式，例如使用安全集中式钱包服务。现在，我们展示我们的系统如何防范攻击者破坏了系统中的节点。目前，我们 较少关注会改变协议或记录以前读取的数据，因为它们很可能是信誉良好，但我们为此类行为提供了可能的解决方案在V-A节中。在这种模型下，只有用户可以控制她的数据。区块链的分散性质与数字签名交易可确保对手无法冒充用户或破坏网络，这意味着对手伪造了数字签名，或者控制了网络的大部分资源。同样，一个对手无法从公共分类帐中学习任何东西，因为仅散列指针存储在其中。控制一个或多个DHT节点的对手不能了解有关原始数据的任何信息，因为它是用密钥加密的没有节点。请注意，虽然数据完整性是不能在每个节点中确保，因为单个节点可以篡改它的本地副本或以拜占庭方式行事，我们仍然可以在实践中通过充分分发和复制使风险最小化数据。

最后，为每个用户服务对生成新的复合标识可确保仅一小部分数据如果对手同时获得两个

签名和加密密钥。如果对手仅获得一名键，则数据仍然安全。注意在实践中我们可以进一步分离身份，以限制单一受损的化合物身份。例如，我们可以为每存储一百条记录生成新密钥。在本节中，我们略微介绍了可能的方法未来对区块链的扩展。这些可以发挥重要作用在塑造更成熟的分布式可信计算中的作用平台，而不是当前最先进的系统。更多具体来说，它们将大大提高较早提出的平台。

A.从存储到处理

本文的主要贡献之一是演示如何克服区块链的公共性。所以到目前为止，我们的分析重点是存储指向加密数据的指针。尽管此方法适用于存储和随机查询，它对于处理数据不是很有效。更重要的是，服务一旦查询了一些原始数据，便可以将其存储为未来分析。

更好的方法可能是永远不要让服务观察原始数据，但允许其运行计算直接在网络上获取最终结果。如果我们分裂数据共享（例如，使用Shamir的秘密共享[23]），而不是加密它们，然后我们可以使用安全的多方计算（MPC）安全地评估任何功能

[3]。

在图2中，我们说明了MPC如何与区块链，特别是在我们的框架中。考虑一个一个简单的例子，其中一个城市举行选举并祝愿允许在线秘密投票。它开发了一个移动应用程序利用我们的系统进行投票，现在增强了具有建议的MPC功能。在线选举后发生后，城市随后提交其后端代码汇总结果。网络选择节点的子集随机地，解释器将代码转换成安全的MPC协议。最终，结果被公开存储分类帐，可以安全地防止篡改。结果，没有人们可以了解个人投票是什么，但每个人都可以查看选举结果。

图2.区块链网络中安全计算流程的示例。的左上方的块（EVote过程）是不安全的代码，其中的参数（\*）中标记的是私有的，并作为共享存储在DHT中。网络随机选择一个节点子集以计算EVote的安全版本，然后将结果广播回整个网络，然后将其存储在分类帐中。

B.区块链中的信任和决策

比特币或一般的区块链假定所有节点都是同样不受信任，并且他们在集体中所占的比例

决策过程完全基于他们的计算资源（称为工作量证明算法）[17]。在换句话说–对于每个节点n，信任∝ resources（n）（根据概率）确定节点的权重。这个导致不良反应，最明显的是易患紫杉醇攻击，过多的能量消耗和高延迟通过此措施，网络可以将更多权重分配给受信任的节点和计算块更有效。由于它需要时间来赢得对系统的信任，它应该对sybil攻击。这种机制可能会吸引其他人攻击类型，例如节点增加其声誉在以后进行恶意行为。这可以通过缓解随机选择几个节点，并对其信任进行加权，以对每个街区进行投票，然后取均等的多数

投票。这应该防止单个演员过多影响力，无论他们的信任度如何。

**VI。结论**

个人数据和一般敏感数据不应受到第三方的信任，它们很容易受到攻击和滥用。相反，用户应拥有和控制其数据而不会损害安全性或限制公司和当局提供个性化服务的能力服务。我们的平台通过结合区块链来实现这一目标，通过链外存储解决方案重新用作访问控制主持人。无需用户信任任何第三方，并且始终了解被收集关于他们以及如何使用它。此外，区块链将用户识别为他们的所有者个人资料。反过来，公司可以专注于利用数据不必过于担心正确的保护和划分他们。此外，借助去中心化平台，和有关收集，存储和共享的法规决策敏感数据应该更简单。此外，法律法规可以被编程到区块链本身中，这样它们是自动执行的。在其他情况下，分类帐可以作为访问（或存储）数据的法律依据，因为它（按计算）是防篡改的。