卞立平,孙爱东,孙晓明,等. 基于区块链技术的农产品深度溯源系统建设思考和设计方案[J].江苏农业学报,2022,38(4):1092-1098.
doi:10.3969/j.issn.1000-4440.2022.04.028

基于区块链技术的农产品深度溯源系统建设思考和设计方案

卞立平, 孙爱东, 孙晓明, 刘贤金
(江苏省农业科学院农产品质量安全与营养研究所,江苏 南京 210014)

摘要: 针对常规农产品追溯过程中存在的数据可信度低、关联追溯难等问题,以联盟链为基础,设计构建了一套适用于农产品的追溯系统。通过节点权限管理、数据加密上链、数据存储、基于节点关联度的哈希匹配检索等模块,能够在整个供应链流程中实现真实可信、全面完整、安全高效的多级深度追溯,有效提高了农产品追溯信息的可信度、深度和广度,具有的广泛的适用范围和较好的应用前景。

关键词: 溯源;联盟链;农产品;多级关联检索
中图分类号: S126 **文献标识码:** A **文章编号:** 1000-4440(2022)04-1092-07

Thought of agricultural product deep-traceability system construction and scheme design based on blockchain technology

BIAN Li-ping, SUN Ai-dong, SUN Xiao-ming, LIU Xian-jin
(*Institute of Quality, Safety and Nutrition of Agricultural Products, Jiangsu Academy of Agricultural Sciences, Nanjing* 210014, *China*)

Abstract: Aiming at the problems of low reliability and difficult correlating in the process of conventional agricultural products traceability, a set of traceability system suitable for agricultural products was designed and constructed based on the consortium blockchain. Through functional modules such as node authority management, data encryption and chaining, data storage, hash matching retrieval module based on node association degree, it can achieve real, reliable, comprehensive, safe and efficient multi-level in-depth traceability in the whole supply chain process. It has significantly improved the reliability, depth and breadth of agricultural product traceability information, and has wide scope of application and good prospects.

Key words: traceability; consortium blockchain; agricultural product; multi-level association retrieval

当前行业内常用的溯源方法和技术,大部分仅限于读取单段农产品生产加工流通过程中存储进数据库的信息数据,并以此完成一个过程溯源查询请求。该溯源方法原理,既无法进行对供应链上涉及多企业的多级产出、投入品的多级溯源检索,也不能实现高效读取追溯信息。其方法存在三方面风险:①为某些蓄意造假的企业留下了替换与篡改质量安全信息数据的隐性漏洞;②难以实现真正的全程追溯示踪,大部分情况下会遗缺部分溯源信息;③因信息不对称,监管部门较难评判质量问题发生在哪一个具体环节[1]。

伴随区块链有关技术的不断发展,基于联盟链的可信完整供应链全环节、全过程、多级关联溯源,为监管部门和消费者全方位监督农产品的质量安全提供了可能性。本文所述产品深度溯源系统将对现行溯源方法、技术和系统所能达到的追溯能力进行升级,实现对全供应链进行真实可信、全面完整、安

收稿日期:2021-12-27
基金项目:国家重点研发计划项目(2017YFC1601000);江苏省自主创新基金项目[CX(21)3054]
作者简介:卞立平(1985-),男,江苏东台人,硕士,助理研究员,主要从事农产品质量安全管控与农业信息化相关技术研究。(Tel)025-84391572;(E-mail)blp_njau@163.com
通讯作者:刘贤金,(Tel)025-84390036;(E-mail)jaasliu@163.com

全高效的多级深度追溯,服务终端市场和监管部门。因此,对本系统的设计研究,主要所需解决的技术问题是提供一种高效智能的基于节点关联度的哈希匹配检索的实现方法,解决现有技术的弱点,如低效率运作、无法智能多级深度追溯等。

1 区块链技术

1.1 联盟链技术

联盟链和公链是当前区块链技术发展的两大方向,两者最大的差异在于是否存在“准入机制”,即针对访问权限的管理机制。因为联盟链存在“准入机制”,并非任意个人或组织都能加入,同时,数据只在联盟链内部公开和共享。而公链,则是完全公开,没有准入门槛,所有个人或组织都能加入,且数据是对所有人公开的[2]。

联盟链和公链在技术方面的异同点见表 1。整体来说,除了与公链一样具备区块链标志性的分布式账本、共识算法、防篡改等技术能力之外,联盟链还重点实现了对实名制准入机制的严格管理[3],将数据的互联互通限制在指定的可信范围内。相对于公链节点过多而造成的多节点同步效率低下、区块链性能整体偏低的问题,联盟链具有较少节点,可以快速实现信息同步,大大提高业务性能。基于这些联盟链所具备的优势,越来越多的国内外企业机构在选择过程中,选择了联盟链技术而非公链技术,比如沃尔玛、雀巢、都乐和黄金食品使用的 IBM 食品溯源链,以及用于有机食品溯源和防伪的中兴云链[4-5]。

表 1 联盟链和公链异同点比较

Table 1 Comparison of similarities and differences between alliance chain and public chain

技术性能	联盟链	公链
DLT 分布式账本	有	有
数据防篡改	有	有
采用共识算法	是	是
准入机制	是	否
信任基础	需要	不需要
节点数量	少	多
是否发币	否	是

1.2 数字签名和数字证书

数字签名是一串不可伪造的字符串,仅当发送方发出数据时才能生成,既具备不可抵赖性,还能够验证数据完整性,是确认发送方身份是否真实的有效凭据。区块链上的数字签名是基于非对称加密算法和哈希算法的融合应用[6],签名时使用私钥,验证签名时用公钥。目前,使用最为广泛的数字签名为公钥数字签名。

数字证书则用来证明某个公钥属于谁,内容是否正确。数字证书是由证书认证机构 CA 来签发的,作用就像一本真实的证书,可以证明信息真实性和合法性。数字证书内容信息包括证书序列号、版本号、签名算法类型、签发方信息、证书有效期、被签发人、公开密钥、CA 数字签名等,其中,最重要的是公开密钥和 CA 数字签名 2 个信息,因其可证明某一公钥的合法性。

1.3 区块时间戳

区块链领域的“时间戳”是一个专有名词,它能记录区块链任意时间内的交易记录和交易值。(说明:此处所说的区块“交易”,指在区块中写入信息后将区块上链,包括但并不限于金融交易)。时间戳本质具有时序性,从而使链上的区块也具有了时序性。时间戳是区块交易中的必要信息之一,它由第一个节点计算出新区块高度时就立刻标记时间戳,由该节点向全链节点广播这个区块高度及加盖的时间戳。可以说,时间戳证明了交易数据在此刻已经存在。因时间具有唯一性,因此每笔加盖时间戳的区块交易都具有唯一性,使整个区块链分布式网络能够确定地验证某个区块交易的真实性。

1.4 链上数据检索

以联盟链架构 Hypeiledger Fabric 为例,其账本数据库以文件系统为基础,且实际区块数据存储在文件块中,使用 LevelDB[7](一种高效的 key-value 数据库)用于存储区块交易的定位索引,即对应的文件块与其偏移,此设计加快了定位索引的速度。索引的内容为文件位置指针。单个指针由 3 部分构成,分别是文件编号、文件内偏移量和区块所占用字节数。但该技术方案目前仅支持用区块编号、区块哈希、交易 ID 等有限字段进行检索。

2 系统设计方案

2.1 系统功能模块设计

基于团队现有技术研究成果和实践应用结果[8-9],本文所述的农产品深度溯源系统包含如下主

要功能模块,即:联盟链成员管理、企业数据记录及管理、数据上链加密、多级关联检索与验证。产品深度溯源系统技术方案流程如图 1 所示。

“联盟链成员管理”模块,对联盟链上节点(即现实中的成员机构)进行授权与管理,同时对联盟链上节点的属性进行定义,为后续提取、计算节点关联度,实现数据定向共享和关联检索提供前提条件。

“企业数据记录及管理”模块,提供各类可溯源数据的录入与管理功能,且该模块不一定存在于链上,可在独立的应用平台上通过接口将数据接入到区块链中。

“数据上链加密”模块,将联盟链成员希望在链内共享可见的产品信息基于共识算法打包上链并创造加密区块,同时,向指定类型属性的节点签发数字证书,通过数字证书的认证可获取加密上链的信息数据。

在“多级关联检索与验证”模块中着重体现了本系统的核心技术和与其他相似功能系统的区别度。在该模块中,可实现相关数据的多级高效关联检索和哈希值快速匹配验证功能。原理为:通过对产品溯源记录的哈希值所匹配到的区块进行基础信息解析,得到该区块所对应的唯一 CORE 节点和属性有关的 RELATED 节点的历史溯源数据记档表,并以区块形成时刻排序,组成待哈希匹配、验证的区块池,用于检索下一级的相关投入品/产出品的溯源信息。

2.2 联盟链成员管理

“联盟链成员管理”模块的功能通过“联盟链管理平台”实现。该平台可对联盟链中的成员机构节点进行管理,包括:①当一个联盟成员加入联盟链后,在联盟链中创建具有对应机构主体信息节点,并标记其属性信息,如农产品生产企业、监管单位、原材料投入品供应商等。②经链上管理者审核节点信息和属性信息的正确性,激活节点可用状态。③由链上管理者设置该节点的权重等级。④基于权重等级和节点属性,联盟链将自动地为该节点授权相应功能权限。

2.3 企业数据记录及管理

“企业数据记录及管理”模块的功能通过“企业应用端”实现。企业在应用端需按照如下顺序依次录入数据:①企业主体与往来企业数据、员工数据、基础设施(地块、仓库、操作间等)数据、物联网相关数据;②投入品数据、产出品数据、生产加工标准数据、生产加工项目数据、检测报告数据、仓储物流数据、销售数据、供应链流通数据、财务数据。最后,系统将根据以上相关信息数据生成可追溯的产品对象与对应的溯源二维码。

2.4 数据上链与加密

“数据上链加密”模块涉及 2 个操作应用端,分别为“联盟链管理平台”和“企业应用端”。对数据进行打包上链和加密存证的技术方案如下:①为待上链数据打标签(数据由“企业数据记录及管理”模块采集)。全部标签类别包含数据传输头节点属性、尾节点属性、数据类型、数据摘要、数据采集方、数据采集时间;②基于共识算法,将数据打包上链并更新最新区块高度。联盟链共识算法类型为 PoS(Proof of Stake 权益证明型共识算法),其核心公式为:$\frac{Hash(K_{pub},Hash_{pb})}{B_e}<D_{cb}\times(T_{cur}-T_{pb})$,式中,$Hash$ 为哈希函数,K_{pub} 为账号公钥,$Hash_{pb}$ 为上一区块哈希值,B_e 为账户可用额度,D_{cb} 为当前区块难度系数,T_{cur} 为当前时刻,T_{pb} 为前一区块产生时刻。仅当最新区块高度的值符合上述公式时,才可确认新区块被成功创建。③将上链的信息数据加密。使用哈希算法进行数据加密操作,该算法核心是哈希函数,其主要作用是:权益证明计算,生成区块地址,将不同长度的字符串映射为较短且固定长度的比特串。核心公式为:$h=H(M)$,式中 M 为任意长度消息字符串 M,$H(M)$ 为固定长度的比特串。④向获得有关授权的联盟链成员,发放此上链数据的有效数字证书,用于验证身份、读取数据内容。

2.5 基于节点关联度的哈希匹配检索验证

“多级关联检索与验证”模块涉及 3 个操作端,分别为“企业应用端”、“深度溯源消费者查询端”、“溯源监管平台”。多级关联检索与验证模块的技术方案如图 2 所示,原理如下:

①基于节点关联度的哈希匹配检索验证,分步骤为:i)检索第一级溯源信息的哈希值所匹配的目标区块。一级溯源信息是指首次查询的起始追溯信息,次级溯源信息是指第 2 次、第 3 次……直至第 N 次查询的关联追溯信息。ii)找到步骤 i)中成功匹配验证的目标区块对应的节点,作为 CORE 节点。iii)通过 CORE 节点的节点属性找到它所有的 RELATED 节点。假设 CORE 节点 X 是 A 产品的批发商,那么 X 的 RELATED 节点包括 A 产品的生产商 Y 和零售商 Z、A 产品的投入品原料的生产商 W 等。CORE 节点的属性包含节点之间的供应关系、生产关系的相关程

度。比如,与零售商这一属性相关度最高的是批发商,而不是生产商;与原料供应商这一属性相关度最高的则为生产商,而非零售商。联盟链根据上述节点属性相关紧密度来计算节点间的关联度。iv)新区块诞生必定同时形成一个时间戳,每个节点上生成的所有具有上链溯源信息的区块,其哈希值按自然时序从近到远排列,即组成历史溯源数据记档表。同理可得一个 CORE 节点的全部 RELATED 节点的历史溯源数据记档表集合。v)基于交易合约的时间先后,排序历史溯源数据记档表,以此形成待匹配上链溯源信息哈希值的区块池。vi)用户发起对下一级的相关投入品/产出品溯源信息查询时,将在步骤 v)中形成的区块池中,寻找匹配某一哈希值(仅当查询的哈希值等于原记录数据的哈希值,视为匹配成功,反之则为匹配失败),并得到验证结果。下一级的相关投入品/产出品的溯源信息是指第 2 级及以后关联检索到的相关产品,比如第 1 级检索的产品是甲,第 2 级检索甲的投入品乙,第 3 级检索乙的投入品丙……以此类推,直至用户无须查询再下一级原料或投入品,则检索过程终止。

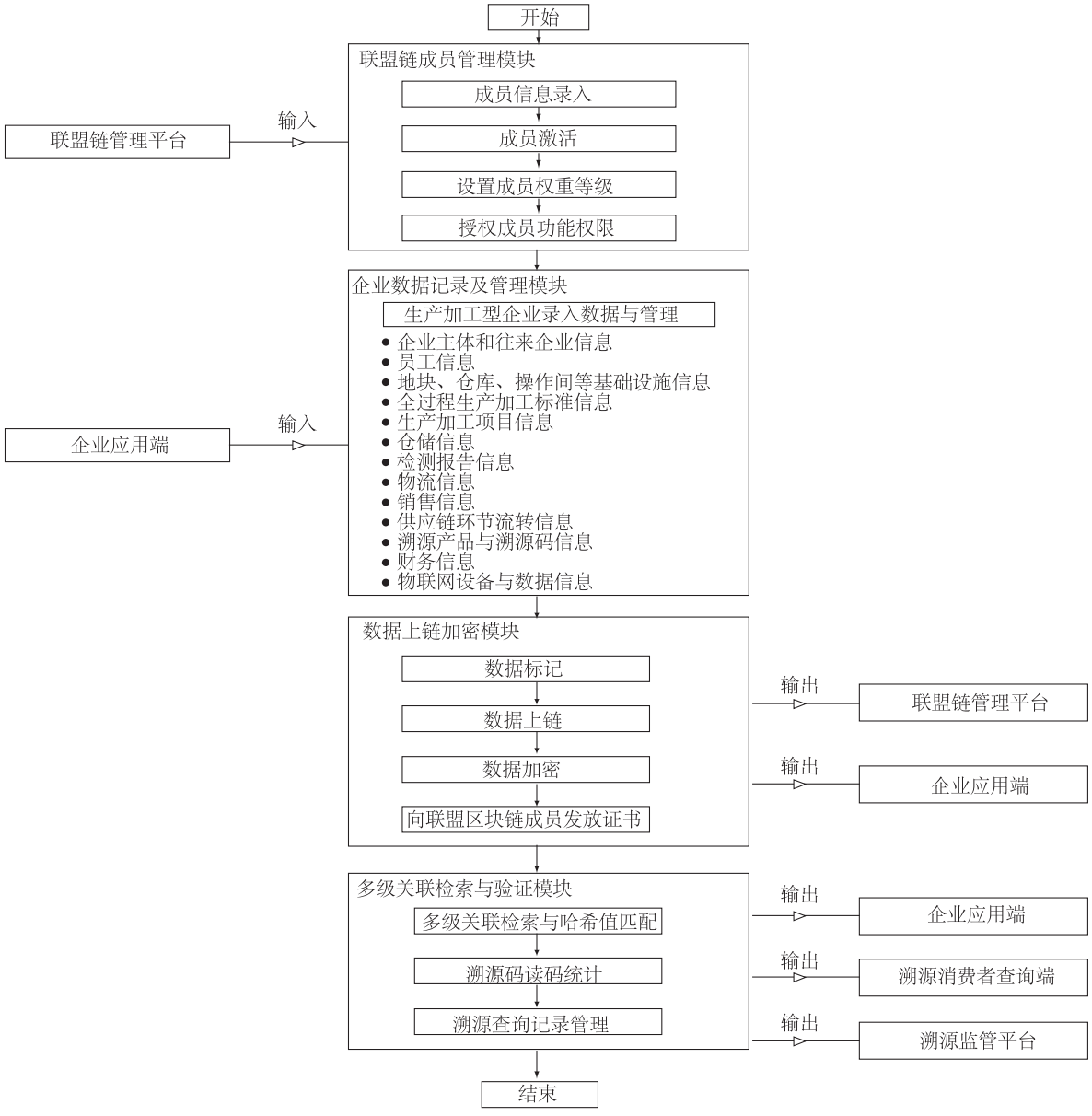

图 1　产品深度溯源系统技术方案流程图

Fig.1　Technical flow chart of product traceability system

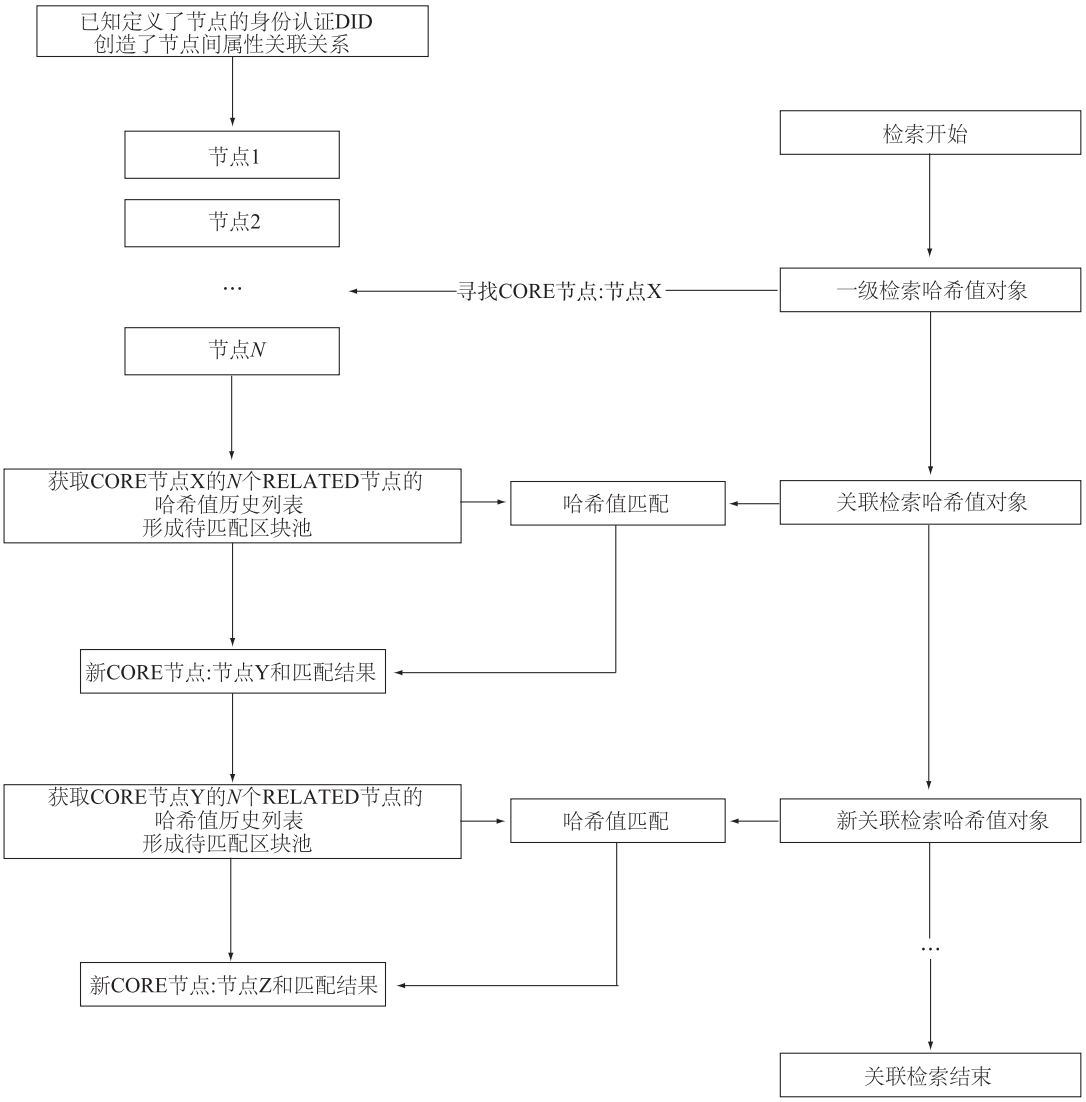

图 2 多级关联检索技术原理示意图

Fig.2 Schematic diagram of multi-level association retrieval technology

②每当完整地完成一次有最终匹配验证结果的溯源查询,就实现溯源码读码统计累加值刷新,且该溯源码的被读码记录将实时更新。

③每当完整地完成一次有最终匹配验证结果的溯源查询,就创建一条追溯查询记录,并向溯源监管平台推送这条数据。这里所说的溯源查询记录,是指用户完成一次溯源查询所得数据记录,该记录至少包括查询时间、查询者 IP、所查询的多级对象及其对应的相关溯源信息等关键部分。

基于上述节点关联度实现的多级关联检索技术,相比于常规仅通过区块链哈希匹配进行检索或基于关系型数据库进行检索的方案,具有较明显的优势,具体特点分析与优劣度对比如表 2 所示。

3 系统应用实施方案与结果分析

3.1 实施方案举例说明

假设现有农业投入品公司 A、农产品生产经营公司 B、监管单位 C 3 家主体通过本系统进行管理,具体实施方案如下:

首先,3 家主体向联盟链平台提交入链申请,提供包括主体基本信息和主体属性(如原材料投入品供应商、农产品生产企业、监管单位)等在内的信息。系统以自动和人工 2 种方式对上述信息进行审核,一旦审核通过,联盟链平台将自动为相关主体创建账户并激活,并向其颁发唯一数字证书,同时设定其在链内的治理权重等级,授权相应的功能权限。

表 2 基于节点关联度的哈希匹配检索和普通多级验证检索对比

Table 2 Comparison between hash matching retrieval based on node relevance and ordinary multi-level verification retrieval

项目	基于节点关联度的哈希匹配检索	普通区块链哈希匹配检索（公链或联盟链）	基于数据库查询的普通验证检索（非区块链）
多级验证检索步骤	(1)输入待验证的第一级溯源信息的哈希值;(2)遍历区块,在目标区块验证哈希值,返回验证结果;(3)找到目标区块对应的节点为 CORE 节点;(4)基于 CORE 节点属性,得到 RELATED 节点,并基于时序得到历史溯源数据记档表和待匹配区块池;(5)在历史溯源数据记档表中查看关联溯源信息或直接输入待验证的第二级溯源信息的哈希值;(6)在待匹配区块池中验证哈希值,返回验证结果;(7)循环或结束。	(1)输入待验证的第一级溯源信息的哈希值;(2)遍历区块,在目标区块验证哈希值,返回验证结果;(3)在第一级溯源信息中定位需要进行验证的第二级溯源信息关键词、时间范围等;(4)在应用系统数据库中基于查询条件遍历检索,匹配返回已上链的第二级溯源信息;(5)输入待验证的第二级溯源信息的哈希值;(6)遍历区块,在目标区块验证哈希值,返回验证结果;(7)循环或结束。	(1)输入待验证的第一级溯源信息;(2)解析该溯源信息的日期、关键词、操作方等查询条件;(3)在应用系统数据库中基于查询条件遍历检索,匹配最符合的查询结果;(4)返回匹配查询结果,对第一级溯源信息作出验证;(5)从返回的查询结果中定位需要进行验证的第二级溯源信息;(6)解析该溯源信息的日期、关键词、操作方等查询条件;(7)在应用系统数据库中基于查询条件遍历检索,匹配最符合的查询结果,返回匹配查询结果,对第二级溯源信息作出验证;(8)循环或结束。
优点	(1)第一次匹配检索后即可缩小后续几级关联检索的搜索范围;(2)保证溯源数据的现实关联度可被证实;保证溯源数据的真实性和不可篡改性。	保证溯源数据的真实性和不可篡改性。	(1)开发难度低,不需要区块链技术知识;(2)数据库内实际业务数据记录证明现实关联度。
缺点		(1)较难证实溯源数据现实关联度;(2)每一次检索均需遍历所有搜索范围,如在公链上验证检索,耗时会更长。	(1)无法保证溯源数据的真实性和不可篡改性;(2)每一次检索均需遍历所有搜索范围。

日常生产经营过程中,公司 A 和 B 利用企业应用端,分别录入各自的企业内部管理数据并申请上链。申请上链的数据须具备以下记号标签:数据传输头节点属性、尾节点属性、数据类型、数据摘要、数据采集方、数据采集时间。联盟链基于共识算法生成上链数据访问请求,并将其通过哈希加密算法加密为非明文字符串类型数据。

假设公司 A 的产品被公司 B 作为生产投入品原料,那么系统将自动对该产品进行链上标记,使公司 A 与公司 B 都拥有该原料的数据查询权,但具体查询范围和深度由数据所有者公司 A 进行设置。原料一旦进入公司 B 进行投入生产,则后续产生的相关数据都将归公司 B 所有。

由于公司 A 和公司 B 的相关数据均已上链,不可篡改,监管单位 C 在对公司 A 或公司 B 进行监管时,可在链上提交针对被监管主体相关数据进行查询的请求申请。系统会对监管单位和被监管主体进行身份核实,校验通过后将从链上摘取可供监管单位 C 访问的公司 A 或公司 B 具体业务数据,并自动生成相应的报表或溯源信息,通过监管查询专用的 API 返回。

公司 B 的农产品在市场销售过程中,消费者也可以通过本系统查询产品的相关信息。该功能由本联盟链的公共查询 API 提供,入口是该产品的可追溯二维码,消费者只需通过微信扫码即可访问。消费者在查询产品信息过程中,若希望进一步查看相关的投入品或原料信息,则需要由公司 A 和公司 B 在系统中分别对原料数据进行授权(该授权操作可在产品销售流通前完成)。其中公司 B 可对原料的采购、物流等数据进行授权,公司 A 可对原料的生产、加工、销售等数据进行授权。通过授权的数据将由系统进行组装后通过 API 返回给消费者,实现产品的深度追溯。

每当系统处理完一次有最终匹配验证结果的溯源查询请求,都会对该溯源码进行查询档案的记录并实时统计更新。公司 A、B 能够在企业应用端分别获知各自产品溯源码查询的综合情况分析。监管单位 C 和消费者可分别在溯源监管平台和溯源消费者端查看各自用户授权范围内的溯源查询记录数据。

3.2 实际应用结果分析

基于上述实施方案,本系统已在江阴市省级现代农业产业园、江苏省苏农科技转移中心有限公司等相关主体中进行了应用。现分别从系统性能指标、系统功能应用情况两方面进行总结与分析。

3.2.1 *系统性能指标* 在应用过程中,本系统体现出良好的稳定性、可靠性和用户体验,能够较好地满足农业生产管理场景的应用(表 3)。

3.2.2 *系统功能应用情况* 以江阴市省级现代农业园区的实际应用为例,其生产经营主体为华西都市农业有限公司,通过应用本系统,该公司进货贮存管理、田块管理、农产品质量安全过程管理、物流销售管理等核心业务全部实现了数字化转型,有效提

高了企业的生产管理效率。公司在不增加管理人员的情况下,生产面积较应用系统前翻了 1 倍,产量由 6 375 kg/hm^{2}提升到7 500 kg/hm^2,平均管理成本由 1 hm^2 6 000元降低到 1 hm^2 4 725元,公司经济效益得到了显著提升。尤其针对华西系列品牌大米产品,公司通过本系统实现了稻米全生命周期管理与追溯的从无到有,且相对市场上其他常见的追溯系统,数据可信度更高,品牌科技含量更高,因此华西稻米产品的市场认可度也得到了一定的提升。

表 3 区块链系统性能测试结果

Table 3 Results of blockchain system performance test

主要功能接口	请求次数	最小响应时间 (ms)	最大响应时间 (ms)	平均响应时间 (ms)	成功率 (%)
数据上链	50	15	22	16.7	100
链上数据查询	50	13	21	15.6	100
多级关联检索查询	50	91	121	105.7	100

4 结 语

综上所述,本文所述的基于联盟链和相关区块链技术的农产品深度溯源系统,核心是通过基于节点关联度的哈希匹配检索技术来实现对追溯信息的检索,并且能够在链上实现加密共享、高速检索追溯数据。该系统能普遍适用于现代智慧农业信息化管理体系、农产品全供应链大数据共享与监管治理体系[10]。其实际应用成本不高,维护管理方式简易,且有显著的功能性优势与开放性的拓展开发前景。相较于目前常见溯源技术和系统,本系统主要优势总结如下:①联盟链上读、写、存的任何机构的重要业务数据都是可信加密、高度安全的,譬如:产品成交价、合同编号、交易者详细信息等。这些数据仅限在联盟链内成员之间共享,不易被他人通过溯源监管平台或消费者查询端发起的网络安全攻击等而盗得重要数据。②监管单位和消费者能够利用溯源码查询,轻松获取到公开的农产品溯源数据信息,并实现高效的多级关联追溯查询。而现有的其他溯源技术,还不能够达到具有智能化多级深度溯源的能力,仅仅只能实现在溯源信息页面之间逐层跳转的手动操作。③因溯源信息上链存证而具有权威公信力,一旦发生质量安全事件,可快速追溯定位问题环节,该环节无法篡改历史数据,只能接受惩处并进行整改。该系统的应用可以有效提升区域范围内农产品安全合规生产水平,形成诚信经营的社会风气。

同时我们也意识到,未来可继续在以下技术方面升级与优化该系统[11]:①进一步优化数据上链记号标签的数据结构和响应检索算法,持续提高单次溯源信息检索的速度和准确率,不断升级多级关联追溯查询的综合效率。②溯源数据获取终端逐步升级过渡到物联网设备,以自动获取数据上链逐渐代替手工录入数据,增加溯源信息的真实性和可信度。③进一步提升联盟链共识机制和节点治理权限,提高节点上传溯源数据的频率。

参考文献:

[1] 张瑞星. 基于区块链技术的食品溯源平台关键技术研究[D]. 成都:电子科技大学,2021.

[2] 姚 前,张大伟.区块链系统中身份管理技术研究综述[J].软件学报, 2021,32(7):2260-2286.

[3] 赵锦波. 面向区块链数据隐私保护的可搜索加密研究[D]. 西安:西安电子科技大学,2019.

[4] 田 阳,陈智罡,宋新霞,等.区块链在供应链管理中的应用综述[J].计算机工程与应用,2021,57(19):70-83.

[5] RONAGHI M H. A blockchain maturity model in agricultural supply chain[J]. Information Processing in Agriculture, 2020,8:398-408.

[6] 沈志宏,黎建辉,张晓林. 关联数据互联技术研究综述:应用、方法与框架[J].图书情报工作,2013,57(14):125-133.

[7] ZHI H W. A log-structured file system based on levelDB[J]. Applied Mechanics and Materials,2014(602-605):3481-3484.

[8] 江苏省农业科学院. 一种基于数据关联模型的产品深度溯源方法:202010829762.8[P]. 2020-08-18.

[9] 江苏省农业科学院. 一种基于区块链技术的产品深度溯源系统:201911163786.8 [P].2019-11-23.

[10] FENG H, WANG X, DUAN Y, et al. Applying blockchain technology to improve agri-food traceability: A review of development methods, benefits and challenges[J]. Journal of Cleaner Production, 2020, 260:121031.

[11] 赵 奕,施鹏飞. 基于频繁集的多层次交互式关联规则挖掘[J].微电子学与计算机,2000(3):54-58.

(责任编辑:张震林)