DOI:10.13718/j.cnki.xsxb.2021.07.018

大数据环境下的安全多维数据聚合协议①

邹劲松[1], 李 芳[2]

1. 重庆水利电力职业技术学院 普天大数据产业学院, 重庆 402160;
2. 重庆大学 计算机学院, 重庆 400044

摘要: 针对大数据无线传感器网络(Wireless Sensor Networks, WSN)对路由聚合的同时难以保持数据机密性和完整性的问题进行研究, 提出一种新的无线传感器网络隐式多维数据聚合协议, 用于大数据环境下安全地聚合 WSN 生成的大量多维数据. 该协议通过采用一种简单、轻量级的超递增序列和加性同态加密方案构造多维数据以保护数据的机密性, 采用同态签名方案来保护数据的完整性, 通过检查中间节点的完整性过滤虚假数据包实现数据的新鲜度. 安全性分析显示该协议实现了端到端的安全性, 实验结果表明: 该协议降低了通信开销和能耗, 提高了传感器网络的生存周期.

关 键 词: 隐式多维数据聚合; 无线传感网络; 大数据; 加性同态加密; 数字签名

中图分类号: TP393 **文献标志码:** A **文章编号:** 1000-5471(2021)07-0125-05

无线传感器网络(WSN)是当前世界上备受关注、技术高度集成、众多学科交叉的前沿和热点研究领域之一, 它是由大量微型传感器节点组成的分布式传感网络[1-2], 在安全监控、事件跟踪和目标检测等与监控相关的应用中起着至关重要的作用[3]. 传感器节点由一组传感器组成以检测不同类型的数据, 且由此生成的数据成为多维数据. 部署在众多应用中的 WSN 产生了大量的大数据, 大数据由于数据复杂性、异构性、安全性、可伸缩性和大规模数据量导致复杂的数据分析[4-5].

数据聚合是大数据 WSN 中最实用、最重要的方式[6], 它将相似的数据组合起来并消除数据冗余, 减少了数据传输, 从而增加了网络的生命周期[7-8]. 但这些协议会影响准确性、容错性、延迟和安全性等性能指标. 因此, 需要以安全的方式执行数据聚合为目的的安全数据聚合协议. 安全数据聚合协议分为两类: 逐跳安全数据聚合、端到端或隐藏数据聚合. 在逐跳安全数据聚合中, 加密数据执行聚合之前在中间节点解密, 这增加了中间节点传感器的敏感度, 需要保护聚合器节点上数据的隐私[9]. 在端到端或隐藏的数据聚合中, 加密后的数据只能在基站解密. 考虑到数据的多维性、高收集频率和节点数量, 现有的安全数据聚合方案不仅带来了较高的通信成本, 而且会给聚合节点带来巨大的负载[10].

为了节省传感器网络的计算和通信资源, 文献[11]提出一种有效的数据集隐私保护聚合方案, 在某些应用程序中该方案聚合器可以验证加密消息以检测数据污染攻击, 而无需访问消息以保护隐私. 可采用完全解密和部分解密两种类型保证安全性. 文献[12]提出一种具有多汇无线传感器网络中的高效数据聚合算法, 该算法针对多汇无线传感器网络提出了两种数据聚合算法: 基于最小生成树算法和基于最短路径树算法, 使数据收集过程中的数据包传输次数最小化, 该算法解决了少量数据包传输的数据聚合问题. 文献[13]提出一种基于访问控制和认证的无线传感器网络安全数据聚合协议. 该协议由安全数据碎片算法和节点加入授权算法两个算法组成, 前者通过将数据分割成小块来对数据进行隐藏, 后者对节点证书进行认证, 以提

① 收稿日期: 2020-07-13
基金项目: 重庆市教育科学"十三五"规划 2020 年度重点课题(2020-GX-169); 重庆市职业教育学会 2020—2021 年度职业教育科研课题(2020ZJXH282086).
作者简介: 邹劲松, 硕士, 讲师, 主要从事计算机软件与理论研究.

高服务质量参数. 设计访问控制方案，通过减少通信开销和保证通信真实性的过程来支持准确性、能量效率、新鲜度和认证. 该协议针对坑攻击和 Sybil 攻击，不可扩展.

在研究了上述 WSN 中数据聚合算法的基础上，为了进一步减少通信开销，本文提出一种新的无线传感器网络隐式多维数据聚合协议，用于大数据环境下安全聚合 WSN 生成的大量多维数据. 该协议分别采用隐性同态加密方案和数字签名方案来保护聚合数据的机密性和完整性，通过检查中间节点的完整性来过滤虚假数据包，确保消息的身份验证和数据的新鲜性. 该协议具有重量轻、可扩展性强、抗各种攻击能力强等特点. 安全性分析和实验结果表明，本文协议实现了端到端的安全性，减少了通信开销和能量消耗，提高了传感器网络的生存期.

1 大数据环境下的 WSN 多维数据聚合

本文提出的多维数据聚合协议使用数字签名方案保护数据的完整性，使用超递增序列同态加密实现数据机密性，通过果粒橙虚假数据包来确保数据新鲜度.

1.1 网络设置

假设 WSN 由单个基站和大量微型传感器节点组成，每个节点由一组测量不同现象的传感器组成，因此在 WSN 中生成的数据本质上是多维的. 传感器节点体积小，且资源有限. 本文假设网络拓扑是基于树的网络拓扑，并且节点平稳，为了支持安全的数据聚合，基站配备了更多的内存、带宽、处理器和能量. 根据微聚合服务协议构造了聚合树(平衡 k 元树)，聚合树中的节点有 3 种类型：源(叶)节点、聚合器(中间)节点和基站. 源节点生成要聚合的数据并将其转发给聚合器，聚合器节点聚合来自其子节点的数据，并将聚合结果转发到聚合器或基站，基站被指定为聚合树的根. 一旦构建了聚合树，基站使用改进的定时、高效、流式、容错(μTESLA)广播认证协议将查询传播到网络. 多跳通信模型用于将数据从源节点或聚合节点传输到基站.

设每个节点都有唯一的身份标准号(ID)和一组传感器，所有的传感器节点都是同质且静态的. 基站配备了一个功能强大且值得信赖的防篡改设备，有足够的内存和电源来保证通信安全，链接对称. WSN 是同步的，即所有源、聚合器基站(BS)的时钟值均同步. 叶节点仅感知数据，聚合器仅聚合数据，遵循周期性简单聚合.

在大数据 WSN 中，一个主要问题是在进行路由数据聚合的同时实现端到端的安全. 在数据聚合过程中，攻击者可以窃听传输数据侵犯数据隐私，发起攻击破坏数据完整性或将伪造的数据注入网络使聚合结果无效. 因此，拟议的协议必须达到以下安全要求：保密性以确保信息不会透露给任何非授权接收者；完整性以确保数据在传输过程中不被更改；认证以确保消息的正确来源，并检查中间节点的数据完整性；隐私性以确保每个传感器节点数据仅为其自身所知.

1.2 采用数字签名保护数据完整性

在无线传感器网络中，必须证明传感信息的准确性，以便基站能够做出正确的决策. 在公钥密码学中，签名是一种信息，可确保传感器节点之间信息的完整性和真实性，而无需共享任何秘密信息. 签名是由发送者的私钥创建的，接收者可以使用发送者的公钥进行验证.

1.3 同态加密确保数据机密性

同态加密可以对加密数据进行直接计算，又称为“隐性同态”. 在无线传感器网络中，隐性同态加密方案聚合加密数据的方式与聚合原始数据的方式相同，同态加密保证了数据的安全聚合.

1.4 大数据环境下的多维数据聚合协议

本文提出的协议将简单加性同态加密与基于身份的签名方案相结合，实现了端到端的安全数据聚合. 它保护了聚合结果的机密性和完整性，并增强了主动和被动攻击的安全性.

2 安全性分析

本节分析所提出协议的安全性，根据 1.1 节的安全要求，重点分析该协议如何实现数据机密性、数据完整性和消息认证.

数据机密性：数据机密性由攻击者无法在不损害源节点的情况下窃听传输数据和对手不能透露所传输数据的机密性两类组成. 在本文协议中，由源节点感测的多维数据被表示为 $d_i=(d_{i1},d_{i2},\cdots,d_{il})$，并被加密为

$$C_i=(a_1d_{i1}+a_2d_{i2}+\cdots+a_ld_{il})\bmod M+k_i \tag{13}$$

令 $m_i=(a_1d_{i1}+a_2d_{i2}+\cdots+a_ld_{il})$，那么 $C_i=m_i\bmod M+k_i$ 仍然是加同态密码系统的有效密文. 由于可加性同态密码系统在语义上对选择明文攻击是安全的，因此其数据在语义上也安全并能保护隐私. 如果攻击者偷听密文，将无法识别相应的消息. 用于加密的密钥仅为相应的源节点和基站所知，源节点用于加密的密钥是唯一的，基站在部署网络之前将其分发给源节点，攻击者无法在不损害源节点的情况下窃听传输数据. 因此，该协议在源节点上保留了数据隐私.

在本文协议中，源节点感知到的观测数据使用一种语义安全加密方案进行加密，该加密方案在不同时间为同一消息产生不同的密文. 由于明文仅为源节点和基站所知的密钥加密，因此对手不能推断出明文，也不能通过比较密文来获得任何重要信息. 假设攻击者破坏了聚合器节点，也无法推断单个消息，因为它们是经过加密和聚合的，而且由于聚合密文附加了一个难以破解的签名，攻击者很难访问聚合器传输数据. 因此，本文方案中实现了数据机密性.

数据完整性和身份验证：在本文协议中，密文和聚合密文在传送到它们对应的父节点之前都附加了签名. 该协议使用单向哈希函数生成的签名在椭圆曲线密码体制下具有可证明的安全性，这些签名由基站的主密钥和节点的私钥生成，这两个密钥很难推导. 秘密随着时间推移而演变，攻击者无法推断出秘密并破坏系统. 此外，密钥不会共享给其他节点，造成签名生成困难，使得对手伪造签名的任务变得不可行. 通过验证签名，聚合器在密文/聚合密文之前检查其真实性，即使对手泄露了一个节点，也无法解密密文，因为本文协议使用不同的密钥进行解密. 因此，本文协议保证了数据完整性和身份验证.

虚假数据攻击：在本文协议中，每个源节点生成一个签名，并在其转发到中间节点之前将其附加到密文中. 每个中间节点检查签名的时间戳，如果签名有效，则验证签名. 如果签名成功传递，则密文/聚合密文将被聚合，否则将确认数据包来自对手，从而丢弃该数据包. 通过删除密文，本文方法可以防止虚假数据注入攻击.

3　实验结果与分析

在大数据 WSN 中，节能是多维数据聚合需要解决的关键问题. 本节将根据计算复杂性、通信开销和能源消耗评估性能. 将本文协议与文献[14]提出的多维隐私保护聚合(MDPA)协议、文献[15]提出的提供精确查询的安全网络内处理(SIES)协议进行分析比对.

所有实验均在一台配置为 Intel Core i7 CPU @3.60 GHz 和 8 GB RAM 的机器上进行，所有测试均在 Matlab 2014a 环境下实现. 实验设定在 200 m×200 m 的区域内随机分布 2 500 个普通传感器节点和 150 个传输半径为 20 m 的汇聚节点. 本节将 F_p 中元素的长度预先考虑为 160 位、1 024 位和 160 位. 此外，假设单向散列函数的长度为 160 位，而标识和时间戳的长度为 32 位.

3.1　通信开销

在无线传感器网络中，源节点首先将数据传输到聚合器节点，聚合器节点再将数据汇聚并传输到基站进一步处理和分析. 因此，可以基于源到聚合器之间以及聚合器到聚合器 / 基站之间的通信来计算通信成本. 令有限域 G,G_T 和 F_p^* 中元素的长度分别为 160 位、1 024 位和 160 位，单向哈希函数的长度为 160 位，身份和时间戳的长度为 32 位.

在本文协议中，源节点以$\{C_i, ID_i, ts, T_i, z_i\}$的形式生成密文和签名，并将其发送到聚合器节点，其中 C_i 为节点 i 的密文，$C_i\in F_p^*$，ID_i 为节点身份标识号，ts 为时间戳，T_i 为公钥，$T_i\in G$，z_i 为使用单向哈希生成的签名，$z_i\in F_p^*$. 因此，源到聚合器的通信成本计算为 $|C_i|+|ID_i|+|ts|+|T_i|+|z_i|=544$ 位. 在 MDPA 协议中，$\{ID_i, ts, C_i, \sigma_i\}$ 从源发送到聚合器，其中 σ_i 为节点 i 的签名，$C_i\in G_T$，$\sigma_i\in G$，通信成本为 $|ID_i|+|ts|+|C_i|+|\sigma_i|=1\,248$ 位.

在本文协议中，聚合器与聚合器 / 基站之间的通信是将$\{C_j, ID_j, ts, T_j, z_j\}$发送到聚合器 / 基站，因

此通信成本为 $|C_j|+|ID_j|+|ts|+|T_j|+|z_j|=544$ 位. 在MDPA协议中,$\{ID_j, ts, C_j, \sigma_j\}$ 从聚合器发送到聚合器/基站,其中 $C_j \in G_T$, $\sigma_j \in G$,通信成本为 $|ID_i|+|ts|+|C_i|+|\sigma_i|=1\ 248$ 位.

与MDPA协议相比,本文协议通信成本更低,这是因为本文使用的加密和签名方案是轻量级的,而在MPDA中生成的密文非常大.

3.2 复杂性分析

为了评估计算复杂度,将一个乘法运算的代价表示为 M,将一个加法运算的代价表示为 A,将一个哈希运算的代价表示为 H,用于聚合的维度数表示为 l. 在本文协议中,源节点计算密文和签名. 要计算密文,需要 $lM+A$ 操作;要生成签名,需要 $M+H$ 操作. 在源节点总共需要 $(l+1)M+H+A$ 个操作. 在聚合器节点,本文协议需要 $(x+2)M+(l-1)A$ 操作来聚合来自 x 子节点的密文/聚合密文,并且需要 $(M+H)$ 操作来签名. 因此,在聚合器处,总共需要 $(x+3)M+H+(l-1)$ 操作.

3.3 能量消耗

能耗是资源受限无线传感器网络的核心问题,能耗量直接影响无线传感器网络的寿命. 本文针对Mica2dot和TelosB传感器平台计算所提协议需要的能量. 计算和通信是影响无线传感器网络能耗的两个因素. 为了分析协议的能耗,在数据速率分别为12.4 kb/s和75 kb/s时,将Mica2dot和TelosB的能耗用于不同的操作. 由于执行加法运算的能量成本非常低,本文只考虑乘法运算. 设子树的子节点数(x)为8,维数(l)为3. 节点标识符和时间戳的大小为8字节.

与SIES协议和MDPA协议相比,本文协议由于节点中所需的模乘运算较少,所需能耗更低. 这是因为本文协议通过检查中间节点的完整性过滤虚假数据包,实现了数据新鲜度.

4 结 语

为了在大数据环境下安全聚合WSN生成的大量多维数据,本文提出一种新的无线传感器网络隐式多维数据聚合协议. 该协议具有重量轻、可扩展性强、抗各种攻击能力强等特点,采用数字签名方案来保护数据的完整性,使用一种简单、轻量级的超递增序列加性同态加密来实现数据的保密性,实现端到端的安全数据聚合,增强了主动和被动攻击的安全性. 此外,通过检查中间节点的完整性过滤虚假数据包,保证消息的身份验证和数据的新鲜度. 实验结果表明,该协议在能量消耗方面有明显改善,提高了传感器网络的生存期. 这是多维数据聚合的第一项工作,提供了端到端的安全性,实现了隐私性、完整性、高效性之间的优化平衡,但是其主要关注感知数据本身的隐私性,未来的工作是将研究扩展到基于上下文的隐私保护数据聚合,更好地保护监测对象的安全.

参考文献:

[1] 孙 倩,陈 昊,李 超. 基于改进人工蜂群算法与MapReduce的大数据聚类算法[J]. 计算机应用研究,2020,37(6):1707-1710,1764.

[2] MODIEGINYANE K M, LETSWAMOTSE B B, MALEKIAN R, et al. Software Defined Wireless Sensor Networks Application Opportunities for Efficient Network Management: a Survey [J]. Computers & Electrical Engineering, 2018, 66: 274-287.

[3] ELSAYED W, ELHOSENY M, SABBEH S, et al. Self-Maintenance Model for Wireless Sensor Networks [J]. Computers & Electrical Engineering, 2018, 70: 799-812.

[4] OUSSOUS A, BENJELLOUN F Z, AIT LAHCEN A, et al. Big Data Technologies: a Survey [J]. Journal of King Saud University-Computer and Information Sciences, 2018, 30(4): 431-448.

[5] DAI H N, WONG R C W, WANG H, et al. Big Data Analytics for Large-Scale Wireless Networks [J]. ACM Computing Surveys, 2019, 52(5): 1-36.

[6] 陈 琪,陈宏滨. 无线传感器网络中移动节点辅助的数据采集效率优化研究[J]. 计算机应用研究,2020,37(11):3467-3471.

[7] TONYALI S, AKKAYA K, SAPUTRO N, et al. Privacy-Preserving Protocols for Secure and Reliable Data Aggregation in IoT-Enabled Smart Metering Systems [J]. Future Generation Computer Systems, 2018, 78: 547-557.

[8] LI X, LIU S P, WU F, et al. Privacy Preserving Data Aggregation Scheme for Mobile Edge Computing Assisted IoT Applications [J]. IEEE Internet of Things Journal, 2019, 6(3): 4755-4763.

[9] SHEN X D, ZHU L H, XU C, et al. A Privacy-Preserving Data Aggregation Scheme for Dynamic Groups in Fog Computing [J]. Information Sciences, 2020, 514: 118-130.

[10] ZHONG H, SHAO L L, CUI J, et al. An Efficient and Secure Recoverable Data Aggregation Scheme for Heterogeneous Wireless Sensor Networks [J]. Journal of Parallel and Distributed Computing, 2018, 111: 1-12.

[11] SHERIFL A, ALSHARIF A, MAHMOUD M, et al. Efficient Privacy-Preserving Aggregation Scheme for Data Sets [C]// 2018 25th International Conference on Telecommunications (ICT). Saint-Malo: IEEE, 2018.

[12] YESTEMIROVA G, SAGINBEKOV S. Efficient Data Aggregation in Wireless Sensor Networks with Multiple Sinks [C]// 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA). Krakow: IEEE, 2018.

[13] RAZAQUE A, RIZVI S S. Secure Data Aggregation Using Access Control and Authentication for Wireless Sensor Networks [J]. Computers & Security, 2017, 70: 532-545.

[14] 刘振鹏，董亚伟，赵　璇，等. MDPA：基于 MCL 的社会网络差分隐私数据发布算法 [J]. 郑州大学学报(理学版)，2018，50(1)：1-7.

[15] ZHOU Q, YANG G, HE L. A Secure-Enhanced Data Aggregation Based on ECC in Wireless Sensor Networks [J]. Sensors (Basel), 2014, 14(4): 6701-6721.

Secure Multidimensional Data Aggregation Protocol in Big Data Environment

ZOU Jin-song[1], LI Fang[2]

1. *Putian Big Data Industry School, Chongqing College of Water Resources & Electric Engineering, Chongqing 402160, China;*

2. *School of Computer Science, Chongqing University, Chongqing 400044, China*

Abstract: Aiming at the problem that it is difficult to maintain data confidentiality and integrity while routing aggregation in Wireless Sensor Networks (WSN), in this paper, a new implicit multi-dimensional data aggregation protocol has been proposed for wireless sensor networks, which can safely aggregate a large number of multidimensional data generated by WSN in big data environment. The protocol uses a simple, lightweight super increasing sequence and additive homomorphic encryption scheme to construct multi-dimensional data to protect the confidentiality of the data, uses homomorphic signature scheme to protect the integrity of the data, and filters the false data packets by checking the integrity of the intermediate nodes to achieve the freshness of the data. Security analysis shows that the protocol achieves end-to-end security. Experimental results show that the protocol reduces communication overhead and energy consumption, and improves the life cycle of sensor networks.

Key words: concealed multidimensional data aggregation; wireless sensor networks; big data; additive homomorphic encryption; digital signature

责任编辑　夏　娟