面向双层传感网的隐私保护 k-NN 查询处理协议

彭 辉 陈 红 张晓莹 曾菊儒 吴云乘 王 珊

(中国人民大学数据工程与知识工程教育部重点实验室 北京 100872)

(中国人民大学信息学院 北京 100872)

摘 要 无线传感器网络作为物联网感知层的核心组成部分,具有广阔的应用前景.然而,隐私泄露问题严重阻碍了传感器网络的发展.目前,传感器网络隐私保护技术已成为研究热点,其中隐私保护 k-NN(k-Nearest Neighbor)查询协议是富有挑战性的问题.文中提出了面向双层传感器网络的高效的隐私保护 k-NN 查询协议.首先,为提升查询效率,基于定向存储策略给出了适用于双层传感网的 k-NN 查询架构.其次,针对管理节点俘获攻击,提出了一种新颖的隐私保护数据编码机制,通过为真实数据附加编码的方式,保证在不泄露数据隐私的同时精确地完成查询处理.再次,针对节点共谋攻击,设计了基于代理节点的单向数据隐藏机制,通过破坏普通节点与管理节点间数据的关联性实现抵御共谋攻击的目标.理论分析和仿真实验验证了协议的安全性和有效性.

关键词 物联网;无线传感器网络;隐私保护;k-NN 查询;节点俘获;共谋攻击

中图法分类号 TP301 **DOI 号** 10.11897/SP.J.1016.2016.00872

Privacy-Preserving k-NN Query Protocol for Two-Tiered Wireless Sensor Networks

PENG Hui CHEN Hong ZHANG Xiao-Ying ZENG Ju-Ru WU Yun-Cheng WANG Shan

(*Key Laboratory of Data Engineering and Knowledge Engineering of Ministry of Education, Beijing* 100872)

(*School of Information, Renmin University of China, Beijing* 100872)

Abstract As an important part of Internet of Things (IoTs), Wireless Sensor Networks (WSNs) can be widely deployed in many physical environments for data collection and monitoring, such as oceans, volcanoes, animal habitats and battlefields. In these applications of WSNs, k Nearest Neighbors (k-NN) query which returns the nearest k objects of the given query location or value is very important for users to get the required information of monitoring areas. In consideration of the built-in resource limitation problems of single tier sensor networks, k-NN query can be more efficient in two tiered WSNs which contain a large number of resource constrained sensor nodes and fewer resource rich management nodes. However, relying on management nodes for data storage and query processing raises significant security concerns. Frequently, the deployment environments of WSNs are hostile and unpredictable, which threaten the data security and privacy. What is worse, the open nature of wireless communication makes it easy for adversaries to eavesdrop data packets or compromise nodes of sensor networks. Management nodes are more vulnerable to be

收稿日期:2015-02-09;在线出版日期:2015-12-08.本课题得到国家"九七三"重点基础研究发展规划项目基金(2014CB340402,2012CB316205)、国家"八六三"高技术研究发展计划项目基金(2014AA015204)、国家自然科学基金(61532021,61272137,61202114)资助.**彭 辉**,男,1986 年生,博士研究生,主要研究方向为无线传感器网络、隐私保护、物联网数据管理.E-mail:huipeng@ruc.edu.cn.**陈 红**(通信作者),女,1965 年生,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为数据库、数据仓库、无线传感器网络.E-mail:chong@ruc.edu.cn.**张晓莹**,女,1987 年生,博士,主要研究方向为无线传感器网络、隐私保护.**曾菊儒**,男,1987 年生,博士研究生,主要研究方向为参与式感知、隐私保护.**吴云乘**,男,1989 年生,博士研究生,主要研究方向为物联网数据管理、隐私保护.**王 珊**,女,1944 年生,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为高性能数据库、知识工程、数据仓库、数据分析.

compromised as they store massive data and serve as an important role for responding queries. The compromised management nodes bring serious security risks to the network and significantly restrict the development of two tiered WSNs. Therefore, in recent years, privacy preservation in two tiered WSNs has attracted more and more attention. Accurately answer k-NN query while preserving data privacy is a challenging problem. This paper presents an efficient privacy-preserving k-NN query protocol for two tiered WSNs. To the best of our knowledge, this paper is the first to consider collusion attacks for a value-based k-NN query in two tiered WSNs while fulfilling the preservation of privacy. To improve query efficiency, the architecture of k-NN query is proposed based on the directional storage strategy for two tiered WSNs. To resist management nodes compromising attacks, this paper proposes a novel privacy-preserving data encoding mechanism to conceal sensitive information to response the accurate query result without data privacy leakage. Then, to resist nodes collusion attacks, we present a proxy-nodes-based one-way data hiding mechanism to disorder the data correlation between common sensor nodes and management nodes. In addition, a series of improve schemes are further introduced to enhance the performance of proposed privacy-preserving approaches, including group-based random storage mechanism, dynamic query mechanism, subfields partition mechanism and data code compression mechanism. Finally, theoretical analysis and experimental results of real dataset confirm the privacy and efficiency of our proposals.

Keywords　Internet of Things; wireless sensor networks; privacy-preservation; k-Nearest Neighbor query; node compromising; collusion attack

1　引　言

无线传感器网络[1](Wireless Sensor Network, WSN)作为物联网[2]感知层的核心组成部分,在国防军事、医疗卫生、智慧交通、环境监测、智能家居等领域有着广阔的应用前景.传感器网络可以在任何时间、地点和环境下获取大量重要的信息,为人们的科研、工作和生活提供极大的便利.然而,近年来,传感器网络在多个关键性领域的实际应用与部署过程中,也暴露出严重的隐私威胁.例如,在军事领域,传感器网络收集和查询的数据往往携带着重要情报,这些数据一旦泄漏很容易造成重大决策失误;在医疗应用领域,使用便携式无线传感器采集患者实时数据时,容易泄露患者的敏感体征数据;在车载传感网应用中,用户的位置和行动轨迹等隐私信息容易被泄露等.这些在传感器网络查询应用中暴露出的隐私问题会导致用户和监测对象的重要信息泄露,甚至威胁到用户和监测对象的安全,极大地阻碍了传感器网络的应用.研究和解决传感器网络中的隐私保护问题,对传感器网络的大规模应用具有重要意义.

目前,无线传感器网络中的隐私保护问题[3]已经成为研究的热点.现有隐私保护技术研究主要面向求和、极值、中位数等数据聚集[4-11]操作以及范围查询[12-21]和 Top-k 查询[22-26].对传感器网络中另一类重要的查询类型 k-NN(k-Nearest Neighbor)查询的隐私保护技术研究还很少.k-NN 查询要求网络返回距离查询点最近的前 k 个数据对象值,在传感器网络中有着广泛的应用.例如,在水污染监测网络中,可以通过查询距离标准污染指数最近的前 k 个监测数据值及其节点来确定具有相同污染水平的区域.根据查询点设置的不同,可以将现有的 k-NN 查询划分为两种类型:一是基于地理位置的 k-NN 查询,即给定一个地理位置作为查询点,查找距离该位置最近且满足一定条件的前 k 个节点;二是基于值的 k-NN 查询,即用户以一个设定的数值作为查询点,查找离给定值最近且满足一定条件的前 k 个节点.本文主要关注第 2 种基于值的 k-NN 查询.

由于单层传感器网络所固有的资源受限等问题,现有的数据查询处理研究大多采用双层传感器网络结构[27].如图 1 所示,在双层无线传感器网络

中,网络的下层由大量能量和计算能力有限的普通节点组成,网络的上层由少量资源充足的管理节点组成.管理节点的引入可以给网络带来明显的优势:(1)延长网络生命周期,普通节点只需要通过一跳或多跳的方式将数据传输到最近的管理节点,而不需要发送给更远的基站,明显地减少了数据包的传输跳数,降低了网络的通信能耗;(2)提高查询效率,管理节点拥有更强的计算和存储能力,能够更好地执行复杂的查询操作,有效地缩短查询响应时延;(3)增强可伸缩性,管理节点可以简化节点的加入/离开以及网络拓扑管理操作,对各种规模的应用场景都有较好的适用性.

然而,管理节点在给网络数据存储和查询处理带来便利的同时也导致了严重的隐私泄露风险.传感器网络无线通信的开放性以及部署环境的无人值守等特性使得攻击者可以较为容易地窃听通信链路或者直接俘获网络节点.管理节点由于存储着大量的感知数据,并且在查询处理过程中起着关键作用,因此是攻击者俘获的重点目标.俘获后的管理节点会给网络数据安全带来严重的影响.首先,由于管理节点是不可信的,所有感知数据在发送到管理节点之前都必须进行加密操作.其次,管理节点在无法获得数据真实值的前提下,必须正确高效地在加密数据上完成查询处理操作.再次,简单的端到端加密和逐跳加密都无法适用于管理节点被俘获的双层传感器网络.在端到端加密中,管理节点无法执行网内数据查询操作;在逐跳加密中,攻击者可以通过俘获管理节点来获得密钥,进而获得敏感数据信息.最后,攻击者可以同时俘获管理节点和普通节点,并通过两者的共谋攻击来获得密钥、数据组织格式、数据隐藏机制等关键安全信息,进一步破坏数据的隐私性.因此,针对攻击者的窃听、俘获和共谋攻击,研究既能够保证数据隐私安全又能够高效率低能耗地完成查询的隐私保护 k-NN 查询处理协议具有很强的必要性.

本文研究了双层传感网中的隐私保护 k-NN 查询处理问题,分别提出了面向俘获攻击的 PKN(Privacy-preserving k-NN Query)协议和面向共谋攻击的 CPKN(Collusion-resistant Privacy-preserving k-NN Query)协议.为提高查询执行效率,文中首先设计了基于定向存储机制的 KNQ(k-NN Query)查询框架,有效缩短查询响应时延;为保护数据隐私,PKN 协议引入了一种新颖的数据编码方法来分别对感知数据和查询请求进行编码,保证管理节点可以在不获知数据真实值的前提下完成 k-NN 数据查询操作;为抵御共谋攻击,CPKN 通过在普通节点和存储节点之间设置一组特殊的代理节点来对初始编码进行处理,从而破坏普通节点与管理节点的数据关联性,达到使共谋攻击失效的目的.

本文的主要贡献有:(1)提出双层传感器网络中基于 Bloom filter 编码的隐私保护 k-NN 查询协议,在不泄露数据隐私的情况下精确地完成 k-NN 查询;(2)针对俘获后的管理节点和普通节点之间较高的共谋风险,提出基于代理节点的编码隐藏机制,能够有效地抵御共谋攻击;(3)在不泄露数据隐私的前提下,设计了双层网络上基于定向存储机制的隐私保护查询架构,保证了查询的高效执行.

本文第 2 节介绍相关研究工作;第 3 节给出研究模型和问题定义;第 4 节和第 5 节分别描述面向俘获攻击的 PKN 协议和面向共谋攻击的 CPKN 协议;第 6 节通过真实数据集上的实验证明协议的有效性;第 7 节对全文进行总结.

2 相关工作

现有的传感器网络隐私保护技术研究大多集中于隐私保护数据聚集[4-11]、范围查询[12-21]、Top-k 查询[22-26]等,对隐私保护 k-NN 查询的研究还很少.同时,已有研究所针对的攻击类型多为节点俘获攻击,对于隐私危害性更大的共谋攻击的研究也较少.

现有隐私保护数据聚集协议主要研究求和聚集与极值聚集,所使用的隐私保护技术有数据扰动技术[4-6]、切分重组技术[4-5,7-8]和同态加密技术[9-11]等.CPDA[4-5]结合了数据扰动和多项式性质,在隐藏节点真实数据的前提下,通过多项式变换求解出精确的求和聚集结果.SMART[4-5]提出了基于切分重组机制的隐私保护求和聚集协议,节点将每个敏感数据切分成多个数据片,不同数据片发送给不同的邻居节点,节点将收到数据片重新组合来完成求和聚集.ESPART[7]对 SMART 进行了改进,通过引入多种优化因子,有效提高了网络效率.SESDA[8]采用基于路线的方式执行数据聚集,将数据片通过安全通道传输来保证数据的隐私性,由于不采用加密处理,保证了较高的通信性能和能耗性能.IPHCDA[10]根据传感器网络的特点,对基于椭圆曲线密码机制的同态加密技术进行了改进,提高了节点数据在求

和聚集时的安全性. 文献[11]在使用同态加密技术保护节点数据隐私性的同时，提出了一种基于离散时间控制环的动态数据聚集机制，以降低传感器节点的通信能耗.

现有隐私保护范围查询协议研究主要在双层传感器网络上展开，主要方法有桶模式[12-16]、数据编码技术[17-19]和保序加密技术[20-21]. 文献[12-13]采用桶模式保护数据的隐私性，首先将值域划分成多个连续的桶，然后通过将用户查询和节点数据分别表示成对应的桶号来完成查询，攻击者无法从桶号推断出节点准确的真实数据. 文献[16]给出了面向隐私保护单维和多维范围查询的 VQuery 协议，VQuery 基于多项式技术对数据范围和查询条件进行编码以隐藏真实敏感信息，基于水印链完成结果完整性认证，并提出了多维区间树结构以表示多维数据. SafeQ[17-18]使用前缀成员验证技术在数据编码上实现双层网络中的隐私保护范围查询. 在 ZOSR[19]中，感知数据和查询请求被分别转化为 Z-O 编码，管理节点利用 Z-O 编码的数值比较特性，在不需要感知数据明文的前提下隐私保护地完成查询操作. 文献[21]给出了一种新颖的保序加密机制，通过数据的有序映射，管理节点可以不泄露隐私地完成查询任务，同时还构建了基于多维邻居树和环形冗余校验码的链接水印机制以检验结果的完整性.

在隐私保护 Top-k 查询处理技术中，现有研究所使用的主要技术包括安全多方计算[22-23]、保序加密技术[24]、前缀成员验证技术[25-26]等. SafeTQ[22]为每一个管理节点搭配一个高资源的辅助节点，管理节点与辅助节点使用安全比较机制[28]求解出 Top-k 结果的阈值. PI-TQ[23]对 SafeTQ 进行了改进，消除了对辅助节点的依赖，算法通过两轮查询来得出 Top-k 结果. 文献[24]在保序加密机制上分别考虑了攻击者的密文攻击、背景知识推测攻击和恶意节点攻击，通过引入扰动机制和基于抽样假设检验的验证机制来保证数据的安全性和完整性. SVTQ[25]协议首先将节点数据通过前缀成员验证技术表示成前缀族和范围前缀的编码集合，再使用素数聚集机制完成数据大小关系判定，从而在保护隐私的同时完成查询. PPTQ[26]协议中，存储节点借助前缀成员验证技术的数值比较特性，在不获取原始数据的情况下计算包含查询结果的最小候选密文数据集，以实现隐私保护 Top-k 查询.

在传统数据库和云计算环境中，隐私保护 k-NN 查询研究已经取得了一些成果. 然而，这些研究成果往往基于高强度的加密操作[29-32]或复杂的分布式多方计算[33-34]来达到保护数据隐私的目的，既要求很强的计算能力又要求很强的网络通信能力，因此并不适用于能量以及计算能力、存储能力受限的无线传感器网络.

现有的传感器网络隐私保护技术研究对 k-NN 查询的关注很少，由于查询类型存在明显差异，以上这些方法无法直接应用在传感器网络隐私保护 k-NN 查询中. 同时，现有成果大多仅针对管理节点俘获攻击，而没有考虑对数据隐私危害更大的管理节点与普通传感器节点间的共谋攻击. 针对以上问题，本文分别给出了面向节点俘获攻击的 PKN 协议和面向节点共谋攻击的 CPKN 协议，以在双层传感器网络中实现高效的隐私保护 k-NN 查询处理.

3　研究模型与问题定义

3.1　研究模型

如图 1 所示，双层网络中包含 N 个能量、计算和存储都受限的普通传感器节点 s_i $(1\leq i\leq N)$以及 M 个与基站类似的资源充足的管理节点 $s_{\mathcal{M}}$. 传感器节点通过多跳方式与管理节点通信，而管理节点可以通过一跳方式直接将数据发送给传感器节点. 简明起见，我们假设网络通信状况良好，不存在路由阻塞、链路失效等问题. 传感器节点在每个上传周期中执行 1 次数据采集操作，采集到的数据包含多个属性，例如温度、光照、湿度等. 在任意周期 t 中，普通节点 $s_i(1\leq i\leq N)$采集到的数据表示为 $V_{s_i}=\{v_i[1], v_i[2],\cdots, v_i[d]\}$，$d$ 为数据的维数. 在周期结束时，

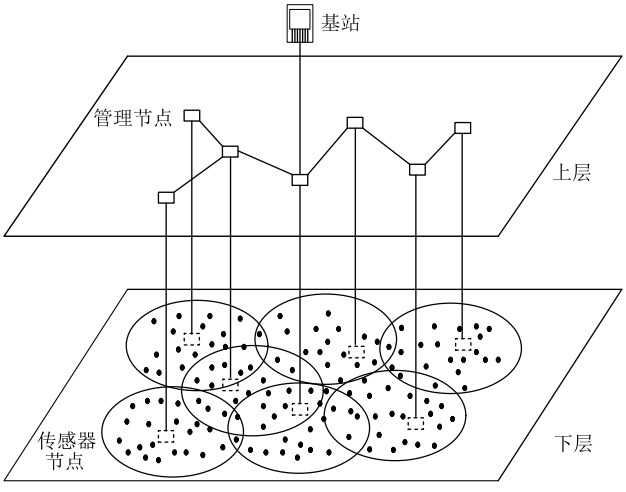


图 1　双层无线传感器网络示意图

普通节点通过多跳通信的形式将采集到的数据上传到最近的管理节点.数据上传结束后,基站直接与管理节点通信以完成 k-NN 查询.接下来,我们给出基于值的 k-NN 查询的相关定义.

定义 1. 值的距离函数.对于任意给定的值 G ($G=\{G[1],G[2],\cdots,G[d]\}$),$G$ 与节点 s_i 之间的值的距离 $Dis(G,s_i)$ 为

$$Dis(G,s_i)=\sqrt{\sum_{j=1}^{d}w_j(G[j]-v_i[j])^2},$$

其中 w_j 为第 j 维属性的权值,由用户按照应用的不同以及各属性值的值域和重要性进行指定.

定义 2. 基于值的 k-NN 查询.对于任意给定的包含 N 个对象的集合 S 以及任意给定的值 G,查找一个大小为 k 的子集 $S'(S'\subseteq S,k\leq N)$,使得对于 $\forall o_e\in S'$ 以及 $\forall o_f\in S-S'$,有 $Dis(G,o_e)\leq Dis(G,o_f)$.其中,$Dis$ 为值的距离函数.

以图 2 为例,当 $d=1$ 时,下发查询($k=3,G=5.0$),则查询结果为 $\{\langle s_1,5.1\rangle,\langle s_6,5.2\rangle,\langle s_8,4.7\rangle\}$.

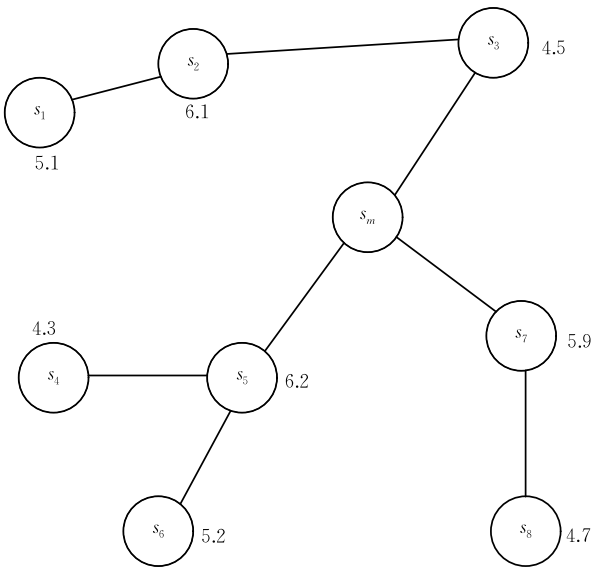


图 2 基于值的 k-NN 查询示意图

3.2 攻击模型

在双层传感器网络攻击模型中,对隐私数据的攻击可以分为外部攻击和内部攻击两种.外部攻击通过无线通信链路窃听的方式来获得敏感数据信息.内部攻击主要通过俘获单个或多个节点的方式实现,包括:

普通节点俘获攻击.在俘获单个普通节点后,攻击者可以获得该节点上的真实数据,对整体网络的危害较小.

管理节点俘获攻击.与普通节点相比,攻击者更倾向于俘获管理节点,一是因为管理节点存储着大量的感知数据;二是管理节点可以基于接收到的数据和查询请求进行推测攻击,以获得更多隐私数据信息.

多个节点共谋攻击.攻击者通过管理节点与普通节点之间的共谋来获得敏感数据信息.可以看出,普通节点间的共谋攻击只会泄露被俘获节点的数据信息,对整体网络的影响较小;而管理节点与普通节点的共谋攻击有很大可能会泄露存储在管理节点上的所有的数据信息,造成极大的危害.

本文的研究基于诚信但好奇(honest-but-curious)模型[35],攻击者通过外部攻击和内部攻击来记录或推测敏感数据信息,但其遵守协议且不篡改数据,仅破坏数据的隐私性但不破坏数据的完整性.因此,其他非诚信类型的攻击,包括 DoS(Denial of Service)、数据篡改删除等,不在本文的考虑范围内.我们假设基站节点是可信的,并且不可被俘获,同时攻击者能够俘获节点的数目是有限的,因为当网络中超过一定比例的节点都被俘获时,已经无法保证网络的正常性能.

考虑到管理节点俘获攻击和共谋攻击对数据隐私安全的显著影响,本文首先针对管理节点俘获攻击提出了 PKN 隐私保护协议,接着针对共谋攻击提出了 CPKN 隐私保护协议.

3.3 问题定义

本文的目标是提出适用于双层传感器网络的高效率低能耗的隐私保护 k-NN 查询处理协议.假设基站将查询请求 $\mathcal{Q}$ 发送到管理节点 $s_{\mathcal{M}}$,令 $\mathcal{B}=\{A_i[j]\mid 1\leq i\leq N,1\leq j\leq d\}$ 为网络中所有传感器节点采集的数据集合,$\mathcal{D}=\{Dis(G,s_i)\mid 1\leq i\leq N\}$ 为所有传感器节点与查询点的距离值集合,Ω_k 为 k-NN 查询的结果集合,则在面对俘获攻击和共谋攻击时,隐私保护 k-NN 查询处理协议需要同时满足以下要求:

(1) 数据隐私性.$\mathcal{B}$和$\mathcal{D}$的真实值只能被传感器节点本身和基站获得,而不允许被 $s_{\mathcal{M}}$ 获得.

(2) 查询隐私性.查询请求$\mathcal{Q}$会泄露用户的偏好等隐私信息,因此查询参数 k 只允许被保存在基站上.

(3) 正确性.$\mathrm{Max}(Dis(G,\Omega_k))\leq\mathrm{Min}(Dis(G,\mathcal{B}/\Omega_k))$.

(4) 效率.协议在提供高强度隐私保护的同时

需要保证较高的查询效率和较低的能耗.

4　隐私保护 k-NN 查询协议

管理节点在被攻击者俘获时，容易直接泄露节点上存储的大量的感知数据，严重危害网络安全. 因此，本节中我们给出了面向管理节点俘获攻击的 PKN(Privacy-preserving k-NN query)协议.

PKN 协议的目标是为基于值的 k-NN 查询提供数据隐私保护的同时保持较高的查询和能耗效率. 因此，我们分为 2 个阶段来实现 PKN 协议. 首先，为保证查询效率，构建适用于双层传感器网络的 KNQ 查询架构(k-NN Query framework). 接着，为抵御窃听攻击和管理节点俘获攻击，在 KNQ 查询架构上提出基于 Bloom filter 编码的隐私保护机制.

4.1　KNQ 查询架构

KNQ 查询架构的构建基于两个观察：(1) 将潜在的查询响应数据定向存储到离基站更近的管理节点，可以缩短查询响应时间；(2) 由于管理节点拥有充足的资源，所以应该尽量将数据通信和查询处理操作上推到管理节点执行，以节省普通节点的能量，提高网络的能耗效率，延长生存周期. 本小节首先给出了 KNQ 查询架构的实现步骤，包括语言变量构建、数据定向存储和数据查询处理，接着指出了 KNQ 架构存在的隐私泄露问题. 4.2 节针对 KNQ 架构的隐私泄露风险，给出了基于编码机制的隐私保护方案.

4.1.1　基于值的语言变量

在 KNQ 查询架构中，我们将 k-NN 查询中节点与指定值 G 之间距离的值域划分为非均匀的子域，不同的子域对应与指定值 G 之间不同的距离，越靠近 0 的子域的间隔越小. 每个子域对应不同的基于值的语言变量 VLV(Value based Linguistic Variables). 表 1 给出了基于值的语言变量的划分与表示的实例.

表 1　基于值的语言变量的实例

$VLVs$	Semantic meaning	Subfields
A	Almost the value	$[0,3)$
B	Very close	$[3,8)$
C	Close	$[8,16)$
D	Adjacent	$[16,33)$
E	Noteworthy	$[33,50)$
F	Unremarkable	$[50,+\infty)$

表 1 中将距离的值域划分为 6 个子域，分别用 A 到 F 表示. A 对应最小的子域$[0,3)$以及相应的语义“几乎是指定值”. F 代表离指定值最远的子域并对应语义“不值得关注”，在很多应用中可以根据具体情况直接进行过滤操作. 通常来说，基于值的语言变量的子域参数选择可通过统计方法获得，或者直接由领域专家给定.

4.1.2　数据定向存储

在 KNQ 架构中，距离值域的划分由基站完成，并发送到管理节点上. 传感器节点在上传数据到最近的管理节点时，将落在同一个 VLV 子域的感知数据定向存储到一个指定的管理节点，可以在查询时有效缩短查询响应时延. KNQ 使用 hash 函数将每个 VLV 映射到一个指定的管理节点进行相应的数据存储，子域值较小的 VLV 存储在离基站更近的管理节点. 因此，在 KNQ 中，管理节点既要进行 VLV 判断操作，同时又是某一个或多个指定 VLV 的存储节点. 图 3 给出了表 1 中 VLV 的定向存储的实例. 子域为$[0,3)$的 $VLV(A)$存储在离基站最近的管理节点上. 例如当 $G=10$ 时，感知数据 $V_i=11$ 将存储在 $VLV(A)$ 对应的管理节点上；感知数据 $V_{i+1}=17$ 将存储在 $VLV(B)$对应的管理节点上；而 $V_{i+2}=65$ 将直接被过滤.

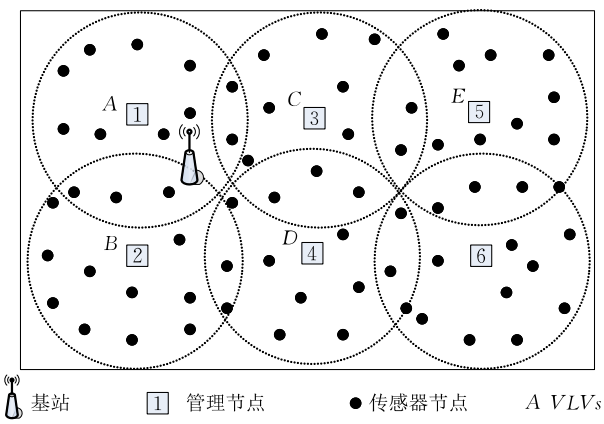


图 3　基于 VLV 的定向存储

4.1.3　数据查询处理

在执行 k-NN 查询的过程中，基站首先根据参数 k 和历史记录确定需要查询的管理节点. 例如，当查询 2-NN 时，基站仅向存储 VLV-A 的管理节点下发查询请求；当查询 8-NN 时，基站同时向 $VLV(A)$、$VLV(B)$、$VLV(C)$的存储节点下发查询请求. 接收到查询请求的管理节点将存储的数据发送到基站. 基站只需要在接收到的数据上进行排序操作，就可

以获得最终的查询结果.在特殊情况下,有很小的概率基站查询到的数据不能满足参数 k 的需求,基站需要再次向下一个 VLV 存储节点下发查询请求来完成查询.

4.1.4 KNQ 架构的隐私泄露风险

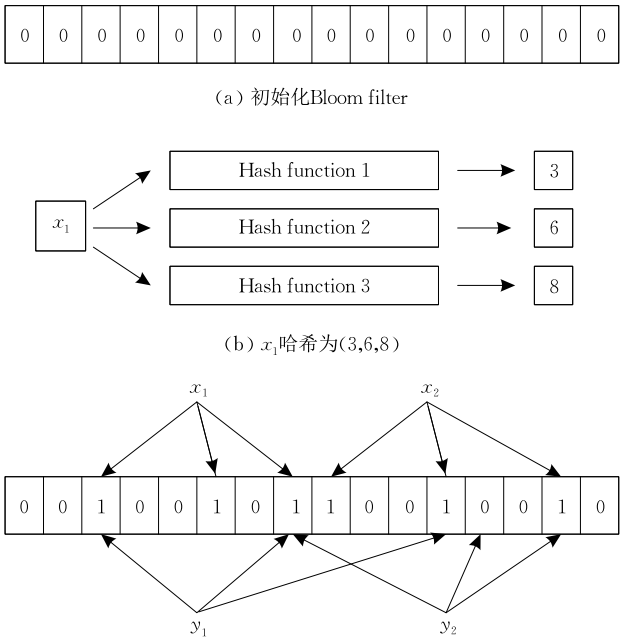
KNQ 架构可以在减少普通传感器节点能耗的同时缩短 k-NN 查询的时延,保证较高的能耗效率和查询效率.然而,考虑到传感器网络中存在的窃听攻击和节点俘获攻击风险,在 KNQ 架构中,管理节点一旦被俘获,将严重泄漏网络的隐私信息.因此,未加密的原始敏感数据都不允许被转发或存储到管理节点,包括:(1)传感器节点感知数据;(2)基站查询请求;(3)语言变量子域划分表.然而,管理节点无法在加密数据上完成 VLV 判定和定向存储操作.本文中我们使用附加编码的方法来反映加密数据的真实值,协助基站和管理节点在不泄露隐私的前提下完成数据查询处理.4.2 节给出了基于改进的 Bloom filter 进行隐私保护 k-NN 查询处理的 PKN 协议.

4.2 隐私保护 k-NN 查询协议(PKN)

本节首先对标准的 Bloom filter[36] 进行概要介绍,接着对 Bloom filter 进行改进以实现隐私保护 k-NN查询.

4.2.1 Bloom filter 概述

Bloom filter 是具有很好的时间和空间效率的随机数据结构,通常用在成员关系判定中.如图 4(a)所示,初始状态时,Bloom filter 是一个包含 m 位的位数组,每一位都置为 0.为了表达 n 个元素的集合 $X=\{x_1,x_2,\cdots,x_n\}$,Bloom filter 使用 l 个相互独立的 hash 函数,将集合中的每个元素映射到 l 个分布在范围内的位置上,并将数组中相应的位置置为 1.如图 4(b),$h_1(x_1)=3$,$h_2(x_1)=6$,$h_3(x_1)=8$,则将数组中的第(3,6,8)位都置为 1.类似的,为表示元素 x_2,也将第(9,12,15)位都置为 1.如果一个位置多次被置为 1,则只有第一次会起作用,其他几次将没有任何效果.当判定元素 y 是否属于集合 X 时,检查是否所有的 $h_i(y)(1\leq i\leq l)$ 位都为 1,如果不成立,则 y 一定不属于 X;如果成立,并不能 100%确定 y 属于 X.这是因为 Bloom filter 存在一定的假阳性(False Positive)概率,当所有的 $h_i(y)(1\leq i\leq l)$ 位都恰好被集合 X 中的其他元素置为 1 时,元素 y 会被错误地判定属于集合 X.例如图 4(c),集合 $X=\{x_1,x_2\}$ 由 Bloom filter 编码 0010010110010010 表示,元素 y_1 的 hash 值为(3,8,12),由于编码中以上 3 位都为 1,y_1 将被判定为属于集合 X,这是一个假阳性误判.而 hash 值为(8,13,15)的 y_2 将直接被判定为不属于 X.



(a) 初始化Bloom filter

(b) x_1哈希为(3,6,8)

(c) 检测y_1, y_2是否属于集合$\{x_1, x_2\}$

图 4 Bloom filter 机制示意图

由于假阳性判定不会对结果的完整性产生影响,同时发生的概率较低且是可控的,Bloom filter 机制可以应用在隐私保护 k-NN 查询处理中.定理 1[36] 给出了 Bloom filter 发生假阳性判定的概率.

定理 1. 设 m 为 Bloom filter 数组的长度,l 为 hash 函数的数目,n 为集合 X 中元素的数目,P 为假阳性(False Positive)判定发生的概率.则

$$P\in\left[0,\left(1-\left(1-\frac{1}{m}\right)^{nl}\right)^{l}\right).$$

简单起见,$\left(1-\left(1-\frac{1}{m}\right)^{nl}\right)^{l}$ 可以用 $(1-\mathrm{e}^{-nl/m})^{l}$ 来表示.

4.2.2 隐私保护数据定向存储

上文已经提到非加密的数据不允许被转发和存储到管理节点,因此,我们首先对本文使用的全局同步的端到端加密机制进行定义.在网络初始化阶段,传感器节点 s_i 与基站共享点对点的密钥种子 $k_{i,0}$.在周期 t 中,s_i 使用 $k_{i,t}$ 对数据进行加密并上传到管理节点,$k_{i,t}=\mathrm{hash}(k_{i,t-1})$ 是节点 s_i 在周期 t 中的密钥,而 $k_{i,t-1}$ 在周期 t 开始时将被节点 s_i 删除.管理节点无法获得传感器节点与基站之间的同步密钥,因

此，即使被攻击者俘获也无法对加密数据进行解密来获得隐私数据信息.

同时，我们对 Bloom filter 机制进行改进，将感知数据和 VLV 子域分别转化成数据编码的形式，从而在加密数据上完成 VLV 判定和数据定向存储操作. 接下来，我们分别给出基站、传感器节点和管理节点上的隐私保护定向存储机制.

基站：在网络初始阶段，基站与普通传感器节点间共享一个由 l 个相互独立的 hash 函数$\{h_1,h_2,\cdots,h_l\}$组成的 hash_{BF} 函数集合. 基站为每个 VLV 子域构建一个 m 位的 Bloom filter 数组，使用 hash 函数集合$\{h_1,h_2,\cdots,h_l\}$将子域中的每个元素映射到 l 个位置，并将数组中相应的位置置为 1，重复这一过程，直到子域中的所有元素都被映射到 Bloom filter 数组中. 这样，每一个 VLV 子域都可以用一个 m 位的 Bloom filter 编码进行表示.

基站将所有 VLV 子域对应的 Bloom filter 编码发送到管理节点. 消息格式如下：

$Sink\rightarrow s_{\mathcal{M}}:\{t;BF_A,BF_B,\cdots,BF_{|VLVs-1|};\text{Hash}_{\text{match}}\}$，

其中：t 为周期，BF_A 为 $VLV(A)$ 子域对应的 Bloom filter 编码数组. BF_{VLVs} 为最后一个 $VLVs$ 的编码，通常会被直接过滤掉，因此，$BF_{|VLVs-1|}$ 为最后一个值得关注的子域的 Bloom filter 编码. $\text{Hash}_{\text{match}}$ 是存储节点分配函数，为每一个 VLV 子域分配指定的存储节点. 距离值较小的 VLV 子域的数据被存储在离基站较近的管理节点上.

普通传感器节点：传感器节点 s_i 采集数据并计算距离值 $Dis(G,s_i)$，使用 hash 函数集合$\{h_1,h_2,\cdots,h_l\}$将 $Dis(G,s_i)$ 映射为一个 m 位的 Bloom filter 编码，接着将 Bloom filter 编码和加密后的真实数据发送到最近的管理节点. 消息格式如下：

sensor $s_i\rightarrow s_{\mathcal{M}}:\{t;ID_i;BF_{s_i};E_{k_{i,t}}(V_{s_i})\}$，

其中：ID_i 为节点 ID，BF_{s_i} 表示 $Dis(G,s_i)$ 的 Bloom filter 编码，$E_{k_{i,t}}(V_{s_i})$ 表示使用密钥 $k_{i,t}$ 对 V_{s_i} 加密后的密文.

管理节点：在基站和传感器节点已有的前期准备的基础上，管理节点可以基于定理 2 和定理 3 完成隐私保护 VLV 判定和数据定向存储操作.

定理 2.　如果 $BF_{s_i}\cap BF_A\neq BF_{s_i}$，一定有 $Dis(G,s_i)\notin VLV(A)$.

证明.　使用反证法来进行证明. 假设当 $BF'_{s_i}\cap BF_A\neq BF'_{s_i}$ 时，$\exists Dis(G,s_i)'\in VLV(A)$，其中 BF'_{s_i} 表示 $Dis(G,s_i)'$ 的 Bloom filter 编码. 如果 $Dis(G,s_i)'$ 的值属于子域 A，那么它对应的 Bloom filter 编码中的 l 个 hash 位置都应该被置为 1，则一定有 $BF'_{s_i}\cap BF_A=BF'_{s_i}$，与假设相矛盾，原命题得证.　　证毕.

定理 3.　当 $BF_{s_i}\cap BF_A=BF_{s_i}$ 时，$Dis(G,s_i)\in VLV(A)$ 以一定概率成立.

证明.　定理 3 是定理 2 的逆反命题，由于定理 2 成立，则定理 3 一定成立. 当 $BF_{s_i}\cap BF_A=BF_{s_i}$ 时，令 $\hat{P}$ 表示 $Dis(G,s_i)\in VLV(A)$ 的概率，由定理 1 可得

$$\hat{P}\in\left[\left(1-\left(1-\left(1-\frac{1}{m}\right)^{nl}\right)^{l}\right),1\right].$$

当管理节点接收到传感器节点发来的消息时，首先提取出 s_i 的 Bloom filter 编码 BF_{s_i}，然后与从基站接收到的每一个 VLV 子域的 Bloom filter 编码进行比较，以求得满足式(1)的参数 U.

$$\exists U(A\leq U\leq|VLVs-1|):BF_{s_i}\cap BF_U=BF_{s_i}\quad(1)$$

如果 U 存在，则 s_i 的数据属于 $VLV(U)$ 子域. 管理节点使用 $\text{Hash}(U)$ 来确定 $VLV(U)$ 的存储节点 $s_{\mathcal{M}}(U)$，并将从 s_i 接收到的消息直接转发到 $s_{\mathcal{M}}(U)$ 节点.

如果 U 不存在，则一定有 $Dis(G,s_i)\in VLV$ (Unremarkable)，传感器节点 s_i 的数据通常被直接被过滤掉.

4.2.3　隐私保护数据查询处理

与 KNQ 查询架构类似，基站根据参数 k 和历史记录确定需要查询的 VLV 子域，并将查询请求下发到这些 VLV 子域对应的管理节点. 管理节点在接收到查询请求后，将存储的数据发送到基站，消息格式如下：

$$s_{\mathcal{M}}\rightarrow Sink:\{t;ID_i;BF_{s_i};E_{k_{i,t}}(V_{s_i})\}.$$

基站接收到查询响应数据后，首先根据 t 和 ID_i 来匹配得到密钥 $k_{i,t}$；接着通过解密得到真实值 V_{s_i}，并求解真实距离值 $Dis(G,s_i)$，从而将假阳性判定引起的误判数据删除掉；最后通过对 $Dis(G,s_i)$ 进行排序来得到精确的查询结果数据. 由于基站可以通过真实 $Dis(G,s_i)$ 值将假阳性判定数据与正确数据区分开，Bloom filter 的假阳性误判问题并不会对 k-NN 查询结果集的正确性产生影响.

4.2.4　PKN 安全性分析

隐私保护 k-NN 查询的要求是节点的真实数据 $\mathcal{B}$ 和真实距离值 $\mathcal{D}$ 只能被节点本身和基站获取. 考虑到节点真实数据使用与基站的端到端密钥加密后传输到管理节点，在无密钥的情况下，管理节点无法通

过解密获得节点真实数据值.另一种攻击方式是被俘获的管理节点通过接收到的传感器节点 s_i 的编码 BF_{s_i},猜测 BF_{s_i} 与真实距离值 $Dis(G,s_i)$ 的一一对应关系.管理节点通过观察能够确定传感器节点使用的哈希函数的个数 l 以及编码数组的长度 m,则攻击者能够正确猜测编码与真实距离值对应关系的概率为 $P_i=\left(\binom{m}{l}\times l!\right)^{-1}$.

4.2.5 PKN 性能分析

(1)PKN 协议计算复杂度.参照文献[12-13,16],我们对 PKN 协议中普通节点和管理节点的计算复杂度进行分析.普通节点需要对数据进行 BF 编码操作和加密操作,BF 编码操作的计算复杂度为 $O(l)$,其中 l 为哈希函数的个数,设加密算法的复杂度为 $O(E)$,则普通节点的计算复杂度为 $O(l)+O(E)$.

管理节点上单次编码比较操作的计算复杂度为 $O(m)$,假设每个管理节点收到 c 个普通节点的数据,VLV 子域的划分数量为 v,则管理节点的计算复杂度为 $O(cvm)$.

(2)传感器节点通信能耗.令 $\ell_{\mathrm{id}},\ell_{\mathrm{BF}},\ell_{\mathrm{d}}$ 分别表示上传消息中 ID,BF_{s_i} 和 $E_{k_{i,t}}(V_{s_i})$ 所占的字节数,$L_{is_{\mathcal{M}}}$ 表示节点 s_i 到 $s_{\mathcal{M}}$ 的路由跳数,则 PKN 协议中传感器节点的总能耗为

$$C_S=\sum_{i=1}^{N}(\ell_{\mathrm{id}}+\ell_{\mathrm{BF}}+\ell_{\mathrm{d}})\times L_{is_{\mathcal{M}}}.$$

(3)管理节点通信能耗.令 $VLV(k)$ 表示完成 k-NN 查询使用到的最大的 VLV 子域,$L_{ss_{\mathcal{M}}}$ 表示数据定向存储到对应管理节点的路由跳数,L_{sink} 表示管理节点向基站返回查询结果的路由跳数.则 PKN 协议中管理节点的总能耗为

$$C_{s_{\mathcal{M}}}=\sum_{i=1}^{N}(\ell_{\mathrm{id}}+\ell_{\mathrm{BF}}+\ell_d)\times L_{ss_{\mathcal{M}}}+\sum_{U=A}^{VLV(\lceil k\rceil)}(\ell_{\mathrm{id}}+\ell_{\mathrm{BF}}+\ell_d)\times|VLV(U)|\times L_{sink}.$$

由以上分析可知,PKN 协议可以在管理节点被俘获的情况下不泄露隐私的完成 k-NN 查询,并且保持较低的能耗.本文第 6 节通过实验对 PKN 协议的性能进行了验证.

4.3 PKN 协议改进与优化

本节中,我们从隐私保护强度、协议适用范围和网络性能等方面对 PKN 协议进行改进,分别提出分组随机存储机制、动态参数 G 查询机制、VLV 子域划分机制、编码长度选择机制和编码数据压缩机制等 5 种针对 PKN 协议的优化策略.

4.3.1 分组随机存储机制

在 PKN 协议中,管理节点存储的子域范围与该管理节点与基站距离存在相关性,攻击者可以根据管理节点与基站的距离远近来推测不同管理节点子域范围的大概值域和大小关系,从而推测多个普通节点距离值的大概取值范围和大小关系.虽然攻击者无法准确推测出管理节点的真实子域范围,也无法推测普通感知节点的准确距离值,但仍然可能泄露数据隐私信息.针对以上问题,我们给出了分组随机存储机制.

在基站为不同 VLV 子域指定存储节点前,首先根据子域范围大小对 VLV 子域进行分组,同时根据管理节点与基站距离对管理节点进行相应分组.以图 3 为例,将 $VLV(A)$、$VLV(B)$ 和 $VLV(C)$ 划分为第 1 组,对应的第 1 组管理节点为 1、2 和 3;将 $VLV(D)$ 和 $VLV(E)$ 划分为第 2 组,对应第 2 组管理节点为 4、5 和 6.基站在指定存储节点时,从第 1 组管理节点中为第 1 组 VLV 随机选择存储节点,从第 2 组管理节点中为第 2 组 VLV 随机选择存储节点.例如,$VLV(A)$,$VLV(B)$ 和 $VLV(C)$ 分别存储在管理节点 2,3 和 1 上,$VLV(D)$ 和 $VLV(E)$ 分别存储在管理节点 6 和 4 上.为提高安全性,基站将周期性的对分组随机存储的分配进行更新.

在未引入分组随机存储机制时,攻击者能推测多个普通节点数据距离值的大概取值范围和大小关系;在引入分组随机存储机制后,攻击者既不能确定子域取值范围也不能确定数据间的大小关系,保证了敏感数据的隐私性.同时,由于分组策略考虑了管理节点与基站间的距离,分组随机存储机制不会对网络性能产生明显影响.

4.3.2 动态参数 G 查询机制

在传感器网络中,k-NN 查询通常应用于数据监测环境中,参数 G 在多数情况下保持固定.然而,随着网络环境和应用需求的变化,用户也可能更改参数 G 的值.当 G 发生动态变化时,PKN 协议按照如下步骤执行:

(1)用户将新的参数 G 及相应的 VLV 子域范围划分方式提交到基站;

(2)基站根据新的 VLV 子域划分范围生成相应的 Bloom filter 编码,并发送到管理节点;

(3)基站使用与普通节点共享的公钥对 G 进行

加密，并通过管理节点下发到普通节点；

(4) 普通节点解密后使用新的 G 值作为标准计算 $Dis(G,s_i)$，并生成 BF 编码；

(5) 普通节点的数据上传操作和管理节点上的定向存储操作与固定参数 G 的情况完全相同.

新的参数 G 的下发操作和新的 VLV 编码计算及传输都由基站和高资源的管理节点完成，参数 G 的更新并不会对普通节点的能耗和生命周期产生明显影响. 由于基站与高资源管理节点的能量都是充足的，因此能够满足参数 G 动态更新的需求.

4.3.3　VLV 子域划分机制

在 PKN 协议中，VLV 子域划分方式对查询响应结果集的大小有直接影响，本节我们对 VLV 的划分方式进行讨论. 与文献[12-13]相同，假设基站通过历史数据估计得出距离值的分布函数 $F(x)$. 设第 i 个 VLV 子域的上下界分别为 l_i 和 h_i. 由 4.2.5 节中的计算复杂度分析可知，管理节点执行的编码比较操作的次数与 VLV 子域的数量 v 成正比，过高的 VLV 子域数量会增加管理节点的负载. 因此，VLV 子域划分粒度不应过小. 常用的划分方式包括等宽划分、等高划分和递增划分 3 种.

(1) 等宽划分方式：所有的 VLV 子域采用相同的值域范围，即 $h_1-l_1=\cdots=h_v-l_v$. 当 V_K 落在第 i 个 VLV 子域时，等宽划分方式冗余数据的上界为

$$\sum_{x=l_i}^{h_i}F(x),\ 1\leq i\leq v.$$

可以看出，当 VLV 子域 i 对应大量的节点数据时，查询时可能导致明显的冗余数据问题.

(2) 等高划分方式：以不同 VLV 子域对应相等数量的节点数据为原则进行划分. 节点数据的数量即为 VLV 子域的高度值. 设高度值为 T，则等高划分方式的冗余数据的上界为 T. 可以通过设置 T，将冗余数据限制在一定范围内，然而当 k 较小时，冗余数据的比例仍旧比较高，如 $k=2$ 时，冗余比例为 $(T-2)/T$.

(3) 递增划分方式：距离值较小的 VLV 的划分粒度较小，随距离值的增大，VLV 的划分粒度逐渐增大，即 $h_1-l_1<\cdots<h_v-l_v$. 当 k 较小时，冗余数据的比例低，然而当查询涉及值域较大的子域时，容易导致大量的冗余数据.

结合 k-NN 查询的特点，兼顾查询结果集较小和较大的情况，我们采用混合划分方式. 对距离值较小的 VLV 子域，采用递增式划分，以尽量减少冗余数据比例. 而考虑查询结果集较大的情形，对距离值较大的 VLV 子域，采用等高划分方式，控制冗余数据的上限. 采用混合划分方式可以保证当 k 较小时，冗余数据比例较小，同时当 k 较大时，冗余数据量也不超过 T.

4.3.4　Bloom filter 数组长度选择机制

在进行 k-NN 查询时，过高的假阳性判定概率会导致查询时的冗余数据问题. 由定理 1 可知，假阳性判定概率与数组长度成反比，然而采用过长的 Bloom filter 数组编码时会增加普通节点的通信代价，缩短网络的生存周期. 定理 4 给出了任意指定假阳性概率和元素数量所对应的最小 Bloom filter 数组编码长度.

定理 4.　设 ε 为指定的假阳性判定概率，m 为 Bloom filter 数组的长度，l 为 hash 函数的数目，n 为集合 X 中元素的数量，P 为假阳性(False Positive)判定发生的概率. 满足 $P\leq\varepsilon$ 的最小 Bloom filter 数组长度 m 为

$$m\geq n\log_2 e\times\log_2(1/\varepsilon).$$

证明.　由定理 1 可知[36]，对于指定的 m 和 n，保证最低假阳性概率的最优 hash 函数个数为 $l=\left(\frac{m}{n}\right)\ln 2$，因此，假阳性判定概率可以表示为

$$P=(1-e^{-((\frac{m}{n})\ln 2)(\frac{n}{m})})^{(\frac{m}{n})\ln 2}.$$

由 $P\leq\varepsilon$ 可得，$m\geq n\log_2 e\times\log_2(1/\varepsilon)$，得证. 证毕.

第 6 节实验部分对 Bloom filter 数组长度与元素个数之间的关系进行了评估.

4.3.5　数据压缩机制

在传感器节点数据生成 BF 编码的过程中，由于仅使用 l 个 hash 函数，编码中仅有 l 位为“1”，其他位全部为“0”. 为减少普通节点上传的数据量，需要对 BF 编码进行数据压缩. 本小节根据数据压缩方式的不同，分别给出了两种数据压缩机制：绝对位置数据压缩和相对位置数据压缩.

(1) 绝对位置数据压缩机制

用“1”的绝对位置的集合来表示 BF 编码. 例如，用{1,4,12}表示 BF 编码〈1001000000010000〉，其中 1、4、12 分别表示“1”在 BF 编码中的绝对位置. 相比于初始的 16 位编码，压缩后只需要 8 位就可以表示 BF 编码，压缩率为 2. 设 l 为哈希函数的数量，m 为 BF 编码的长度，B^j 为第 j 个“1”在 BF 编码中的绝对位置，Bit 为编码压缩后所需的位数. 以下公式分别给出了基于绝对位置的压缩机制所需

位数的最大值和最小值.

$$Bit_{\max}=\max\Big\{\sum_{j=1}^{l}(\log_2 B^j+1)\Big\}$$

$$=\sum_{i=m-l+1}^{m}(\log_2 i+1),$$

$$Bit_{\min}=\min\Big\{\sum_{j=1}^{l}(\log_2 B^j+1)\Big\}$$

$$=\sum_{i=1}^{l}(\log_2 i+1).$$

可以看出,当所有的"1"都落在编码的最后 l 位时,压缩后的位数最多;当所有的"1"都落在编码的最前 l 位时,压缩后的位数最少.

(2)相对位置数据压缩机制

用相邻两个"1"的位置之差的集合来表示 BF 编码. 如用{1,3,8}表示 BF 编码〈1001000000010000〉,其中 1 是第一个"1"在 BF 编码中的绝对位置,3 和 8 分别是 BF 编码中相邻两个"1"的位置之差. 基于相对位置的压缩机制只用 7 位就可以表示 16 位的 BF 编码,压缩率为 2.286. 以下公式给出了基于相对位置的数据压缩机制所需位数的最大值.

$$Bit'_{\max}=\max\Big\{\sum_{j=1}^{l}(\log_2 B^j-B^{j-1}+1)\Big\}$$

$$=l\times\Big(\log_2\frac{m}{l}+1\Big).$$

可以看出,当所有的"1"均匀分布在 BF 编码中时,压缩后的位数最多;当所有的"1"都落在编码的最前 l 位时,压缩后的位数最少.

显然,无论是绝对位置压缩还是相对位置压缩,管理节点都可以还原出初始 BF 编码,并正确地完成 *VLV* 判定. 数据压缩机制仅应用于普通节点数据,以减少数据上传量,节省普通节点能量. 由于基站与管理节点能量充足,因此 *VLV* 编码不做压缩处理. 6.2.1 节实验在对两种编码压缩算法的性能进行了评估.

5 面向共谋攻击的 k-NN 查询协议

双层传感器网络中另一类危害较大的攻击类型是普通传感器节点与管理节点间的共谋攻击. 攻击者通过共谋攻击能够直接获得传感器节点上的 Bloom filter 编码 hash_{BF} 函数集合. 基于这些函数,攻击者可以枚举得到:(1)管理节点上存储的感知数据的真实距离值$\mathcal{D}$;(2)基站下发的 *VLV* 子域的真实范围与定向存储信息. 因此,传感器节点与管理节点间的共谋攻击将使 PKN 协议的保护机制失效,并严重影响网络的数据安全.

我们在 PKN 协议的基础上提出了抵御共谋攻击的 CPKN 协议,首先给出了节点匿名方法,接着基于迭代处理机制给出了基础的 CPKN-naive 协议,然后对它进行了改进,提出了基于编码分片传输机制的 Slice-CPKN(CPKN-s)协议.

5.1 节点匿名空间

为抵御共谋攻击,基站在网络初始化阶段为每个传感器节点构建匿名空间. 如图 5 所示,基站上建立由 $N\times\eta$ 个虚假 ID 建立的总体匿名空间,然后采用随机选择的方式为每个传感器节点分配一个由 η 个虚假 ID 组成的匿名子空间,并直接内置在传感器节点上. 传感器节点上传数据时,每次从子匿名空间中随机选择一个虚假 ID 来替代真实 ID 进行数据传输.

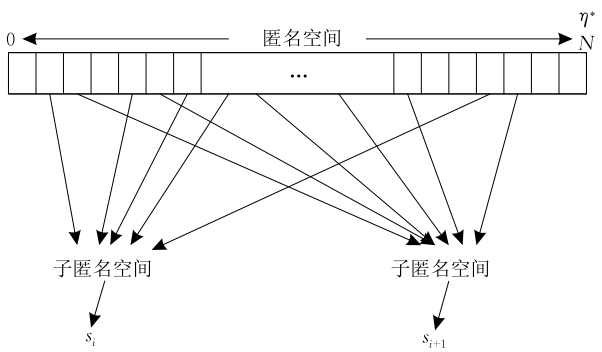


图 5 匿名空间机制示意图

5.2 Naïve CPKN 协议

5.2.1 Naïve CPKN 协议实现

Naïve CPKN 协议的核心思想是基于节点间的迭代处理来达到抵御共谋攻击的目的. 首先,在网络中选择一系列传感器节点作为代理节点(Agent Nodes),用 $A_\theta(\theta=1,2,\cdots,w)$ 表示,为保证数据精度,代理节点仍然正常进行数据采集操作. 代理节点的选择由基站控制,并根据能耗状况进行周期性的更新. 基站通过端到端加密向选定的代理节点发送如下参数:ID 单路完美哈希函数 $h^{\text{ID}}_{A_\theta}$、编码单路哈希函数 $h^{C}_{A_\theta}$ 和端到端密钥 k_{A_θ}. 完美哈希函数的性质是能够保证避免哈希碰撞问题. $h^{\text{ID}}_{A_\theta}$ 的作用是将收到的 ID 哈希为新的 ID,$h^{C}_{A_\theta}$ 的作用是将收到的编码中的每个 1 位随机哈希到其他位,从而产生一个新的编码. k_{A_θ} 仅在 A_θ 与基站间共享,用来加密接收到的

数据.

算法 1 给出了 Naïve CPKN 协议中基于代理节点迭代处理的数据隐藏机制的详细步骤. 当 A_1 节点接收到传感器节点 s_i 发送来的数据 $\{t;pID_i;BF_{s_i};E_{k_{i,t}}(V_{s_i})\}$ 后，令 $pID_i^0=pID_i$，$BF_{s_i}^0=BF_{s_i}$，$E^0(V_{s_i})=E_{k_{i,t}}(V_{s_i})$. A_1 节点分别对数据进行如下处理：$pID_i^1=h_{A_1}^{\mathrm{ID}}(pID_i^0)$，$BF_{s_i}^1=h_{A_1}^{C}(BF_{s_i}^0)$，$E^1(V_{s_i})=E_{k_{A_1}}(E^0(V_{s_i}))$. 然后，将处理后的数据发送至下一个代理节点 A_2. A_2 接收到消息后重复数据处理过程，直到消息到达 A_w 节点.

代理节点 A_w 在对接收到的数据处理后，直接发送到管理节点. 因此，管理节点接收到的 s_i 的数据为 $\{t;pID_i^w;BF_{s_i}^w;E^w(V_{s_i})\}$.

基站将对所有 VLV 子域进行重编码，将每一个元素对应的编码进行与代理节点相同的哈希处理后，组成新的 VLV 子域 Bloom filter 编码，并发送到管理节点. 消息格式如下：

$Sink\rightarrow s_{\mathcal{M}}$：$\{t;BF_A^w,BF_B^w,\cdots,BF_{|VLVs-1|}^w;\mathrm{Hash}\}$.

管理节点采用与 PKN 协议完全相同的定向存储和数据查询处理机制来完成隐私保护 k-NN 查询.

算法 1.　Naïve CPKN 协议代理节点数据迭代隐藏机制.

输入：$\{t;pID_i;BF_{s_i};E_{k_{i,t}}(V_{s_i})\}$，$A_\theta(\theta=1,2,\cdots,w)$，$h_{A_\theta}^{\mathrm{ID}},h_{A_\theta}^{C},k_{A_\theta}$

输出：$\{t;pID_i^w;BF_{s_i}^w;E^w(V_{s_i})\}$

1. $pID_i^0\leftarrow pID_i$
2. $BF_{s_i}^0\leftarrow BF_{s_i}$
3. $E^0(V_{s_i})\leftarrow E_{k_{i,t}}(V_{s_i})$
4. FOR $1\leq\theta\leq w$
5. $\quad pID_i^\theta\leftarrow h_{A_\theta}^{\mathrm{ID}}(pID_i^{\theta-1})$
6. $\quad BF_{s_i}^\theta\leftarrow h_{A_\theta}^{C}(BF_{s_i}^{\theta-1})$
7. $\quad E^\theta(V_{s_i})\leftarrow E_{k_{A_\theta}}(E^{\theta-1}(V_{s_i}))$
8. $\quad A_\theta\rightarrow A_{\theta+1}:\{t;pID_i^\theta;BF_{s_i}^\theta;E^\theta(V_{s_i})\}$
   //代理节点 A_θ 将消息发送给 $A_{\theta+1}$ 代理节点
9. END FOR
10. RETURN $\{t;pID_i^w;BF_{s_i}^w;E^w(V_{s_i})\}$
    //代理节点将隐藏后的数据发送到管理节点

5.2.2　Naïve CPKN 协议安全性分析

在 Naïve CPKN 协议中，攻击者根据俘获节点的不同，可以实施两种形式的共谋攻击：管理节点与传感器节点的共谋攻击；管理节点、代理节点与传感器节点的共谋攻击. 下面针对这两种情况分别进行安全性分析：

(1) 管理节点 $s_{\mathcal{M}}$ 与传感器节点 s_i 的共谋攻击. 经过 w 个代理节点对传感器节点数据的处理后，即使 s_i 也无法从 $s_{\mathcal{M}}$ 上确定其自身数据 $\{t;pID_i^w;BF_{s_i}^w;E^w(V_{s_i})\}$. 管理节点虽然获得了传感器节点所使用的 Bloom filter 哈希函数，但是由于传感器节点的原始 BF_{s_i} 已经被 w 个哈希函数转化为 $BF_{s_i}^w$，管理节点很难从 $BF_{s_i}^w$ 逆向推测出原始的 BF_{s_i}. 设 l 为 s_i 节点上 Bloom filter 哈希函数的个数，则管理节点能够正确地逆向推测原始 BF_{s_i} 的概率为 $P_i=\binom{m}{l}^{-1}$. 因此，Naïve CPKN 协议能够有效的抵御管理节点与传感器节点的共谋攻击.

(2) 管理节点 $s_{\mathcal{M}}$、代理节点 A_θ 与传感器节点 s_i 的共谋攻击. 假设攻击者同时俘获了多个传感器节点、代理节点和管理节点，如果被俘获的代理节点不能构成从传感器节点到管理节点的完整路径，则仍旧不能完成节点原始数据与管理节点数据的一一对应. 只有当所有的代理节点都被俘获时，传感器节点才可以通过重复代理节点完整的数据处理操作来进行枚举攻击，在原始值与管理节点存储值之间建立联系. 然而，由于攻击者能够俘获的节点数目是有限的，并且代理节点在网络中的分布是不确定的，攻击者俘获完整路径的几率很小. 假设攻击者能够俘获的节点数为 Γ，代理节点数目为 w，网络中传感器节点数目为 N，攻击者俘获的任意一个节点为普通节点的概率为 $(1-w/N)$，刚好为第 θ 个代理节点的概率为 $(1/N)$，则攻击者在俘获 Γ 个节点时，能够同时俘获代理节点完整路径的概率 P_c 为

$$P_c=\begin{cases}0, & \Gamma<w\\ \dfrac{\Gamma!}{(\Gamma-w)!}\left(\dfrac{1}{N}\right)^w\left(1-\dfrac{w}{N}\right)^{\Gamma-w}, & \Gamma\geq w\end{cases}.$$

因此，在代理节点足够多时，攻击者很难通过管理节点、代理节点与传感器节点间的共谋攻击来获得隐私数据.

5.2.3　Naïve CPKN 协议能耗分析

(1) Naïve CPKN 协议计算复杂度. 我们分别对普通节点、代理节点和管理节点的计算复杂度进行分析. 普通节点的计算复杂度与 PKN 协议相同，仍为 $O(l)+O(E)$.

代理节点上 $h_{A_\theta}^{\mathrm{ID}}$ 和 $h_{A_\theta}^{C}$ 的复杂度分别为 $O(1)$ 和 $O(m)$. 假设每个管理节点收到 c 个普通节点的数据，则代理节点的计算复杂度为 $c(O(1)+O(m)+$

$O(E)$). 与 PKN 协议相同,管理节点的计算复杂度为 $O(cvm)$.

(2) 传感器节点通信能耗. 令 L_{iA} 表示从节点 s_i 到代理节点 A 的路由跳数,$L_{AWs_{\mathcal{M}}}$ 表示从代理节点 A 到 W 再到 $s_{\mathcal{M}}$ 的路由跳数,则 Naïve CPKN 协议中传感器节点的总能耗为

$$C_S=\sum_{i=1}^{N}(\ell_{id}+\ell_{BF}+\ell_d)\times(L_{iA}+L_{AWs_{\mathcal{M}}}).$$

(3) 管理节点通信能耗. 由于 Naïve CPKN 协议与 PKN 协议的定向存储和数据查询操作是完全相同的. 因此,两者的管理节点总能耗是相同的.

由以上分析可知,Naïve CPKN 协议能够提供高强度的隐私保护,但代理节点的计算能耗和通信能耗都比较高. 虽然引入代理节点轮转机制能够减轻能耗不均衡问题的影响,但仍然会消耗较多的网络能量. 因此 Naïve CPKN 协议仅适用于安全需求高,能量较为充足的网络中. 为进一步解决代理节点能耗问题,扩大协议的适用范围,我们提出了基于分片机制的 Slice-CPKN (CPKN-s)协议.

5.3 Slice-CPKN 协议

由于 Naïve CPKN 协议存在网络能耗较多和上传时延较长的问题,我们给出了 Slice-CPKN (CPKN-s)协议.

5.3.1 CPKN-s 协议实现

在 CPKN-s 中,与 Naïve-CPKN 不同,基站同时在网络中设置 ϕ 组功能完全相同的代理节点,用 $A_\theta(\theta=1,2,\cdots,w)$,$B_\theta(\theta=1,2,\cdots,w)$,$\cdots$,$\phi_\theta(\theta=1,2,\cdots,w)$表示. 其中每组中对应节点的数据处理函数是相同的,例如 A_1、B_1 和 ϕ_1 的功能相同,A_w、B_w 和 ϕ_w 的功能相同. 因此,同一个编码数据在经过代理节点$(A_1,B_2,B_3,\cdots,\phi_w)$处理后的值与经过$(A_1,A_2,A_3,\cdots,A_w)$处理后的值是完全相同的. 基站通过端到端加密仅向选定的代理节点 A_θ,B_θ和 ϕ_θ发送编码单路哈希函数 h_θ^C.

算法 2 给出了 CPKN-s 协议中基于编码切片的数据隐藏机制的详细步骤. 传感器节点 s_i 上传数据时,将加密后的数据 $E_{k_{i,t}}(V_{s_i})$加上 pID_i后直接发送到管理节点,然后将编码 BF_{s_i} 切分为 w 个编码片 $BF_{s_i}(z_\theta)(\theta=1,2,\cdots,w)$. w 个编码切片满足如下要求:$\|z_1\|>\|z_2\|>\cdots>\|z_w\|$,$\|z_w\|$ 表示第 w 个编码切片的长度. 将第 $\theta(\theta=1,2,\cdots,w)$个编码片加上 pID_i后发送到最近的第 θ 个代理节点上. 第 θ 个代理节点使用 h_θ^C将接收到的 BF_{s_i} 的第 θ 片段 $BF_{s_i}(z_\theta)$转化为 $BF'_{s_i}(z_\theta)$,并直接发送给管理节点.

管理节点在接收到来自 s_i的所有数据后,包括 pID_i,$BF'_{s_i}(z_\theta)(\theta=1,2,\cdots,w)$和 $E_{k_{i,t}}(V_{s_i})$,通过每条消息携带的 pID_i首先将 $BF'_{s_i}(z_\theta)(\theta=1,2,\cdots,w)$合并为整体编码 BF'_{s_i},z_θ编码片的长度信息表示了该编码片在整体编码 BF'_{s_i}中的顺序. 管理节点能够根据以上信息,合并出隐藏后的节点 s_i的完整信息 $\{t;pID_i;BF'_{s_i};E_{k_{i,t}}(V_{s_i})\}$.

基站在下发和更新 VLV 编码表时,也采用与传感器节点对应的方式将每个元素对应的编码切分为 w 个编码片,对每个编码片使用对应的 h_θ^C处理后合并为完整的元素编码,从而得到新的 VLV 子域 Bloom filter 编码,并发送到管理节点. 消息格式如下:

$$Sink\rightarrow s_{\mathcal{M}}:\{t;BF'_A,BF'_B,\cdots,BF'_{|VLVs-1|};\text{Hash}\}.$$

在 CPKN-s 协议中,管理节点采用与 PKN 和 CPKN-naive 协议完全相同的定向存储和数据查询处理机制来完成隐私保护 k-NN 查询.

算法 2. CPKN-s 协议代理节点数据隐藏机制.

输入:$\{t;pID_i;BF_{s_i};E_{k_{i,t}}(V_{s_i})\}$,$\{A_\theta,B_\theta,\cdots,\phi_\theta\,|\,(\theta=1,2,\cdots,w)\}$

输出:$\{t;pID_i;BF'_{s_i};E_{k_{i,t}}(V_{s_i})\}$

1. FOR $1\leq\theta\leq w$
2. $Sink\rightarrow A_\theta,B_\theta,\cdots,\phi_\theta:h_{A_\theta}^C$
   //基站为每组代理节点中对应的节点下发相同的哈希处理函数
3. END FOR
4. $s_i\rightarrow s_{\mathcal{M}}:\{pID_i,E_{k_{i,t}}(V_{s_i})\}$
5. s_i将 BF_{s_i} 切分为 w 个编码片,并将第 θ 个编码片 $BF_{s_i}(z_\theta)$发送到最近的第 θ 个代理节点
6. $A_\theta,B_\theta,\cdots,\phi_\theta:BF'_{s_i}(z_\theta)\leftarrow h_{A_\theta}^C(BF_{s_i}(z_\theta))$
7. $A_\theta,B_\theta,\cdots,\phi_\theta\rightarrow s_{\mathcal{M}}:BF'_{s_i}(z_\theta)$
   //代理节点将处理后的编码片发送到管理节点
8. $s_{\mathcal{M}}$:按照 $BF'_{s_i}(z_\theta)$长度从小到大排序
9. 合并 $BF'_{s_i}(z_\theta)$得到 BF'_{s_i}
10. RETURN $\{t;pID_i;BF'_{s_i};E_{k_{i,t}}(V_{s_i})\}$

//管理节点合并得到隐藏后的数据

5.3.2 CPKN-s 协议安全性分析

(1) 管理节点 $s_{\mathcal{M}}$与传感器节点 s_i的共谋攻击. 经过 w 个代理节点对传感器节点数据编码的处理后,管理节点虽然获得了传感器节点所使用的 Bloom filter 哈希函数,但是由于传感器节点的原始 BF_{s_i}已经被代理节点上的 w 个哈希函数转化为 BF'_{s_i},管理

节点很难从 BF'_{s_i} 逆向推测出原始的 BF_{s_i}. 设 l 为 s_i 节点上 Bloom filter 哈希函数的个数，假设 l_θ 个 1 落在第 $BF_{s_i}(z_\theta)$ 个编码片上，则攻击者能够正确地逆向推测原始 BF_{s_i} 的概率为 $P_i=\left(\prod_{\theta=1}^{w}\begin{pmatrix}\| z_\theta \| \\ l_\theta\end{pmatrix}\right)^{-1}$.

(2) 管理节点 $s_{\mathcal{M}}$、代理节点 A_θ 与传感器节点 s_i 的共谋攻击. 假设攻击者同时俘获了多个传感器节点、代理节点和管理节点，攻击者的目的是获得所有的 $h_\theta^C(\theta=1,2,\cdots,w)$ 函数. 因为即使仅缺少一个 h_θ^C 函数，攻击者也无法重复代理节点完整的数据处理操作，也就无法通过分析 BF'_{s_i} 得出原始的 BF_{s_i}，进而得到数据的真实值. 假设网络中传感器节点数目为 N，攻击者能够俘获的节点数为 Γ，代理节点共有 ϕ 组，每组包含 w 个节点，用 Γ_θ 表示攻击者通过俘获得到的 $h_\theta^C(\theta=1,2,\cdots,w)$ 函数的个数. 要获得所有的 h_θ^C 函数，则一定有 $1\leq\Gamma_\theta\leq\phi(\theta=1,2,\cdots,w)$. 因此，攻击者在俘获 Γ 个节点时，能够得到所有 h_θ^C 函数的概率 P_c 为

$$P_c=\begin{cases}0, & \Gamma<w\\ \sum\limits_{\substack{\forall\theta\in[1,w],\\ 1\leq\Gamma_\theta\leq\phi}}\left(\dfrac{\Gamma!}{\left(\prod\limits_{\theta=1}^{w}\Gamma_\theta!\right)\left(\Gamma-\sum\limits_{\theta=1}^{w}\Gamma_\theta\right)!}\times\right. & \\ \left.\left(1-\dfrac{\phi_w}{N}\right)^{\Gamma-\sum_{\theta=1}^{w}\Gamma_\theta}\left(\dfrac{\phi}{N}\right)^{\sum_{\theta=1}^{w}\Gamma_\theta}\right), & \Gamma\geq w\end{cases}.$$

由上式可以看出，P_c 与俘获节点的数目 Γ 以及代理节点的组数 ϕ 成正比，与每组代理节点的数目 w 成反比. 当 w 足够大，ϕ 取值合理时，攻击者很难获得所有的 h_θ^C 函数. 我们将在实验中对 CPKN-s 协议的隐私保护性能进行验证.

5.3.3　CPKN-s 协议能耗分析

(1) CPKN-s 协议计算复杂度. CPKN-s 协议中普通节点的计算复杂度仍为 $O(l)+O(E)$.

假设每个管理节点收到 c 个普通节点的数据，则每个代理节点接收到的 BF 编码片数量平均为 (c/ϕ)，BF 编码片的长度平均为 (m/w)，代理节点上 $h_{A_\theta}^C$ 的复杂度为 $O(m/w)$. 因此代理节点的计算复杂度为 $(c/\phi)O(m/w)$，明显低于 Naïve CPKN 协议. 而管理节点的计算复杂度仍为 $O(cvm)$.

(2) 传感器节点总能耗. 令 $L_{i\theta_{s_{\mathcal{M}}}}$ 表示从节点 s_i 到代理节点 θ 再到 $s_{\mathcal{M}}$ 的平均路由跳数，w 表示编码切片个数，则 CPKN-s 协议中传感器节点的总能耗为

$$C_S=\sum_{i=1}^{N}((\ell_{\mathrm{id}}+\ell_d)\times L_{is_{\mathcal{M}}}+(w\times\ell_{\mathrm{id}}+\ell_{\mathrm{BF}})\times L_{i\theta_{s_{\mathcal{M}}}}).$$

(3) 管理节点通信能耗. CPKN-s 协议的管理节点总能耗与 PKN 协议以及 Naïve CPKN 协议是相同的.

5.3.4　CPKN-s 协议能耗均衡优化机制

面向能耗均衡的簇头节点选取机制在已有工作中已经得了深入的研究，如文献[37-40]. 在 CPKN-s 协议中，多组代理节点不再需要对数据进行加密操作，而是只对接收到的部分编码片进行哈希操作，计算能耗明显降低，能耗不均衡问题也较 Naïve CPKN 有所缓解. 为了更好的解决 CPKN-s 协议中的代理节点选取和能耗均衡问题，我们引入现有研究中比较成熟的算法[37]，文献[37]中提出的多簇头产生机制适用于 CPKN-s 中多代理节点的选择，并且能够达到很好的网络能耗均衡效果.

6　实验与分析

我们首先对本文核心的隐私保护 Bloom filter 编码机制的性能进行实验验证，并讨论相应的参数选择问题；接着按照非共谋攻击和共谋攻击两种场景，分别对本文提出的隐私保护 k-NN 查询协议的关键性能进行验证，包括网络性能和隐私性能两个方面. 网络性能主要衡量 PKN 和 CPKN 协议的通信能耗以及查询时延两个重要参数；隐私性能主要关注 Naïve-CPKN 和 CPKN-s 协议在共谋攻击下的隐私保护强度. 由于现有研究大多针对基于位置的 k-NN 查询展开，缺少传感器网络中隐私保护基于值的 k-NN 查询方面的工作，实验中选用经典的不提供数据隐私保护的分布式 TinyOS-kNN[41] 算法作为对比协议.

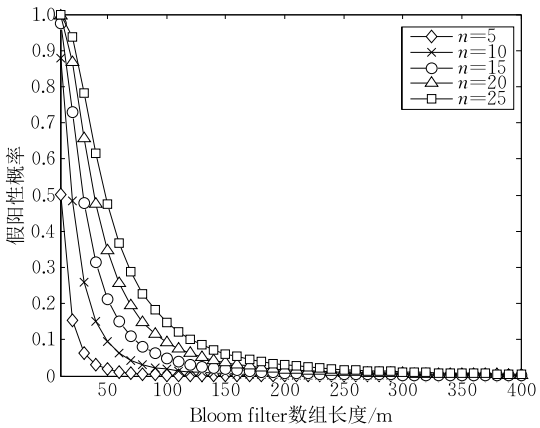
我们使用 OMNet++4.1 来对协议的性能进行仿真. 传感器节点数目 N 从 100～1000 变化，网络覆盖范围为 $100\sqrt{N}\times100\sqrt{N}\,\mathrm{m}^2$ 的矩形区域，基站位于区域中心位置. 文献[42-43]的研究表明，在双层传感网中，管理节点数量的增加通常有利于延长普通节点的生命周期，但也会明显地增加网络的成本. 考虑到实际应用中网络的实用性和可扩展性，与文献[43]相同，实验中管理节点的数量随网络规模的增大而增加，设每个管理节点对应 c 个普通节点(本文中 c 约为 50)，因此管理节点的数目

从 2 个逐渐增加到 20 个，并均匀分布在网络中. 与文献[14-15]相同，设普通节点到管理节点的平均跳数 $L_{is_{\mathcal{M}}} \approx \sqrt{c}/2+1$.

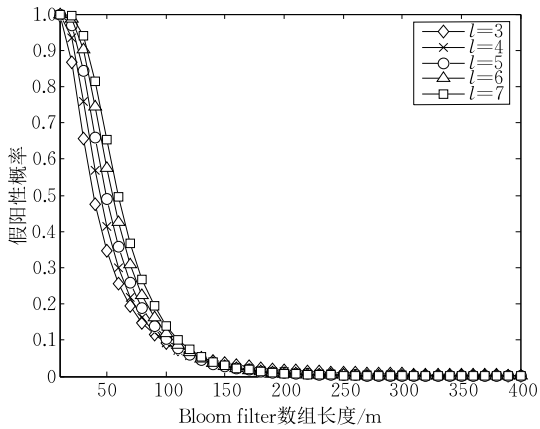
实验使用公开的真实数据集 LUCE dataset①，该数据集为 2006 年 11 月至 2007 年 5 月采集的环境温度、土壤湿度等多维属性数据，共包含约 100 个节点. 实验中不同规模网络的传感器节点数据通过合成 LUCE dataset 得到.

6.1 隐私保护 Bloom filter 编码机制的性能

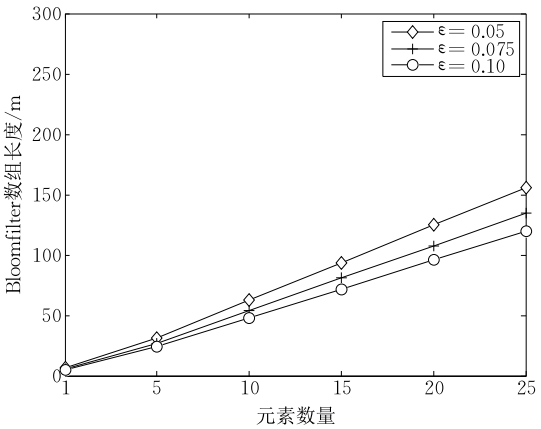
假阳性判定概率是影响 Bloom filter 编码机制性能的关键点，在进行 k-NN 查询时，过高的假阳性判定概率会引发严重的能量和存储空间浪费问题. 本节对 Bloom filter 的假阳性判定概率与编码数组长度以及哈希函数个数之间的关系进行了评估. 如图 6(a)所示，当哈希函数取 4 个时，Bloom filter 的假阳性判定概率随着编码长度的增加快速下降. 在传感器网络当中，假阳性判定概率一般不应超过 5%，因此对于包含 10 个、15 个和 20 个元素的集合，适合的编码数值长度分别为 70 位、95 位和 125 位. 图 6(b)给出了在元素个数为 20 时，哈希函数个数对假阳性判定概率的影响. 可以看到，哈希函数个数越多时，对编码长度的要求越高，同时假阳性判定概率的下降速度也越快. 图 6(c)给出了指定假阳性判定概率下，元素数量与编码数组长度之间的关系. 可以看到，元素数量越多所需要的编码数组长度也就越长. 因此，在划分 VLV 时，子域元素的数量一般不超过 20. 本文中为 Bloom filter 机制选择 4 个哈希函数和 128 位的编码数组.



(a) 假阳性判定概率与编码长度



(b) 假阳性判定概率与哈希函数个数



(c) 元素数量与数组长度

图 6 Bloom filter 编码机制性能

6.2 隐私保护 k-NN 查询协议性能(PKN)

本节中首先对 PKN 协议所采用的数据压缩机制的性能进行验证，接着使用通信能耗来对 PKN 协议的网络性能进行评价，最后评估了非共谋攻击下 PKN 协议的隐私保护性能.

6.2.1 PKN 协议数据压缩机制性能评价

数据压缩机制能够明显地降低普通传感器节点

① LUCE. http://lcav.epfl.ch/cms/lang/en/pid/86035

的通信量，图 7 比较了两种数据压缩机制的压缩性能. 压缩后的位数越少，则通信能耗越低. 设初始 BF 编码为 128 位，使用绝对位置数据压缩机制后，编码被压缩到约 24 位，平均压缩率约为 5；而使用相对位置压缩机制后，编码被压缩到约 18 位，平均压缩率约为 7. 可以看到，相对位置数据压缩机制能够更好地减少普通节点的通信量，在 PKN 中将采用相对位置数据压缩机制.

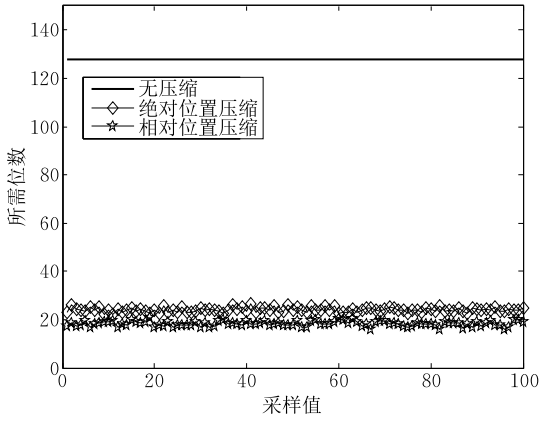


图 7　编码数据压缩机制性能

6.2.2　PKN 协议网络性能评价

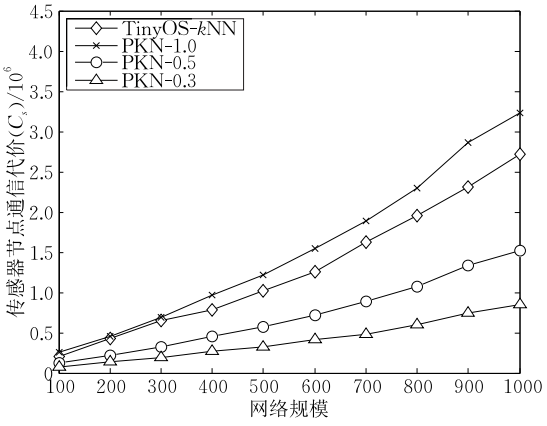
在传感器网络中，绝大部分的能量消耗是由数据通信引起的. 因此，我们使用通信代价来衡量网络的能耗性能. 对于 PKN 协议，VLV 划分表对值域的覆盖率是决定能耗高低的重要因素，我们分别考虑了 VLV 覆盖率为 100%(PKN-1.0)、50%(PKN-0.5)和 30%(PKN-0.3)的情况. 图 8 和图 9 分别给出了普通传感器节点和管理节点在不同场景下的通信代价.

如图 8(a)所示，当网络规模从 100 增加到 1000 时，能耗随着网络规模的扩大而增加. 当 PKN 的 VLV 表覆盖整个值域时，PKN-1.0 的能耗高于 TinyOS，这是因为 PKN 协议为每个数据增加了一个编码数组. 而 PKN-0.5 和 PKN-0.3 由于在传感器节点进行了过滤操作，减少了数据发送量，能耗较低. 由实验结果可以看出，实际应用中在不影响结果精度的前提下基站选择覆盖率尽可能小的 VLV 表是有必要的.

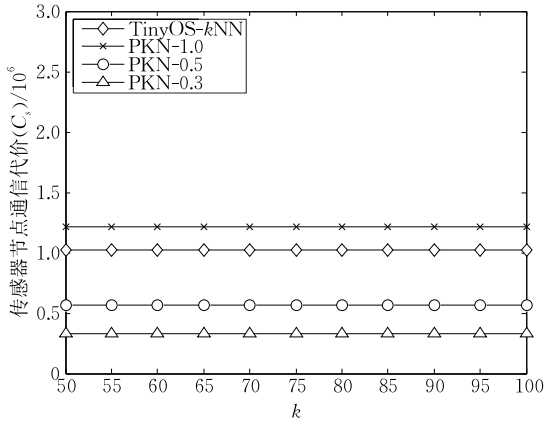
图 8(b)给出了 k-NN 查询中参数 k 的变化与普通传感器节点能耗的关系. 在 500 个节点的网络中，k 值由 50 增大到 100. 由于数据查询过程仅仅与上层管理节点有关，因此各协议的普通传感器节点能耗不随 k 值的变化而发生变化.

图 9(a)对 PKN 协议中管理节点能耗与网络规模的关系进行了衡量. 考虑到查询的合理性，我们将 k 值设置为网络规模的 10%. 可以看出 PKN-0.5 的管理节点总能耗略高于 TinyOS，而 PKN-0.3 的管理节点能耗相对较低. TinyOS 中数据分布存储在多个管理节点，查询响应时的多跳通信增加了总体的能耗.

图 9(b)中，在 500 个节点的网络中，当 k 值由 50 增大到 100 时，PKN-0.5 和 PKN-0.3 的管理节点能耗增长幅度不大，而 TinyOS 的能耗有较明显的增加. 这是因为在 PKN 协议中，k 值的增大对定向存储的能耗没有影响，仅仅增加了结果返回值的规模，而结果数据存储在近基站的节点，数据返回所使用的能耗较低.

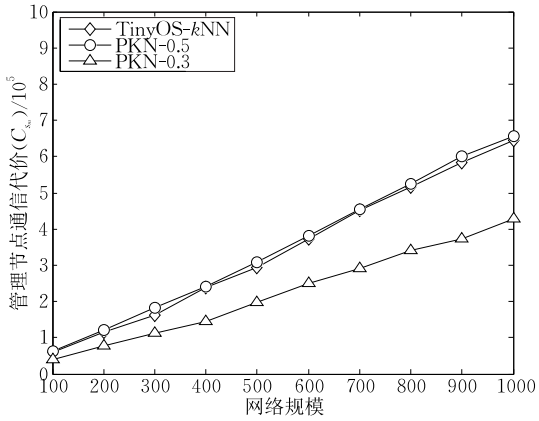


(a) 普通传感器节点能耗与网络规模

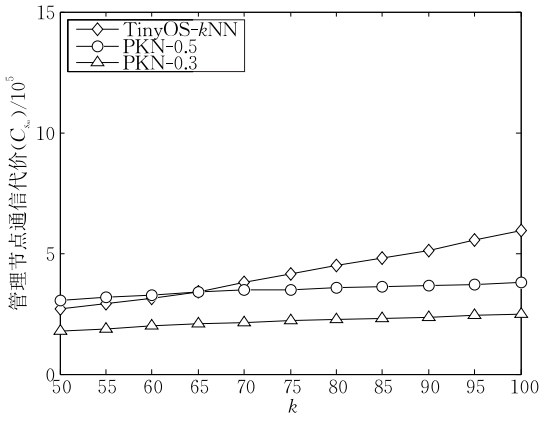


(b) 普通传感器节点能耗与参数k

图 8　PKN 协议普通传感器节点能耗性能



(a) 管理节点能耗与网络规模



(b) 管理节点能耗与参数k

图 9 PKN 协议管理节点能耗性能

6.2.3 PKN 协议隐私性能评价

图 10 给出了 PKN 协议在不同 Bloom filter 哈希函数数目下的隐私泄露概率,即攻击者通过推测攻击正确关联数据编码与距离真实值的概率. 仅当哈希函数数目为 1 时,有不足 0.01 的隐私泄露概率,当哈希函数数目大于等于 3 时,隐私泄露概率接近于 0,攻击者无法获得隐私数据信息. 本文中为 Bloom filter 选择 4 个哈希函数,因此能够保证数据的隐私.

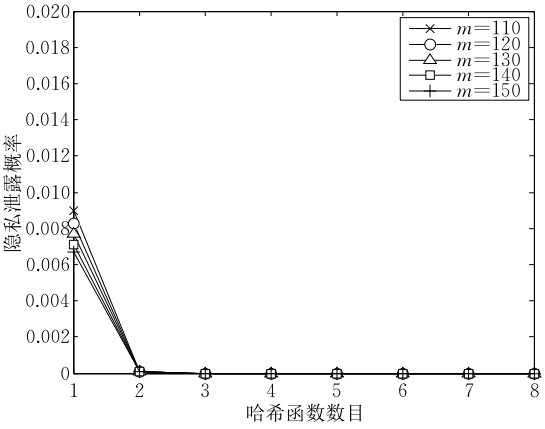


图 10 PKN 协议隐私性能

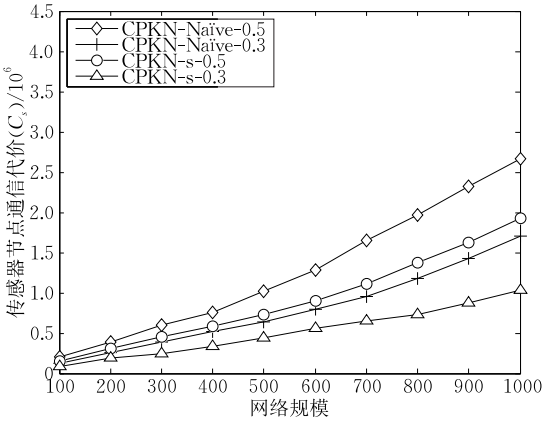
6.3 共谋攻击下的 k-NN 查询协议性能(CPKN)

本节主要关注 CPKN 协议在共谋攻击下的能耗性能和隐私性能.

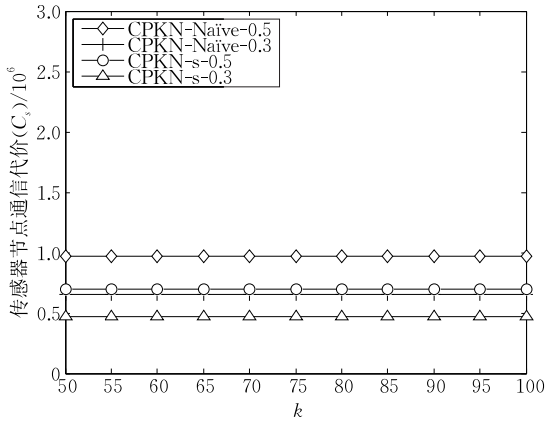
6.3.1 CPKN 协议能耗性能评价

图 11(a)给出了 CPKN-Naïve 协议和 CPKN-s 协议普通传感器节点的能耗. 可以看出,CPKN-Naïve 协议中,代理节点由于采用了串行的数据处理方式,增加了数据上传的平均跳数,能耗明显大于 CPKN-s 协议. 而 CPKN-s 协议中代理节点仅对 BF_{s_i} 进行并行处理的方法有效地降低了传感器节点的能耗.

图 11(b)给出了 CPKN 协议中普通传感器节点能耗与参数 k 的变化的关系. 与 PKN 协议类似,CPKN-Naïve 和 CPKN-s 协议的传感器节点能耗不随 k 值的变化而发生变化.

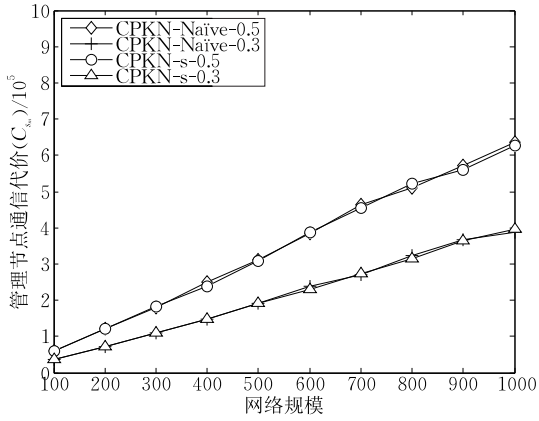


(a) 普通传感器节点能耗与网络规模

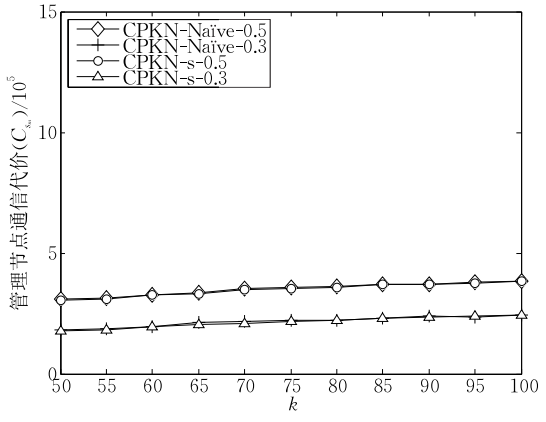


(b) 普通传感器节点能耗与参数k

图 11 CPKN 协议普通传感器节点能耗性能



(a) 管理节点能耗与网络规模



(b) 管理节点能耗与参数k

图 12　CPKN 协议管理节点能耗性能

如图 12(a)所示，CPKN-Naïve 协议和 CPKN-s 协议的管理节点能耗在不同网络规模下是基本相等的，这是因为两个协议在管理节点上的定向存储机制和数据查询机制都是相同的，仅仅在 VLV 子域判定过程中存在细微差别. 数据上传阶段中代理节点的不同数据处理方式对管理节点不会产生影响.

与图 12(a)类似，图 12(b)中 CPKN-Naïve 协议和 CPKN-s 协议的管理节点能耗在不同 k 值条件下也是基本相等的.

6.3.2　CPKN 协议隐私性能评价

对 CPKN 协议隐私性能的评价分为两种情况：一是普通传感器节点与管理节点间的共谋攻击；二是普通传感器节点、代理节点与管理节点间的共谋攻击. 图 13 给出了攻击者通过存储在管理节点上的 BF'_{s_i} 来推测得到正确的原始 BF_{s_i} 的概率. 可以明显地看出，当 Bloom filter 哈希函数大于等于 3 时，CPKN-Naïve 协议和 CPKN-s 协议的隐私泄露概率都趋近于 0，攻击者无法在普通传感器节点与管理节点共谋的情况获得敏感数据信息.

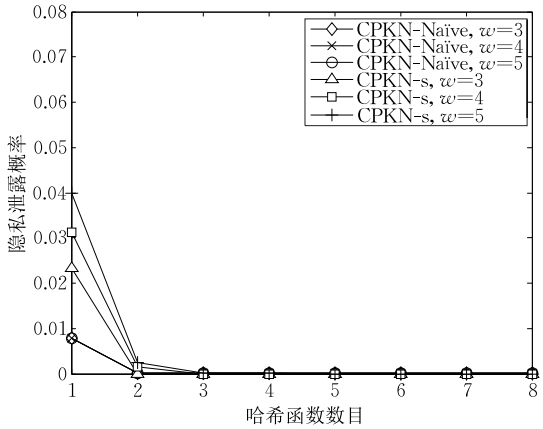
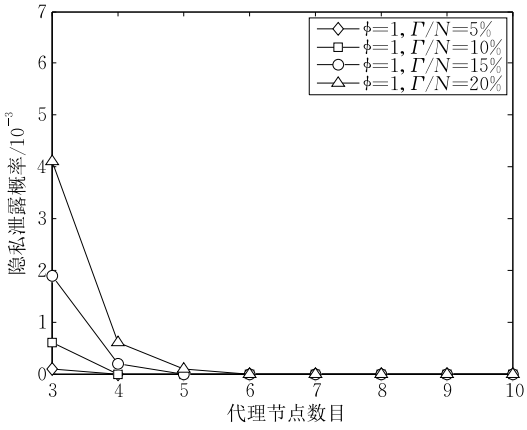
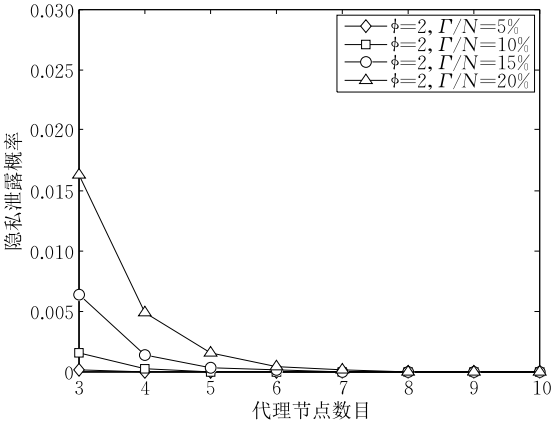


图 13　s_i 与 $s_{\mathcal{M}}$ 共谋时 CPKN 协议隐私性能

攻击者为获得隐私数据，将尝试在俘获普通传感器节点和管理节点的同时继续俘获代理节点. 然而，攻击者必须在俘获所有代理的节点的前提下才能获得所有的编码处理函数 $h_\theta^C(\theta=1,2,\cdots,w)$，进而获取真实数据值. 图 14(a)给出了当 $\phi=1$ 时，攻击者获得所有 h_θ^C 函数的概率. 当网络中存在 4 个代理节点时，攻击者即使俘获整个网络中 20% 的节



(a) ϕ=1的隐私性能



(b) ϕ=2的隐私性能

图 14　s_i，A_θ 与 $s_{\mathcal{M}}$ 共谋时 CPKN 协议隐私性能

点，能够获得所有 h_θ^C 函数的概率依旧小于 0.001. 如图 14(b)所示，当 $\phi=2$ 时，隐私泄露的概率有所增加，这是因为同时存在的两组功能相同的代理节点增大了攻击者获得所有 h_θ^C 函数的概率. 但是，当每组中的代理节点数目大于等于 4 个时，隐私泄露的概率仍然不超过 0.005. 使用 $\phi=2, w=4$ 的 CPKN-s 协议既可以保证较低的网络能耗，又能够保证高强度的隐私保护.

由以上实验结果可知，在管理节点俘获攻击和节点共谋攻击两种情况下，PKN 协议和 CPKN 协议能够在提供高强度隐私保护的同时高效率低能耗地精确完成 k-NN 查询处理.

7 总 结

本文针对双层无线传感器网络中的隐私保护 k-NN 查询处理，分别提出了面向管理节点俘获攻击的 PKN 协议和面向共谋攻击的 CPKN 协议. 在 PKN 协议中，我们首先基于定向存储机制提出了适用于双层传感网的 KNQ 查询架构，接着给出了一种隐私保护数据编码机制，通过为真实数据附加编码的方式，保证在不泄露数据隐私的同时高效的完成查询. 在 CPKN 协议中，针对节点共谋攻击，设计了基于一系列代理节点的数据隐藏机制，通过破坏普通节点编码与管理节点编码间的关联性达到抵御共谋攻击的目的. 理论分析和实验结果表明，PKN 和 CPKN 协议能够在提供高强度隐私保护的同时保证较高的能耗性能和查询时延性能.

致 谢 本文审稿专家和编辑老师提出了宝贵的意见和建议，在此表示感谢！本文部分工作是在作者访问中国人民大学的萨师煊大数据管理和分析中心时完成的，该中心获国家高等学校学科创新引智计划(111 计划)等资助！

参 考 文 献

[1] Ren Feng-Yuan, Huang Hai-Ning, Lin Chuang. Wireless sensor networks. Journal of Software, 2003, 14(7): 1282-1291(in Chinese)
(任丰原，黄海宁，林闯. 无线传感器网络. 软件学报，2003，14(7)：1282-1291)

[2] Atzori L, Iera A, Morabito G. The Internet of Things: A survey. Computer Networks, 2010, 54(15): 2787-2805

[3] Fan Yong-Jian, Chen Hong, Zhang Xiao-Ying. Data privacy preservation in wireless sensor networks. Chinese Journal of Computers, 2012, 35(6): 1131-1146(in Chinese)
(范永健，陈红，张晓莹. 无线传感器网络数据隐私保护技术. 计算机学报，2012，35(6)：1131-1146)

[4] He W, Liu X, Nguyen H, et al. PDA: Privacy-preserving data aggregation in wireless sensor networks//Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM). Anchorage, USA, 2007: 2045-2053

[5] He W, Liu X, Nguyen H V, et al. PDA: Privacy-preserving data aggregation for information collection. ACM Transactions on Sensor Networks, 2011, 8(1): 6

[6] Yao Y, Liu J, Xiong N N. Privacy-preserving data aggregation in two-tiered wireless sensor networks with mobile nodes. Sensors, 2014, 14(11): 21174-21194

[7] Yang Geng, Li Sen, Chen Zheng-Yu, et al. High-accuracy and privacy-preserving oriented data aggregation algorithm in sensor networks. Chinese Journal of Computers, 2013, 36(1): 189-200(in Chinese)
(杨庚，李森，陈正宇等. 传感器网络中面向隐私保护的高精确度数据融合算法. 计算机学报，2013，36(1)：189-200)

[8] Wang Tao-Chun, Qin Xiao-Lin, Liu Liang, Ding You-Wei. Secure and energy-efficient spatial data aggregation algorithm in wireless sensor networks. Journal of Software, 2014, 25(8): 1671-1684(in Chinese)
(王涛春，秦小麟，刘亮，丁有伟. 无线传感器网络中安全高效的空间数据聚集算法. 软件学报，2014，25(8)：1671-1684)

[9] Chen C M, Lin Y H, Lin Y C, et al. RCDA: Recoverable concealed data aggregation for data integrity in wireless sensor networks. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(4): 727-734

[10] Ozdemir S, Xiao Y. Integrity protecting hierarchical concealed data aggregation for wireless sensor networks. Computer Networks, 2011, 55(8): 1735-1746

[11] Sicari S, Grieco L A, Boggia G, et al. DyDAP: A dynamic data aggregation scheme for privacy aware wireless sensor networks. Journal of Systems and Software, 2012, 85(1): 152-166

[12] Sheng B, Li Q. Verifiable privacy-preserving range query in two-tiered sensor networks//Proceedings of the 27th IEEE Conference on Computer Communications (INFOCOM). Phoenix, USA, 2008: 46-50

[13] Sheng B, Li Q. Verifiable privacy-preserving sensor network storage for range query. IEEE Transactions on Mobile Computing, 2011, 10(9): 1312-1326

[14] Shi J, Zhang R, Zhang Y. A spatiotemporal approach for secure range queries in tiered sensor networks. IEEE Transactions on Wireless Communications, 2011, 10(1): 264-273

[15] Zhang R, Shi J, Zhang Y. Secure multidimensional range queries in sensor networks//Proceedings of the 10th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc). New Orleans, USA, 2009: 197-206

[16] Li Rui, Lin Ya-Ping, Yi Ye-Qing, Hu Yu-Peng. A privacy and integrity preserving range query protocol in two-tiered sensor networks. Chinese Journal of Computers, 2013, 36(6): 1194-1209(in Chinese)
(李睿，林亚平，易叶青，胡玉鹏. 两层传感器网络中隐私与完整性保护的范围查询协议. 计算机学报，2013，36(6)：1194-1209)

[17] Chen F, Liu A X. SafeQ: Secure and efficient query processing in sensor networks//Proceedings of the 29th IEEE International Conference on Computer Communications (INFOCOM). San Diego, USA, 2010: 2642-2650

[18] Chen F, Liu A X. Privacy-and integrity-preserving range queries in sensor networks. IEEE/ACM Transactions on Networking, 2012, 20(6): 1774-1787

[19] Dou Yi, Huang Hai-Ping, Wang Rui-Chuan, et al. Secure range query in two-tiered wireless sensor networks. Journal of Computer Research and Development, 2013, 50(6): 1253-1266(in Chinese)
(窦轶，黄海平，王汝传等. 两层无线传感器网络安全范围查询协议. 计算机研究与发展，2013，50(6)：1253-1266)

[20] Tsou Y T, Lu C S, Kuo S Y. Privacy-and integrity-preserving range query in wireless sensor networks// Proceedings of the IEEE Global Communications Conference (GLOBECOM). Anaheim, USA, 2012: 328-334

[21] Yi Y, Li R, Chen F, et al. A digital watermarking approach to secure and precise range query processing in sensor networks// Proceedings of the 32nd IEEE International Conference on Computer Communications (INFOCOM). Turin, Italy, 2013: 1950-1958

[22] Fan Yong-Jian, Chen Hong. Verifiable privacy-preserving Top-k query protocol in two-tiered sensor networks. Chinese Journal of Computers, 2012, 35(3): 423-433(in Chinese)
(范永健，陈红. 两层传感器网络中可验证隐私保护 Top-k 查询协议. 计算机学报，2012，35(3)：423-433)

[23] Chen Wei, Xu Ruo-Mei, Li Yu-Ling. A privacy-preserving integrity-verification-based Top-k query processing. Journal of Computer Research and Development, 2014, 51(12): 2585-2592(in Chinese)
(陈伟，许若妹，李玉岭. 基于隐私保护和完整性验证的 Top-k 查询方法. 计算机研究与发展，2014，51(12)：2585-2592)

[24] Liao X, Li J. Privacy-preserving and secure Top-k query in two-tier wireless sensor network//Proceedings of the IEEE Global Communications Conference (GLOBECOM). Anaheim, USA, 2012: 335-341

[25] Zhou T, Lin Y, Zhang W, et al. Secure and verifiable Top-k query in two-tiered sensor networks. Security and Privacy in Communication Networks, 2013, 127(1): 19-34

[26] Dai Hua, Yang Geng, Qin Xiao-Lin, et al. Privacy-preserving Top-k query processing in two-tiered wireless sensor networks. Journal of Computer Research and Development, 2013, 50(6): 1239-1252(in Chinese)
(戴华，杨庚，秦小麟等. 面向隐私保护的两层传感网 Top-k 查询处理方法. 计算机研究与发展，2013，50(6)：1239-1252)

[27] Gnawali O, Jang K-Y, Paek J, et al. The tenet architecture for tiered sensor networks//Proceedings of the 4th ACM International Conference on Embedded Networked Sensor Systems (SenSys). Boulder, USA, 2006: 153-166

[28] Vaidya J, Clifton C W. Privacy-preserving kth element score over vertically partitioned data. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(2): 253-258

[29] Hu H, Xu J, Ren C, Choi B. Processing private queries over untrusted data cloud through privacy homomorphism// Proceedings of the 27th IEEE International Conference on Data Engineering (ICDE). Hannover, Germany, 2011: 601-612

[30] Wong W K, Cheung D W, Kao B, Mamoulis N. Secure kNN computation on encrypted databases//Proceedings of the ACMSIGMOD International Conference on Management of Data (SIGMOD). Rhode Island, USA, 2009: 139-152

[31] Zhu Y, Xu R, Takagi T. Secure k-NN computation on encrypted cloud data without sharing key with query users// Proceedings of the ACM International Workshop on Security in Cloud Computing. Hangzhou, China, 2013: 55-60

[32] Yao B, Li F, Xiao X. Secure nearest neighbor revisited// Proceedings of the 29th IEEE International Conference on Data Engineering (ICDE). Brisbane, Australia, 2013: 733-744

[33] Shaneck M, Kim Y, Kumar V. Privacy preserving nearest neighbor search. Machine Learning in Cyber Trust. USA: Springer, 2009

[34] Qi Y, Atallah M J. Efficient privacy-preserving k-nearest neighbor search//Proceedings of 28th IEEE International Conference on Distributed Computing Systems (ICDCS). Beijing, China, 2008: 311-319

[35] Groat M M, He W, Forrest S. KIPDA: k-indistinguishable privacy-preserving data aggregation in wireless sensor networks//Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM). Shanghai, China, 2011: 2024-2032

[36] Bloom B H. Space/time trade-offs in hash coding with allowable errors. Communications of the ACM, 1970, 13(7): 422-426

[37] Cai Hai-Bin, Ju Xiao-Ming, Cao Qi-Ying, et al. Energy prediction and reliable clustering routing protocol for multilevel energy heterogeneous wireless sensor networks. Chinese Journal of Computers, 2009, 32(12): 2393-2402(in Chinese)
(蔡海滨，琚小明，曹奇英等. 多级能量异构无线传感器网络的能量预测和可靠聚簇路由协议. 计算机学报，2009，32(12)：2393-2402)

[38] Gupta G, Younis M. Load-balanced clustering of wireless sensor networks//Proceedings of the 2003 IEEE International Conference on Communications (ICC). Anchorage, USA, 2003: 1848-1852

[39] Ye M, Li C, Chen G, et al. EECS: An energy efficient clustering scheme in wireless sensor networks//Proceedings of the 24th IEEE International Performance, Computing, and Communications Conference(IPCCC). Phoenix, USA, 2005: 535-540

[40] Abbasi A A, Younis M. A survey on clustering algorithms for wireless sensor networks. Computer Communications, 2007, 30(14): 2826-2841

[41] Levis P, Madden S, Polastre J, et al. TinyOS: An operating system for sensor networks//Weber W, Rabaey J M, Aarts E eds. Ambient Intelligence. Berlin Heidelberg, Germany: Springer, 2005

[42] Hou Y T, Shi Y, Sherali H D, et al. On energy provisioning and relay node placement for wireless sensor networks. IEEE Transactions on Wireless Communications, 2005, 4(5): 2579-2590

[43] Bari A, Wazed S, Jaekel A, et al. A genetic algorithm based approach for energy efficient routing in two-tiered sensor networks. Ad Hoc Networks, 2009, 7(4): 665-676

PENG Hui, born in 1986, Ph. D. candidate. His research interests include wireless sensor network, privacy preservation and data management of IoTs.

CHEN Hong, born in 1965, Ph. D., professor, Ph. D. supervisor. Her research interests include database, data warehouse and wireless sensor network.

ZHANG Xiao-Ying, born in 1987, Ph. D. Her research interests include wireless sensor network and privacy preservation.

ZENG Ju-Ru, born in 1987, Ph. D. candidate. His research interests include participatory sensing and privacy preservation.

WU Yun-Cheng, born in 1989, Ph. D. candidate. His research interests include data management of IoTs and privacy preservation.

WANG Shan, born in 1944, professor, Ph. D. supervisor. Her research interests include the new technology of high-performance database, data warehouse and data analysis.

Background

Privacy preservation in Wireless Sensor Networks (WSN) has become a hot spot of research. At present, the studies mainly focus on privacy-preserving data aggregation, range query and Top-k query. Research on privacy-preserving k-NN query in WSN is still rare. As the execution procedure of k-NN query is complex, answering accurately k-NN query in WSNs while preserving data privacy is extremely challenging.

This paper presents a privacy-preserving k-NN query protocol for two-tiered sensor networks. Our proposals can complete accurately k-NN query while preventing data leakage evoked by management node compromising attack and node collusion attack. Theoretical analysis and experimental results of real dataset confirm the privacy and efficiency of our proposals.

This work is supported by the National Basic Research Program (973 Program) of China (Nos. 2014CB340402, 2012CB316205), the National High Technology Research and Development Program (863 Program) of China (No. 2014AA015204), the National Natural Science Foundation of China (Grant Nos. 61532021, 61272137, 61202114).