两层传感器网络中可验证隐私保护 Top-k 查询协议

范永健[1),2),3)] 陈 红[1),2)]

[1)](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)

[2)](中国人民大学信息学院 北京 100872)

[3)](河北工程大学信息与电气工程学院 河北 邯郸 056038)

摘 要 无线传感器网络中隐私保护技术已经成为研究热点,其中隐私保护精确 Top-k 查询协议已成为富有挑战性的研究问题.文中提出了一种两层传感器网络中可验证隐私保护 Top-k 查询协议 SafeTQ(Safe Top-k Query),SafeTQ 由隐私保护 Top-k 查询协议和两种完整性验证模式组成.SafeTQ 使用加随机数扰乱、加密和高资源节点之间安全计算第 k 位数据值等策略,能够在不泄漏隐私信息的情况下,精确地完成传感器网络 Top-k 查询.SafeTQ 中两种完整性验证模式分别使用邻居数据项形成加密链和空间邻居节点概率发送验证消息策略,使 Sink 能够检测和拒绝不正确或不完整查询响应.文中通过理论分析和使用真实数据集实验验证了 SafeTQ 的安全性和有效性.

关键词 物联网;无线传感器网络;Top-k 查询;隐私保护;完整性验证;安全计算

中图法分类号 TP393 **DOI 号**: 10.3724/SP.J.1016.2012.00423

Verifiable Privacy-Preserving Top-k Query Protocol in Two-tiered Sensor Networks

FAN Yong-Jian[1),2),3)] CHEN Hong[1),2)]

[1)](*Key Laboratory of Data Engineering and Knowledge Engineering (Remin University of China) of Ministry of Education, Beijing* 100872)

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

[3)](*School of Information and Electrical Engineering, Hebei University of Engineering, Handan, Hebei* 056038)

Abstract Privacy preservation in wireless sensor networks has attracted more and more attentions. Answering accurately Top-k query in wireless sensor networks while preserving data privacy is a challenge. This paper presents a verifiable privacy-preserving Top-k query protocol in two-tiered sensor networks (SafeTQ), which consists of privacy-preserving Top-k query protocol and two integrity verification schemes. SafeTQ can complete accurately Top-k query in two-tiered sensor networks while preventing attackers from gaining collected data. To preserve privacy, SafeTQ uses some strategies such as adding random numbers, encryption and securely computing kth data value between high resource nodes. To verify integrity, two integrity verification schemes use respectively encrypted data chains and check message provided by spatial neighborhoods. Theoretical analysis and simulation results by using real-world data confirm the high efficacy and efficiency of SafeTQ.

Keywords Internet of Things; wireless sensor network; Top-k query; privacy preservation; integrity verification; secure computation

1 引 言

无线传感器网络(以下简称为传感器网络)作为物联网的重要组成部分,在环境监测、医疗卫生、智能交通、国防军事等领域具有广阔的应用前景.随着传感器网络应用发展,在实际应用部署过程中暴露出严重的隐私泄漏威胁.在野生动物监测应用中,珍

收稿日期:2011-08-24;最终修改稿收到日期:2012-01-02.本课题得到国家自然科学基金(61070056,61075053)资助.**范永健**,男,1978 年生,博士研究生,讲师,中国计算机学会(CCF)会员,主要研究方向为无线传感器网络、隐私保护、数据库.E-mail: fanyj_ruc@ruc.edu.cn.
陈 红(通信作者),女,1965 年生,博士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为数据库、数据仓库、无线传感器网络.E-mail: chong@ruc.edu.cn.

贵野生动物的位置信息可能被用于非法捕猎;在医疗应用领域,患者的体征敏感数据可能被泄漏;在车载传感网应用中,车载实体的位置和运动轨迹等敏感信息可能被泄漏;在军事应用领域,重要数据、基站或事件源位置等敏感信息泄漏可能造成严重后果.传感器网络中这些严重的隐私泄漏威胁,极大地影响了传感器网络的应用发展.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.

传感器网络中数据隐私保护技术已经成为研究热点,现有研究内容主要集中在 SUM、MIN、MAX 或 Median 等聚集操作中隐私保护技术[1-3]和范围查询隐私保护和验证技术[4-6].文献[7]研究了 Top-k 查询完整性验证技术,但该文献没有考虑数据隐私问题.Top-k 查询是传感器网络中重要的查询类型.由于传感器网络可验证隐私保护精确 Top-k 查询协议面临诸多挑战.据作者所知,目前相关成果基本还是空白.

传感器网络 Top-k 查询隐私保护和验证技术面临的挑战主要表现在以下方面:(1) Top-k 查询执行需要查询区域的全局数据信息,并且执行过程中不能泄漏数据值等敏感信息;(2) 应同时保证 Top-k 查询的正确性和精确性;(3) 传感器节点资源严重受限,应减少传感器节点能量消耗,尽量利用高资源节点(两层传感器网络中);(4) 将 Top-k 查询隐私保护和验证技术结合,同时保证数据的隐私性和完整性.

本文研究两层传感器网络中 Top-k 查询的安全性问题.两层传感器网络具有寿命长等优点,已成为研究热点[4-7].在两层传感器网络中,由于单元头节点需要聚集数据和执行查询,往往成为被攻击的对象.单元头节点被俘获后常进行破坏隐私性和破坏完整性两种模式的攻击.本文在单元头节点不安全的情况下,为应对破坏隐私性攻击威胁,采用加随机数扰乱、加密和高资源节点之间进行安全计算等策略,能够在不泄漏隐私信息的情况下,精确地完成传感器网络 Top-k 查询;为应对破坏完整性攻击威胁,采用邻居数据项形成加密链和空间邻居节点概率发送验证消息策略,能够使 Sink 检测和拒绝不正确或不完整查询响应.

本文主要贡献有:(1) 提出两层传感器网络中隐私保护 Top-k 查询协议,首次实现了在不泄漏隐私信息的情况下,精确地完成传感器网络 Top-k 查询;(2) 提出两种新颖的两层传感器网络中 Top-k 查询完整性验证模式,能够使 Sink 检测和拒绝不正确或不完整查询响应;(3) 将隐私保护 Top-k 查询协议与完整性验证模式结合,形成两层传感器网络中可验证隐私保护 Top-k 查询协议 SafeTQ(Safe Top-k Query);(4) 通过理论分析和使用真实数据集仿真实验,分析和验证了 SafeTQ 的安全性和有效性.

本文第 2 节介绍相关研究工作;第 3 节给出本文研究所基于的模型和安全设计目标;第 4 节对 SafeTQ 进行形式化描述;第 5 节给出安全性分析和证明;第 6 节使用公开真实数据集进行仿真实验证明了 SafeTQ 的有效性;第 7 节进行总结.

2 相关工作

(1) 隐私保护技术

目前,隐私保护技术研究成果主要集中在数据挖掘和数据发布两个领域[8].综合现有的研究成果,隐私保护技术主要可分为三类:数据扰乱技术、数据加密技术和数据匿名化技术.数据扰乱技术使用加随机数、交换等技术对原始数据进行扰动,但需要保证处理后数据能够满足相关应用需求;在数据加密技术中,安全多方计算(Secure Multiparty Computation, SMC)是目前研究热点之一;数据匿名化技术主要成果有 k-anonymity、l-diversity、t-Closeness 等模型.

没有任何一种隐私保护技术适用于所有应用.适用于多方参与通信环境的安全多方计算是基于可交换加密等技术实现的,通信和计算开销都较大,不适合直接在资源严重受限的传感器节点间运行,传感器节点能量消耗量直接影响传感器网络的寿命.本文将消耗能量较高的安全比较计算设计在两个高资源节点之间完成,在性能上适用于两层传感器网络环境.k-anonymity 等匿名化技术是基于泛化等技术实现的,往往在隐私披漏风险和数据精度之间进行折中.为保证 Top-k 查询的精确度,本文采用加随机数扰乱技术,通过安全比较算法能够精确完成 Top-k 查询,加随机数扰乱技术的隐私安全程度高于 k-anonymity 模型.

(2) 传感器网络数据隐私保护技术

近年来,传感器网络数据隐私保护技术成为研究热点,有许多有价值的研究成果发表.文献[1-3]针对攻击者俘获聚集节点企图获取敏感信息的攻击模型,提出了具有保护隐私功能的 SUM、MIN、MAX 和 Median 等数据聚集算法,文献[4-6]在两

层传感器网络中存储节点被俘获的情况下，研究范围查询隐私保护和可验证技术. 文献[7]采用加密数据关系链技术，嵌入额外附加信息 *outlier*，在两层传感器网络中存储节点被俘获的情况下，实现 Top-k 查询的完整性验证，但没有考虑隐私保护问题.

Top-k 查询是传感器网络中重要的查询类型，用于对污染、干旱或易发生火灾等区域的污染指数、湿度或温度值等数据的查询和监测. 在这方面有一些研究成果[9-11]，这些研究成果集中在以减少能量为目标的查询优化方面，没有考虑隐私保护问题.

(3) 分布式隐私保护 Top-k 查询技术

在分布式 Top-k 查询隐私保护研究方面，文献[12]提出在分布的多站点中安全计算第 k 位值的算法. 文献[13]针对分布式垂直划分的数据集，提出隐私保护 Top-k 算法. 上述算法通信和计算代价大，特别是在多站点中安全发现数据项并集和计算交集数据项个数时，使用安全多方计算技术需要很高的通信和计算代价，该协议不适用于传感器网络中的 Top-k 查询.

3　模　型

3.1　网络模型

两层传感器网络存在大量资源相对较少的传感器节点和少数资源充裕的高资源节点. 传感器节点构成下层网络，将数据传送至高资源节点；高资源节点构成上层网络，完成数据聚集和执行查询任务. 本文采用的两层传感器网络模型如图 1 所示. 两层传感器网络被划分为若干个单元(cell)，单元划分时应保证每个单元区域至少存在 2 个高资源节点. 选择单元内一个高资源节点作为单元头节点(cell header)，另外指定一个高资源节点作为辅助计算节点(computing node)，即每个单元由单元头节点、辅助计算节点和若干个传感器节点组成. 单元头节点与单元头节点之间能够进行长距离、高速率通信，形成多跳的上层网络，Sink 与一些单元头节点进行无线链接，这种无线链接(如卫星链接)往往具有不稳定、高代价、低速率等特点. 这就要求隐私保护 Top-k 查询操作应在单元内进行，以尽量减少单元头节点与 Sink 之间的通信量. 单元头节点拥有本单元内网络拓扑信息，Sink 拥有全局网络拓扑信息. 由于两层传感器网络具有寿命长、易扩展等优点，使其具有良好的发展空间. 目前很多研究[4-7]采用了两层传感器网络模型.

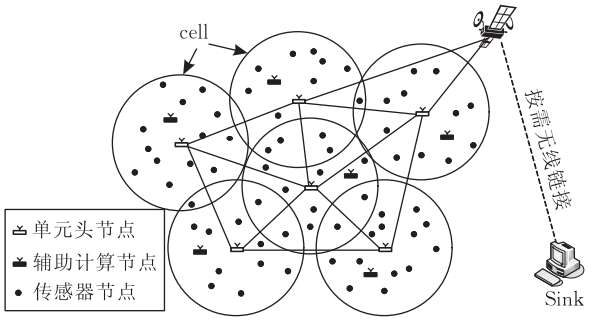

图 1　两层传感器网络模型示意图

3.2　查询模型

本文采用的传感器网络 Top-k 查询 $\mathcal{Q}_t$ 模型形式化表示为

$$\mathcal{Q}_t=(\text{cell}=C)\wedge(\text{query region}=G)\wedge(\text{epoch}=t)\wedge(\text{num}=k),$$

其中 C 为查询单元，G 为查询区域，t 为查询周期，k 为查询需要返回的最大(或最小)的数据项数.

Top-k 查询$\{C,G,t,k\}$在指定查询区域 G 和查询周期 t 内所采集的全体数据中查询 Top-k 数据，本文研究 Top-k 查询 $\mathcal{Q}_t$ 中隐私保护技术和完整性验证机制. 与文献[7]相似，SafeTQ 仅描述查询区域在同一单元内的情况，对于查询区域为多个单元的情况能够进行分解处理.

3.3　攻击模型与安全目标

单元头节点需要完成数据聚集和查询任务，攻击者俘获单元头节点对传感器网络安全性将造成严重的威胁. 本文假设攻击者企图俘获单元头节点，通过单元头节点进行攻击的情况下，研究 Top-k 查询的隐私保护技术和完整性验证机制. 传感器节点被俘获的情况下，攻击者仅能获取或修改被俘获节点采集的数据，对 Top-k 查询隐私性和完整性威胁较小，本文不考虑传感器节点被俘获的情况.

不失一般性，本文假设攻击者俘获单元头节点后，具有的特征和背景知识为：(1) 能够获得单元头节点收到和存储的明文数据，这些数据可能来自传感器节点或 Sink；(2) 不能解密使用传感器节点和 Sink 共享密钥加密的数据；(3) 能够获得单元头节点的背景知识，如网络拓扑等信息.

对传感器网络中 Top-k 查询隐私性和完整性定义如下.

定义 1.　传感器网络中 Top-k 查询隐私性.

传感器网络中 Top-k 查询 $\mathcal{Q}_t=\{C,G,t,k\}$ 正确执行时，如果攻击者不能获得以下敏感信息，则查询 $\mathcal{Q}_t$ 满足隐私性：(1) 查询区域 G 内传感器节点采集数据集 $V_t=\bigcup_{s_i\in G}V_{s_i}$ 中数据项，其中 V_{s_i} 为传感器节点

$s_i \in G$ 采集的数据集；(2) Top-k 数据集 $\mathcal{R}_t = \{vq \mid vq \in V_t$ 且 $vq \geq v_{k\text{th}}\}$ 中数据项，其中 $v_{k\text{th}}$ 为 V_t 中第 k 位数据值；(3) Top-k 节点集 $\{s_i \mid \mathcal{R}_{s_i} \neq \varnothing\}$ 中节点 ID，其中 $\mathcal{R}_{s_i} = \{vq \mid vq \in V_{s_i}$ 且 $vq \geq v_{k\text{th}}\}$.

定义 2. 传感器网络中 Top-k 查询完整性.

传感器网络中单元头节点返回的响应 Top-k 查询 $\mathcal{Q}_t = \{C, G, t, k\}$ 结果集 $\mathcal{R}_t$ 满足以下条件，则查询 $\mathcal{Q}_t$ 满足完整性：(1) 结果集 $\mathcal{R}_t$ 中任何数据项都是查询区域 G 内传感器节点采集的真实数据，即 $\mathcal{R}_t \subseteq V_t$，其中 V_t 为采集数据集；(2) 结果集 $\mathcal{R}_t$ 中任何数据都大于等于第 k 位数据值 $v_{k\text{th}}$，即 $\forall v \in \mathcal{R}_t$，有 $v \geq v_{k\text{th}}$；(3) 采集数据集 V_t 中任何大于等于第 k 位数据值 $v_{k\text{th}}$ 的数据都在结果集 $\mathcal{R}_t$ 中，即 $\forall v \in V_t$ 且 $v \geq v_{k\text{th}}$，有 $v \in \mathcal{R}_t$.

本文假设单元头节点被俘获后采用如下两种模式进行攻击：(1) 破坏隐私性攻击模式，攻击者企图记录或推测定义 1 中描述的敏感信息，破坏传感器网络中 Top-k 查询隐私性；(2) 破坏完整性攻击模式，攻击者企图通过插入、篡改或删除查询响应结果，破坏传感器网络中 Top-k 查询完整性. 本文的完全目标为：SafeTQ 能够应对以上两种模式的攻击，在不泄漏隐私信息的情况下精确地完成 Top-k 查询，并且能够使 Sink 检测和拒绝不正确或不完整查询响应.

4 SafeTQ 协议

两层传感器网络中可验证隐私保护 Top-k 查询协议 SafeTQ 由隐私保护 Top-k 查询协议和两种完整性验证模式组成. 两种完整性验证模式分别为数据项加密链验证模式和概率空间邻居验证模式. 根据使用的完整性验证模式不同，SafeTQ 分为 SafeTQ-L 和 SafeTQ-N，其中 SafeTQ-L 使用数据项加密链验证模式，SafeTQ-N 使用概率空间邻居验证模式.

4.1 隐私保护 Top-k 查询协议

隐私保护 Top-k 查询协议的基本思路为：接到 Top-k 查询 $\mathcal{Q}_t$ 后，传感器节点 $s_i \in G$ 产生随机数 r_i，将周期 t 内采集的数据集 V_{s_i} 中前 k 个数据 $\{v_{i,1}, v_{i,2}, \cdots, v_{i,k}\}$ 分别与随机数 r_i 求和，将求和结果传送到单元头节点，同时将该随机数 r_i 传送到辅助计算节点，由单元头节点与辅助计算节点通过安全比较，计算出周期 t 内查询区域 G 中采集的全体数据集 $V_t = \bigcup_{s_i \in G} V_{s_i}$ 中第 k 位数据值 $v_{k\text{th}}$，单元头节点将 $v_{k\text{th}}$ 作为阈值下发至查询区域内所有传感器节点，传感器节点 $s_i \in G$ 将 V_{s_i} 中每个数据 $v_{i,j}$ 与 $v_{k\text{th}}$ 比较，大于等于 $v_{k\text{th}}$ 的数据即为 Top-k 数据，这些 Top-k 数据构成传感器节点 s_i 查询响应结果集 $\mathcal{R}_{s_i}$，根据所选用的完整性验证模式不同进行相应处理后，使用传感器节点与 Sink 共享的密钥加密并传送到单元头节点，由单元头节点上传至 Sink，Sink 解密并确定查询 $\mathcal{Q}_t$ 结果集 $\mathcal{R}_t = \bigcup_{s_i \in G} \mathcal{R}_{s_i}$（即 Top-$k$ 数据集）.

结合图 2，对隐私保护 Top-k 查询协议进行描述如下：

(1) 传感器节点数据加随机数处理

Sink 将 Top-k 查询 $\mathcal{Q}_t = \{C, G, t, k\}$ 发送给单元头节点，单元头节点将查询参数 $\{t, k\}$ 下发给查询区域 G 内的传感器节点 $s_i \in G$. 传感器节点 $s_i \in G$ 对在周期 t 内采集的数据集 V_{s_i} 进行排序，并取前 k 个数据构成 $\{v_{i,1}, v_{i,2}, \cdots, v_{i,k}\}$ ($v_{i,1} \geq v_{i,2} \geq \cdots \geq v_{i,k}$ 且 $\forall l \neq j (j = 1, 2, \cdots, k), v_{i,l} \leq v_{i,j}$)，传感器节点 $s_i \in G$ 产生随机数 r_i，将 r_i 与前 k 个数据集 $\{v_{i,1}, v_{i,2}, \cdots, v_{i,k}\}$ 中每一项数据分别进行求和，数据项 $v_{i,j} \in \{v_{i,1}, v_{i,2}, \cdots, v_{i,k}\}$ ($j = 1, 2, \cdots, k$) 与 r_i 求和结果记为 $vr_{i,j} = v_{i,j} + r_i$，传感器节点 s_i 将求和结果 $\{vr_{i,1}, vr_{i,2}, \cdots, vr_{i,k}\}$，发送到单元头节点：

$$s_i (s_i \in G) \rightarrow \text{cell header}: i, \{vr_{i,1}, vr_{i,2}, \cdots, vr_{i,k}\}.$$

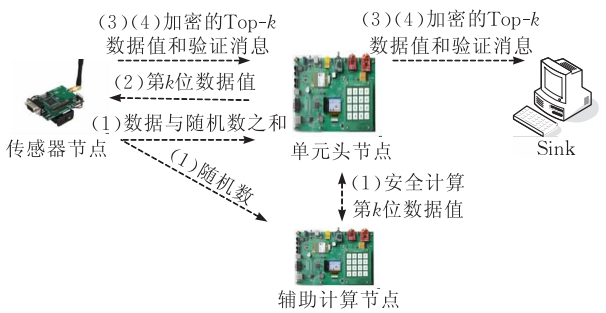

图 2 隐私保护 Top-k 查询协议演示图

同时，传感器节点 s_i 将本节点产生的随机数 r_i，发送到辅助计算节点：

$$s_i (s_i \in G) \rightarrow \text{computing node}: i, \{r_i\}.$$

(2) 安全计算第 k 位数据值

单元头节点与辅助计算节点通过 4.2 节描述的算法，在单元头节点和辅助计算节点不能获知任何数据项值的情况下，计算出数据集 $V_t = \bigcup_{s_i \in G} V_{s_i}$ 中第 k 位数据值 $v_{k\text{th}}$，$v_{k\text{th}}$ 将数据集 V_t 中数据值最大的前 k 项与第 $k+1$ 项及其之后的数据区分开来. 单元头节点将阈值 $v_{k\text{th}}$ 发送至查询区域内所有传感器节点 $s_i \in G$：

$$\text{cell header}\rightarrow s_i(s_i\in G):\{v_{k\text{th}}\}.$$

(3) 生成查询响应结果集并加密上传

传感器节点 $s_i\in G$ 将 $v_{k\text{th}}$ 与周期 t 内采集的数据集 V_{s_i} 进行比较，大于等于 $v_{k\text{th}}$ 的数据(即为 Top-k 数据)形成节点 s_i 的查询响应结果集 $\mathcal{R}_{s_i}=\{vq\mid vq\in V_{s_i}\text{ 且 }vq\geq v_{k\text{th}}\}$.

SafeTQ 使用传感器节点与 Sink 共享的密钥，对需要上传的数据进行加密，单元头节点没有解密能力. 为了增加安全性，本协议采用文献[4]中周期更换密钥方式，用 $k_{i,t}$ 表示周期 t 内传感器节点 s_i 与 Sink 共享的密钥，则有 $k_{i,t}=\text{hash}(k_{i,t-1})$.

对 $\mathcal{R}_{s_i}$ 分情况进行加密处理，使用数据项加密链验证模式时，需要按 4.3.1 节要求进行加密并形成数据项加密链；使用概率空间邻居验证模式时，对 $\mathcal{R}_{s_i}$ 直接进行整体加密. 加密后经单元头节点上传至 Sink：

$$s_i(\mathcal{R}_{s_i}\neq\varnothing)\rightarrow\text{cell header}\rightarrow\text{Sink}:\ i,E_{k_{i,t}}\{\mathcal{R}_{s_i}\}.$$

(4) 使用完整性验证模式生成验证消息

根据所选用完整性验证模式进行不同的处理. 使用数据项加密链验证模式，需要所有非 Top-k 节点 $s_i(\mathcal{R}_{s_i}=\varnothing)$)产生验证消息 $\mathcal{M}_{vl}$(详见 4.3.1 节)；使用概率空间邻居验证模式，需要 Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$的邻居节点以概率 p 发送验证消息 $\mathcal{M}_{vn}$(详见 4.3.2 节).

不使用完整性验证模式的情况下，隐私保护 Top-k 查询协议需要部分非 Top-k 节点发送虚假消息 $\mathcal{M}_f$，以避免单元头节点能够将上传加密数据的节点准确地判断为 Top-k 节点，造成敏感信息泄漏.

为了下式描述方便，将 $\mathcal{M}_{vn}$、$\mathcal{M}_{vl}$ 或 $\mathcal{M}_f$ 统一表示为 $\mathcal{M}$. 使用传感器节点 s_i 与 Sink 共享的密钥 $k_{i,t}$ 对 $\mathcal{M}$ 进行直接加密后，经单元头节点上传至 Sink：

$$s_i(s_i\in G)\rightarrow\text{cell header}\rightarrow\text{Sink}:\ i,E_{k_{i,t}}\{\mathcal{M}\}.$$

(5) Sink 解密并处理数据

Sink 解密数据，并进行如下处理：根据消息标志判断验证消息或虚假数据，根据完整性验证模式进行相应的完整性验证(详见 4.3.1 节和 4.3.2 节)，求所有 $\mathcal{R}_{s_i}$ 并集得到查询 $\mathcal{Q}_t$ 结果集 $\mathcal{R}_t=\bigcup\limits_{s_i\in G}\mathcal{R}_{s_i}$，即 Top-$k$ 数据集.

4.2　安全计算第 k 位数据值算法

本节给出安全计算第 k 位数据值算法. 安全计算第 k 位数据值算法利用二分查找和安全比较算法[12-14]，在单元头节点和辅助计算节点不能获知任何数据项值的情况下，安全计算出数据集 V_t 中第 k 位的数据值 $v_{k\text{th}}$. 1986 年发表的文献[14]提出的安全比较算法，实现了在参与计算双方不能获知对方数据值的情况下进行安全比较计算.

传感器节点 $s_i\in G$ 将产生的随机数 r_i 与前 k 个数据项求和 $vr_{i,j}=v_{i,j}+r_i(j=1,2,\cdots,k)$，并将求和结果 $\{vr_{i,1},vr_{i,2},\cdots,vr_{i,k}\}$ 传送给单元头节点，同时将随机数 r_i 传送到辅助计算节点，显然有 $v_{i,j}=vr_{i,j}-r_i$. 为了表述简单，将单元头节点收到的所有求和结果 $\bigcup\{vr_{i,1},vr_{i,2},\cdots,vr_{i,k}\}$ 统一表示为 $\{a_1,a_2,\cdots,a_n\}(n=|\bigcup\{vr_{i,1},vr_{i,2},\cdots,vr_{i,k}\}|)$，辅助计算节点将收到的所有随机数求相反数后得集合 $\bigcup\{-r_i\}$，将集合 $\bigcup\{-r_i\}$ 相应项表示为 $\{b_1,b_2,\cdots,b_n\}(n=|\bigcup\{vr_{i,1},vr_{i,2},\cdots,vr_{i,k}\}|)$，则 $\forall m,v_m=a_m+b_m$，有 $v_m\in\{v_{i,j}\}$.

安全计算第 k 位数据值算法，使用二分查找法搜索第 k 位数据值 $v_{k\text{th}}$. 用 *middle* 表示中间探测值，对于每一个数据项，单元头节点与辅助计算节点运用安全比较算法，比较 *middle* 与 $v_m=a_m+b_m$ 的大小关系，根据 *middle* 与 v_m 的大小关系，产生相应输出. 在全部数据项安全比较完成后，单元头节点与辅助计算节点对各自所得输出结果求和，使用安全比较算法将该求和结果与 k 进行比较，依此判断 *middle* 与 $v_{k\text{th}}$ 的大小关系，如相等则返回 $v_{k\text{th}}$，否则，按二分查找算法进行下一轮处理，直到找到并返回 $v_{k\text{th}}$. 假设共有 x 项数据上传至单元头节点，用 M 表示数据值域上界，算法能够在 $\log|M|$ 轮内找到第 k 位数据值，在每轮中所有 x 项数据都需要参与安全比较，算法复杂度为 $O(x\log|M|)$.

算法 1.　安全计算第 k 位数据值.

输入：单元头节点和辅助计算节点各自输入数据集 $\{a_1,a_2,\cdots,a_n\}$ 和 $\{b_1,b_2,\cdots,b_n\}$

输出：第 k 位数据值 $v_{k\text{th}}$

1. $lbound\leftarrow 0$　　//数据值域下界
2. $ubound\leftarrow M$　　//数据值域上界 M
3. while true do
4. 　$middle\leftarrow\lceil(lbound+ubound)/2\rceil$
5. 　for $m=1,2,\cdots,n$ do
6. 　　使用文献[14]的安全比较算法，输入 a_m,b_m，*middle*，输出 la_m,lb_m，如 $a_m+b_m\leq middle$，则有 $la_m+lb_m=1\pmod{|M|}$，反之，如果 $a_m+b_m>middle$，则有 $la_m+lb_m=0\pmod{|M|}$
7. 　end for
8. 　单元头节点：$lla\leftarrow\sum\limits_{m=1}^{n}la_m$
9. 　辅助计算节点：$llb\leftarrow\sum\limits_{m=1}^{n}lb_m$

```
10. if lla_m + llb_m (mod |M|) > k then  //middle > v_kth
11.     ubound ← middle
12. else if lla_m + llb_m (mod |M|) < k then //middle < v_kth
13.     lbound ← middle
14. else                                  //middle = v_kth
15.     return middle //返回第 k 位数据值 v_kth
16. end if
17. end while
```

4.3 完整性验证模式

单元头节点可能通过插入虚假数据、篡改和删除数据，破坏 3.3 节中定义 2 定义的传感器网络中 Top-k 查询完整性. 插入虚假数据破坏定义 2 中条件 1，同时可能破坏条件 2；篡改数据破坏条件 1 和条件 3，可能破坏条件 2；删除数据破坏条件 3. SafeTQ 使用传感器节点与 Sink 共享的密钥进行加密，在假设传感器节点没有被俘获的前提下，单元头节点不能获得密钥即没有解密和加密能力，所以单元头节点插入虚假数据和篡改数据时，破坏了加密数据包的完整性，Sink 不能正确解密，从而能够检测到该攻击. 但是，如果被俘获的单元头节点删除传感器节点上传的加密数据包，则破坏了查询 $\mathcal{Q}_t$ 的完整性，需要使用完整性验证模式进行完整性验证.

本文设计的两种完整性验证模式分别为数据项加密链验证模式和概率空间邻居验证模式，分别与隐私保护 Top-k 查询协议结合形成 SafeTQ-L 和 SafeTQ-N.

4.3.1 数据项加密链验证模式

数据项加密链验证模式使用数据项加密链技术检测对 $\mathcal{R}_{s_i}$ 的部分删除攻击，并使用非 Top-k 节点发送验证消息策略检测对 $\mathcal{R}_{s_i}$ 的整体删除攻击，由数据项加密链和验证消息共同完成 Top-k 查询完整性验证. 本文针对隐私保护 Top-k 查询协议设计了相应的数据项加密链数据结构. 数据项加密链技术[7-8]通过对有序数据集填充冗余数据，分段加密形成数据项加密链，使得解密后能够检测数据集的完整性. 通过示例对数据项加密链技术进行演示，假设有序数据集为 $\{2,4,6,8\}$，使用密钥 $k_{i,t}$ 加密后形成的数据项加密链为 $\{(\mathrm{Min}\mid 2)_{k_{i,t}},(2\mid 4)_{k_{i,t}},(4\mid 6)_{k_{i,t}},(6\mid 8)_{k_{i,t}},(8\mid \mathrm{Max})_{k_{i,t}}\}$.

分别使用 $\underline{\mathcal{X}}$ 和 $\overline{\mathcal{X}}$ 表示公共极小值和极大值，使用 n_i 表示 Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$ 形成查询响应结果集 $\mathcal{R}_{s_i}$ 中数据项个数，即 $n_i=|\mathcal{R}_{s_i}|$.

对数据项加密链验证模式进行描述如下：

(1) Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$ 对 $\mathcal{R}_{s_i}$ 排序，排序后表示为 $\{vq_{i,1},vq_{i,2},\cdots,vq_{i,n_i}\}$ $(vq_{i,1}>vq_{i,2}>\cdots>vq_{i,n_i})$，使用传感器节点与 Sink 共享的密钥 $k_{i,t}$，对排序后的 $\mathcal{R}_{s_i}$ 加入冗余数据和公共极值 $\underline{\mathcal{X}}$ 和 $\overline{\mathcal{X}}$，分段加密后形成数据项加密链 $\{(\underline{\mathcal{X}}\mid vq_{i,1})_{k_{i,t}},(vq_{i,1}\mid vq_{i,2})_{k_{i,t}},\cdots,(vq_{i,n_i-1}\mid vq_{i,n_i})_{k_{i,t}},(vq_{i,n_i}\mid\overline{\mathcal{X}})_{k_{i,t}}\}$，将此数据项加密链经过单元头节点上传至 Sink：

$$
\begin{aligned}
&s_i(\mathcal{R}_{s_i}\neq\varnothing)\rightarrow\text{cell header}\rightarrow\text{Sink}:\\
&\quad i,\langle(\underline{\mathcal{X}}\mid vq_{i,1})_{k_{i,t}}\rangle,\\
&\quad\langle(vq_{i,1}\mid vq_{i,2})_{k_{i,t}}\rangle,\\
&\quad\vdots\\
&\quad\langle(vq_{i,n_i-1}\mid vq_{i,n_i})_{k_{i,t}}\rangle,\\
&\quad\langle(vq_{i,n_i}\mid\overline{\mathcal{X}})_{k_{i,t}}\rangle.
\end{aligned}
$$

(2) 查询区域 G 内所有非 Top-k 节点 $s_i(\mathcal{R}_{s_i}=\varnothing)$ 生成验证消息 $\mathcal{M}_{vl}$，采用传感器节点与 Sink 共享的密钥 $k_{i,t}$ 加密，经单元头节点上传至 Sink：

$$s_i(\mathcal{R}_{s_i}=\varnothing)\rightarrow\text{cell header}\rightarrow\text{Sink}:\ i,\{(\mathcal{M}_{vl})_{k_{i,t}}\}.$$

(3) Sink 解密并进行完整性验证

Sink 对所有收到的加密数据进行解密，本模式需要 Sink 进行两类验证：(i) 所有数据项加密链的完整性；(ii) 是否查询区域 G 内所有节点均上传了加密数据. 如果有数据项加密链不完整或查询区域 G 内有节点未上传加密数据，则判断为 Top-k 查询 $\mathcal{Q}_t$ 不完整.

下面讨论攻击者进行不同方式的完整性攻击时，数据项加密链验证模式进行完整性验证的情况. 传感器节点和 Sink 共享密钥的加密方式，使 Sink 能够检测出攻击者插入虚假数据和篡改数据的攻击方式. 所以，完整性验证模式仅需要考虑攻击者部分或全部删除加密数据的情况：

(i) 如果 $\{(vq_{i,j}\mid vq_{i,j-1})_{k_{i,t}}\}$ $(1<j\leq n_i)$ 被单元头节点删除，因不能形成完整的数据项加密链，而能够被 Sink 检测出；

(ii) 如果 $\{(\underline{\mathcal{X}}\mid vq_{i,1})_{k_{i,t}}\}$ 或 $\{(vq_{i,n_i}\mid\overline{\mathcal{X}})_{k_{i,t}}\}$ 被单元头节点删除，因数据项加密链缺少公共极少值 $\underline{\mathcal{X}}$ 或公共极大值 $\overline{\mathcal{X}}$，而能够被 Sink 检测出；

(iii) 如果某数据项加密链被整体删除，Sink 根据网络拓扑信息，能够检测出未收到加密数据的节点，因为本模式要求查询区域 G 内所有节点发送加密数据，Sink 据此判断受到了完整性攻击.

此模式需要查询区域 G 内所有节点发送加密数据，即查询区域 G 内非 Top-k 节点也需要发送加密的验证消息，并且需要冗余数据形成数据项加密

链，则此模式需要较高的通信代价. 下面给出通信代价较小的概率空间邻居验证模式.

4.3.2　概率空间邻居验证模式

概率空间邻居验证模式的基本步骤为：需要响应查询的 Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$，对查询响应结果集 $\mathcal{R}_{s_i}$ 整体加密后上传，同时向邻居节点广播消息 $\mathcal{M}_{\mathcal{Q}}$，表明本节点为 Top-$k$ 节点，收到 $\mathcal{M}_{\mathcal{Q}}$ 的邻居节点生成验证消息 $\mathcal{M}_{vn}$，将 $\mathcal{M}_{vn}$ 与 Top-k 节点 s_i 的编号 i 一起加密后，以给定概率 p 上传，Sink 解密后使用验证消息进行完整性验证.

对概率空间邻居验证模式进行描述如下：

在传感器网络隐私保护 Top-k 算法中，Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$ 形成查询响应结果集 $\mathcal{R}_{s_i}$ 后，进行以下操作：

(1) Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$ 使用与 Sink 共享密钥 $k_{i,t}$，对 $\mathcal{R}_{s_i}$ 进行整体加密后，经单元头节点上传至 Sink：

$$s_i(\mathcal{R}_{s_i}\neq\varnothing)\rightarrow\text{cell header}\rightarrow\text{Sink}:\ i,\{\mathcal{R}_{s_i}\}_{k_{i,t}}.$$

(2) Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$ 向空间邻居节点广播消息 $\mathcal{M}_{\mathcal{Q}}$，表明本节点为 Top-$k$ 节点：

$$s_i(\mathcal{R}_{s_i}\neq\varnothing)\rightarrow\text{neighbourhood}:\ i,\ \mathcal{M}_{\mathcal{Q}}.$$

(3) Top-k 节点 s_i 的邻居节点 s_j 收到广播消息 $\mathcal{M}_{\mathcal{Q}}$ 后，生成验证消息 $\mathcal{M}_{vn}$，使用与 Sink 共享的密钥 $k_{j,t}$，将 $\mathcal{M}_{vn}$ 与 Top-k 节点 s_i 的编号 i 一起加密，为了减少通信量，以给定概率 p 经单元头节点上传至 Sink：

$$p:\ s_j\rightarrow\text{cell header}\rightarrow\text{Sink}:\ j,\{i,\mathcal{M}_{vn}\}_{k_{j,t}}.$$

(4) Sink 解密并进行完整性验证

Sink 对加密数据进行解密. 对于与验证消息 $\mathcal{M}_{vn}$ 一起加密的节点编号 i，Sink 检测是否收到其对应节点 $s_i\in G$ 的查询响应结果集 $\mathcal{R}_{s_i}$，如没收到，则判断为查询 $\mathcal{Q}_t$ 结果集 $\mathcal{R}_t=\bigcup_{s_i\in G}\mathcal{R}_{s_i}$ 不完整.

在概率空间邻居验证模式中，传感器节点 s_i 使用与 Sink 节点共享的密钥 $k_{i,t}$，对查询结果集 $\mathcal{R}_{s_i}$ 进行整体加密，因此单元头节点插入虚假数据包、篡改数据将被检测出，其仅能够通过整体删除加密后的查询响应结果 $\{\mathcal{R}_{s_i}\}_{k_{i,t}}$ 来破坏查询完整性. Sink 通过空间邻居节点发送的验证消息，能够判断攻击者是否删除 $\{\mathcal{R}_{s_i}\}_{k_{i,t}}$，以完成完整性验证.

5　安全性分析

5.1　隐私性分析

定理 1.　SafeTQ 正确完成 Top-k 查询 $\mathcal{Q}_t$ 过程中，单元头节点不能获得 Top-k 查询隐私性定义(定义 1)中描述的敏感信息.

证明.　SafeTQ 在完成 Top-k 查询 $\mathcal{Q}_t$ 过程中，单元头节点可能在以下环节获取定义 1 中描述的敏感信息：(环节 1)为计算第 k 位数据值 $v_{k\text{th}}$ 而收集数据信息时，获取采集数据值；(环节 2)与辅助计算节点合作计算 k 位数据值 $v_{k\text{th}}$ 时，获取采集数据值和 Top-k 数据值；(环节 3)接收 Top-k 节点上传的查询响应结果集 $\mathcal{R}_{s_i}$ 时，获取 Top-k 数据值并推测 Top-k 节点.

在 SafeTQ 执行时，在环节 1 中传感器节点将数据与本节点产生的随机数求和后上传，单元头节点无法获得真实数据；在环节 2 中使用二分查找和安全比较计算第 k 位数据值，单元头节点和辅助计算节点不能获知任何数据项值，安全比较算法[14]保证了参与计算的双方不能获知对方的数据值；在环节 3 中，一方面，使用传感器节点和 Sink 共享密钥加密数据，单元头节点不能通过解密获取 Top-k 数据值，另一方面，SafeTQ 要求部分非 Top-k 节点产生扰乱数据，使单元头节点不能将上传加密数据的节点准确地推测为 Top-k 节点. 综上所述，在 SafeTQ 执行过程中，单元头节点不能获得定义 1 中描述的敏感信息.　证毕.

定理 2.　SafeTQ 正确完成 Top-k 查询 $\mathcal{Q}_t$ 过程中，辅助计算节点不能获得 Top-k 查询隐私性定义(定义 1)中描述的敏感信息.

证明.　SafeTQ 执行 Top-k 查询 $\mathcal{Q}_t$ 过程中，辅助计算节点仅负责接收传感器节点 $s_i\in G$ 传送的随机数 r_i，与单元头节点使用安全比较算法计算第 k 位数据值 $v_{k\text{th}}$. 安全比较算法[14]保证辅助计算节点不能获得单元头节点的数据信息. SafeTQ 执行过程中采集的数据和查询响应的 Top-k 数据不向辅助计算节点传送. 所以，定理 2 成立.　证毕.

定理 3.　SafeTQ 正确完成 Top-k 查询 $\mathcal{Q}_t$ 过程中，在传感器节点没有被俘获的情况下，SafeTQ 满足传感器网络中 Top-k 查询隐私性定义(定义 1).

证明.　根据定理 1、2，SafeTQ 执行 Top-k 查询 $\mathcal{Q}_t$ 过程中，单元头节点和辅助计算节点不能获得定义 1 中敏感信息. 在传感器节点没有被俘获的情况下，SafeTQ 执行过程中攻击者不能获得定义 1 中敏感信息. 所以，SafeTQ 满足传感器网络中 Top-k 查询隐私性定义.　证毕.

5.2　完整性分析

定理 4.　在没有传感器节点被俘获发起攻击

的情况下，SafeTQ-L 检测出单元头节点被俘获发起完整性攻击的概率为 100%.

证明. 单元头节点可能通过插入虚假数据包、篡改数据和删除数据操作进行完整性攻击，破坏定义 2 定义的 Top-k 查询完整性. 在没有传感器节点被俘获发起攻击的情况下，由于 SafeTQ-L 使用传感器节点与 Sink 共享密钥的加密方式，Sink 检测出单元头节点插入虚假数据和篡改数据的概率为 100%；由于 SafeTQ-L 使用数据项加密链技术，Sink 检测出单元头节点删除数据项加密链中任意项的概率为 100%；由于 SafeTQ-L 要求查询区域 G 内所有节点全部发送加密数据，且 Sink 拥有网络拓扑信息，所以 Sink 检测出单元头节点删除某传感器节点上传的全部加密数据的概率为 100%. 综上所述，在没有传感器节点被俘获发起攻击的情况下，SafeTQ-L 检测出单元头节点被俘获发起完整性攻击的概率为 100%.

定理 5. 没有传感器节点被俘获发起攻击的情况下，在 SafeTQ-N 的 Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$ 和其空间邻居节点传送到单元头节点的加密数据中，单元头节点随机删除某加密数据项，使得完整性被破坏且被检测出的概率为 $1-(1-p)^n/(p\times n+1)$，其中 p 为 SafeTQ-N 中邻居节点发送验证消息的概率，n 为 Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$ 空间邻居的个数.

证明. 在 SafeTQ-N 使用的概率空间邻居验证模式中，Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$ 将消息 $\mathcal{M}_Q$ 广播给空间邻居节点，空间邻居节点以概率 p 向单元头节点发送验证消息，设 Top-k 节点 s_i 有 n 个空间邻居节点，则没有任何空间邻居节点发送验证消息的概率为 $(1-p)^n$，向单元头节点发送验证消息的邻居节点个数的期望值为 $p\times n$. SafeTQ-N 将验证消息与响应查询数据项设计为相同长度，使得单元头节点不容易区分加密的查询响应数据项和加密的验证消息项. 只有单元头节点删除加密查询响应数据项，Top-k 查询完整性才被破坏. 在 SafeTQ-N 的 Top-k 节点 $s_i(\mathcal{R}_{s_i}\neq\varnothing)$ 和其空间邻居上传的加密数据包中，单元头节点随机删除的加密数据项为加密查询响应数据项的概率为 $1/(p\times n+1)$，此情况下 Sink 不能检测到的概率为 $(1-p)^n/(p\times n+1)$. 所以，没有传感器节点被俘获发起攻击的情况下，单元头节点随机删除某数据项，使得完整性被破坏且被检测出的概率为 $1-(1-p)^n/(p\times n+1)$. 证毕.

定理 6. SafeTQ 执行 Top-k 查询 $\mathcal{Q}_t$ 时，能够检测出辅助节点被俘获发起的完整性攻击.

证明. SafeTQ 执行过程中，辅助计算节点负责接收传感器节点 $s_i\in G$ 传送的随机数 r_i，与单元头节点使用安全比较算法计算第 k 位数据值 $v_{k\text{th}}$. 辅助计算节点被俘获仅能够通过插入、篡改和删除其接收的随机数，进行完整性攻击. 根据安全计算第 k 位数据值算法(详见 4.2 节)，辅助计算节点接收到的随机数 r_i 与单元头节点收到的求和结果 $vr_{i,j}=v_{i,j}+r_i$ 一一对应. 当辅助计算节点插入或删除随机数项后，安全计算第 k 位数据值算法执行时将出错，从而检测出此攻击. 如果辅助计算节点篡改随机数值，将计算出错误的第 k 位数据值 $v_{k\text{th}}$，使传感器节点筛选出的查询响应结果集 $\mathcal{R}_{s_i}$ 不完整或不正确，进而使 Sink 解密汇总的查询结果集 $\mathcal{R}_t$ 不完整，此时 $\mathcal{R}_t$ 中数据项数不等于 k($|\mathcal{R}_t|\neq k$). 执行 SafeTQ-L 时，Sink 确认所有数据项加密链完整且已收到所有传感器节点上传的加密数据后，如检测 $|\mathcal{R}_t|$ 不等于 k 值，则判断为辅助节点被俘获并发起完整性攻击；执行 SafeTQ-N 时，Sink 确认验证消息携带编号对应节点上传的加密数据全部收到后，如检测 $|\mathcal{R}_t|$ 不等于 k 值，则判断为辅助节点被俘获并发起完整性攻击. 证毕.

6 实验与分析

本文设计的 SafeTQ 协议代价主要由三部分组成：传感器节点采集、计算和传输数据消耗的代价；单元头节点与辅助计算节点两个高资源节点之间安全计算第 k 位数据值 $v_{k\text{th}}$ 的代价；单元头节点与 Sink 之间传输数据的代价. 4.2 节已给出单元头节点与辅助计算节点之间安全计算第 k 位数据值的算法复杂度. 对于传感器节点，通信能量消耗远大于计算能量消耗，如文献[7,15]，本文使用数据通信量评价算法的有效性. 本实验对单元内传感器节点通信量和单元头节点与 Sink 之间的外部通信量分别进行考查，前者影响整个网络寿命，后者使用高代价、低速率的无线链接(如卫星链接)，要求通信量尽量小.

单元内传感器节点传输数据通信量主要由 4 部分构成：

(1) 为安全计算阈值 $v_{k\text{th}}$ 向单元头节点传输采集数据和随机数之和，数据项格式为

节点 ID：4 B	数据和随机数之和：4 B

；

(2) 为安全计算阈值 $v_{k\text{th}}$ 向辅助计算节点传输随机数，数据项格式为

节点 ID：4 B	随机数：4 B

；

(3) 传感器节点对查询响应数据加密后向单元头节点传输，数据格式为

日期:4B	时间:4B	节点 ID:4B	Top-k 数据:4B

(4) 传输验证消息，在数据项加密链验证模式中，非 Top-k 节点对验证消息加密后向单元头节点传输，数据格式为 | 节点 ID:4B | 验证信息:4B |，在概率空间邻居验证模式中，Top-k 节点向邻居节点广播数据的格式为 | 节点 ID:4B | 广播消息:4B |，邻居节点对验证消息加密后向单元头节点传输，数据格式为 | Top-k 节点 ID:4B | 验证信息:12B |，将邻居节点发送的验证消息设计为 16 字节，使得攻击者不容易区分验证消息和查询响应消息. 本文采用文献[4,7]中密钥机制，使用 16 位密钥进行加密.

本文使用真实数据集在 OMNeT++平台上对 SafeTQ 进行仿真实验. 本文使用公开可获得的真实数据集 LUCE data set①，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 11 维属性数据，本文使用环境温度数据进行仿真. 根据 LUCE data set 中节点位置数据，88 个传感器节点分布在半径 280 m 的圆形区域内，这些传感器节点和单元头节点、辅助计算节点一起构成两层传感器网络的划分单元，作为仿真实验查询区域. 本文仿真实验中，单元头节点位于单元中心，假设传感器节点有效传输距离为 60 m，使用 TAG(tiny aggregation service for ad hoc sensor networks)算法建立路由，传感器节点向单元头节点传输数据平均需要 3.13 跳，每个传感器节点平均有 10.97 个邻居节点.

本实验的前两组实验分别考查查询参数 k 值和查询周期 t 对以下 3 种协议单元内传感器节点通信量和单元外通信量影响：使用数据项加密链验证模式的 SafeTQ-L、使用概率空间邻居验证模式的 SafeTQ-N 和没有隐私保护和验证机制的 Top-k 查询协议 Topk-TAG. 为了增加可比性，Topk-TAG 和 SafeTQ 使用同样的 Top-k 查询策略和 TAG 路由. 第 3 组实验考察 SafeTQ-N 中空间邻居节点发送验证消息的概率 p 对 SafeTQ-N 协议通信量的影响. 为便于描述，本实验假设 1 个时间单位传感器节点采集 1 轮数据，即查询周期 t 内采集 t 轮数据.

第 1 组实验考查 k 值对通信量的影响. 实验在参数 $t=10$ 和 $p=0.6$ 下进行. 图 3 显示 k 值对单元内传感器节点通信量影响，可以看出 SafeTQ-L 和 SafeTQ-N 通信量高于没有隐私保护和验证机制的 Topk-TAG，SafeTQ-L 通信量略高于 SafeTQ-N. 同时，三者趋势基本相同，当 $k<10$ 时，通信量随 k 值接近于线性增长，当 $k\geq 10$ 时，通信量增长缓慢. 这是因为单元内传感器节点通信量主要由传感器节点将采集数据的前 k 项或全部数据传输到单元头节点产生，以计算第 k 位数据值 $v_{k\text{th}}$. 在本组实验中 $t=10$，根据假设在查询周期 t 内传感器节点采集 10 项数据. 为计算第 k 位数据值 $v_{k\text{th}}$，当 $k<t$ 时，传感器节点上传前 k 项数据，通信量随 k 值呈近似线性增长，当 $k\geq t$ 时，传感器节点上传全部的 t 项数据，此时通信量随 k 值增长不再显著. SafeTQ 中由于传输随机数、验证消息和对响应查询数据加密，使得 SafeTQ-L 和 SafeTQ-N 通信量高于 Topk-TAG. 图 4 显示 k 值对单元头节点和 Sink 之间通信量的影响，可以看出 SafeTQ-L 通信量明显高于 SafeTQ-N 和 Topk-TAG，这是因为 SafeTQ-L 使用的数据项加密链验证模式要求使用增加冗余数据形成数据项加密链，并且需要全部非 Top-k 节点发送验证消息，使得其通信量高于 SafeTQ-N.

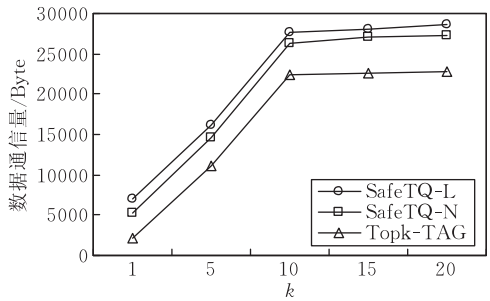

图 3　k 值对单元内传感器节点通信量的影响($t=10$，$p=0.6$)

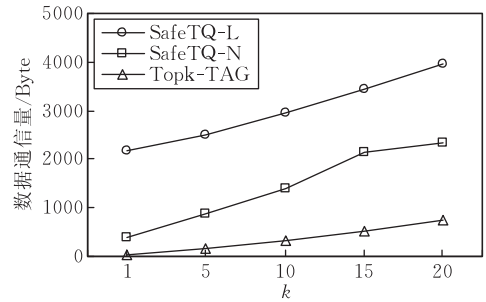

图 4　k 值对单元外部通信量的影响($t=10$，$p=0.6$)

第 2 组实验考查查询周期 t 对通信量的影响. 实验在参数 $k=10$ 和 $p=0.6$ 下进行. 根据假设查询周期 t 内传感器节点采集 t 轮数据. 图 5 考查查询周期 t 对单元内传感器节点通信量影响，通信量增长的主体趋势和原因与图 3 相似. 为计算第 k 位数据值 $v_{k\text{th}}$，当 $t<k$ 时，传感器节点上传全部的 t 项数据，通信量随 t 值接近于线性增长，当 $t\geq k$ 时，传感

① http://sensorscope.epfl.ch/index.php/Environmental_Data

器节点上传前 k 项数据，此时通信量随 t 值增长不再显著. 在图 5、图 6 中，$t=1$ 时，SafeTQ-N 的通信量高于 SafeTQ-L，这是因为从统计意义上看，k 值相同时，t 值越小 Top-k 节点数目越多，当 $t=1$ 时，Top-k 节点个数为 k，而 SafeTQ-N 要求每个 Top-k 节点广播消息给邻居节点，邻居节点再以概率发送验证消息，所以 Top-k 节点数量对 SafeTQ-N 通信量影响较大. 图 6 显示 t 值对单元头节点和 Sink 之间通信量的影响，可以看出 t 值对 SafeTQ-N 影响较大，对 SafeTQ-L 影响较小，而 Topk-TAG 不受 t 值影响.

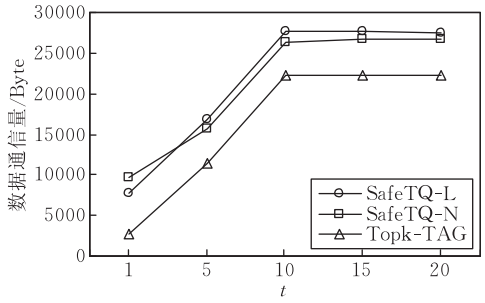

图 5 查询周期 t 对单元内传感器节点通信量的影响 ($k=10$, $p=0.6$)

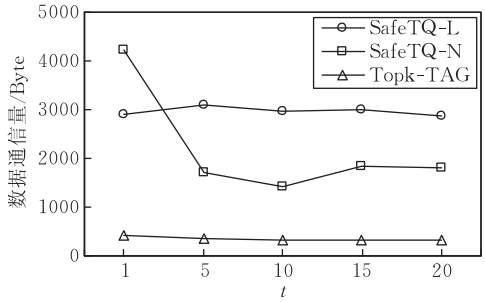

图 6 查询周期 t 对单元外部通信量的影响($k=10$, $p=0.6$)

第 3 组实验考查 SafeTQ-N 发送验证消息概率 p 对通信量的影响. 实验在参数 $k=10$ 和 $t=10$ 下进行. 从图 7、图 8 可以看出，单元内传感器节点通信量和单元外通信量均随概率增大而单调递增. 同时可以看出，无论 p 取何值，SafeTQ-N 通信量均低于 SafeTQ-L，但高于 Topk-TAG.

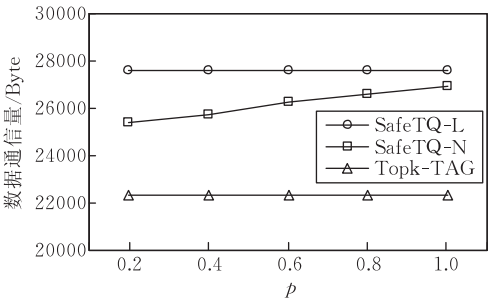

图 7 概率 p 对单元内传感器节点通信量的影响($k=10$, $t=10$)

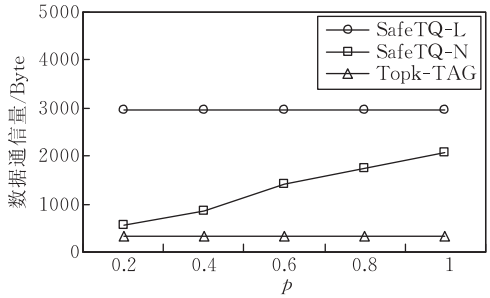

图 8 概率 p 对单元外部通信量的影响($k=10$, $t=10$)

7 总 结

目前传感器网络中隐私保护精确 Top-k 查询协议研究已成为富有挑战性的研究问题. 本文提出了一种两层传感器网络中可验证隐私保护 Top-k 查询协议 SafeTQ. 根据使用的完整性验证模式不同，SafeTQ 分为 SafeTQ-L 和 SafeTQ-N. 对 SafeTQ 的安全性和有效性评价总结如下：

(1) SafeTQ 使用随机数扰乱、加密和高资源节点之间安全计算第 k 位数据值策略，首次实现了在不泄漏隐私信息的情况下，精确地完成传感器网络 Top-k 查询，并能够对完整性攻击进行有效检测. 利用两层传感器网络中高资源节点的资源和通讯优势，使其承担安全计算第 k 位数据值任务，减少传感器节点能量消耗，增强 SafeTQ 有效性和可行性.

(2) 根据安全性分析，SafeTQ 在完成 Top-k 查询过程中，单元头节点和辅助计算节点不能获得定义 1 中描述的 3 类 Top-k 查询敏感信息；SafeTQ-L 检测出破坏完整性攻击的概率高于 SafeTQ-N.

(3) 从真实数据仿真实验结果看，在单元内传感器节点通信量方面，SafeTQ-L、SafeTQ-N 通信量比没有隐私保护的 Topk-TAG 的通信量增加的倍数，远小于文献[1]中使用隐私保护技术的 CPDA、SMART 通信量比 TAG 通信量增加的倍数；在单元外通信量方面，在本实验设定的参数范围内，SafeTQ-L、SafeTQ-N 和 Topk-TAG 通信量都小于 5000 bytes；SafeTQ-N 平均通信量高于 SafeTQ-L.

致 谢 感谢本文审稿专家和编辑老师所提出的宝贵意见和建议！

参 考 文 献

[1] He W, Liu X, Nguyen H, Nahrstedt K, Abdelzaher T. 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

[2] Zhang W S, Wang C, Feng T M. GP^{2}S: Generic privacy-preserving solutions for approximate aggregation of sensor data//Proceedings of the 6th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom). Hong Kong, China, 2008: 179-184

[3] Groat M M, He W B, 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

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

[5] Shi Jing, Zhang Rui, Zhang Yan-Chao. Secure range queries in tiered sensor networks//Proceedings of the 28th IEEE International Conference on Computer Communications (INFOCOM). Rio de Janeiro, Brazil, 2009: 945-953

[6] Chen Fei, Liu Alex 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

[7] Zhang Rui, Shi Jing, Liu Yun-Zhong, Zhang Yan-Chao. Verifiable fine-grained Top-k queries in tiered sensor networks//Proceedings of the 29th IEEE International Conference on Computer Communications (INFOCOM). San Diego, USA, 2010: 1199-1207

[8] Zhou Shui-Geng, Li Feng, Tao Yu-Fei, Xiao Xiao-Kui. Privacy preservation in database applications: A survey. Chinese Journal of Computers, 2009, 32(5): 847-861 (in Chinese)
(周水庚，李丰，陶宇飞，肖小奎. 面向数据库应用的隐私保护研究综述. 计算机学报，2009，32(5)：847-861)

[9] Silberstein A, Braynard R, Ellis C, Munagala K, Yang J. A sampling-based approach to optimizing Top-k queries in sensor networks//Proceedings of the 22nd International Conference on Data Engineering (ICDE). Atlanta, USA, 2006: 68-78

[10] Wu Min-Ji, Xu Jian-Liang, Tang Xue-Yan, Lee Wang-Chien. Top-k monitoring in wireless sensor networks. IEEE Transactions on Knowledge and Data Engineering (TKDE), 2007, 19(6): 962-976

[11] Vlachou A, Doulkeridis C, Nørvåg K, Vazirgiannis M. On efficient Top-k query processing in highly distributed environments//Proceedings of the 28th ACM SIGMOD International Conference on Management of Data (SIGMOD). Vancouver, Canada, 2008: 753-764

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

[13] Vaidya J, Clifton C. Privacy-preserving Top-k queries//Proceedings of the 21st International Conference on Data Engineering (ICDE). Tokyo, Japan, 2005: 545-546

[14] Yao Andrew Chi-Chih. How to generate and exchange secrets//Proceedings of the 27th IEEE Symposium on Foundations of Computer Science. Toronto, Canada, 1986: 162-167

[15] Yang Geng, Wang An-Qi, Chen Zheng-Yu, Xu Jian, Wang Hai-Yong. An energy-saving privacy-preserving data aggregation algorithm. Chinese Journal of Computers, 2011, 34(5): 792-800 (in Chinese)
(杨庚，王安琪，陈正宇，许建，王海勇. 一种低耗能的数据融合隐私保护算法. 计算机学报，2011，34(5)：792-800)

FAN Yong-Jian, born in 1978, Ph. D. candidate, lecturer. His research interests include wireless sensor network, privacy preservation and database.

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

Background

Privacy preservation in Wireless Sensor Networks (WSNs) is necessary for some real-world applications of WSNs, which has attracted more and more attentions. The existing research on data privacy preservation in WSNs mainly focuses on data aggregation and range query in WSNs. As Top-k query needs global information and resource of sensor is constrained, answering accurately Top-k query in WSNs while preserving data privacy is a challenge. Research on privacy-preserving Top-k query in WSNs is still relatively few.

This paper presents a verifiable privacy-preserving Top-k query protocol in two-tiered sensor networks (SafeTQ). SafeTQ can complete accurately Top-k query in two-tiered sensor networks while preventing attackers from gaining sensor data and enable sink to verify the authenticity and completeness of Top-k query results. Theoretical and quantitative results confirm the high efficacy and efficiency of SafeTQ.

This research is supported by the National Natural Science Foundation of China under grant No. 61070056 and No. 61075053.