一种带完整性验证的数据聚集隐私保护算法

石鲁生[1,2] 秦小麟[1]
(南京航空航天大学计算机科学与技术学院 南京 210016)[1] (宿迁学院计算机科学系 宿迁 223800)[2]

摘 要 为使无线传感器网络可以真正满足大规模应用的需求,提出了一种既能保护数据隐私又能验证数据完整性的聚集算法。算法首先构造不相交聚集树,然后让节点在各自对应的时间片内,按不同度数将自身数据分解为数个切片,并将切片分别加密传输至各聚集树中,达到保护节点数据隐私和获取冗余数据的目的,最后采用基于路由树的网内聚集将各聚集树的聚集结果传送至基站,由基站验证最终结果的完整性。仿真实验表明,在资源受限特征突出的无线传感器网络中,算法能够以较低的通信开销获得较高准确度的聚集结果,并具备较好的隐私保护性能和鉴别聚集结果完整性的能力。

关键词 无线传感器网络,数据聚集,隐私保护,完整性
中图法分类号 TP393 **文献标识码** A

Privacy-preserving Data Aggregation Algorithm with Integrity Verification

SHI Lu-sheng[1,2] QIN Xiao-lin[1]
(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)[1]
(Department of Computer Science, Suqian College, Suqian 223800, China)[2]

Abstract In order to satisfy the needs of large-scale applications for wireless sensor networks, this paper proposed an aggregation algorithm which can protect the data privacy and verify integrity. First, the algorithm constructs disjoint aggregation trees, and the nodes divides their own data into several slices in each corresponding time slices according to their different degrees, and then the data slices are encrypted and transmitted in every aggregation trees, so privacy-preserving and data integrity can be addressed. Finally, using in-network aggregation based routing tree, the final result of each tree is sent to the base station, and the integrity of the final result is verified by the base station. Simulation results show that the algorithm can obtain higher data aggregation accuracy with lower communication overhead, has better privacy-preserving performance and the aggregation results integrity.

Keywords WSN, Data aggregation, Privacy-preserving, Integrity

1 引言

无线传感器网络(Wireless Sensor Network,简称 WSN)由散布在空间的大量传感器节点构成,它们自组织成一个多跳网络[1]。作为物联网的重要组成部分,无线传感器网络的基本功能就是通过传感器节点收集并返回监测信息。

目前无线传感器网络的应用领域已快速扩展至军事、航空、反恐、防爆、救灾、环境、医疗、保健、家居、工业、商业等,但由于其常部署于无人值守、条件恶劣的环境当中,加之其以数据为中心、自组织、多跳和无线传输等特性,使得网络中的数据面临着严重的安全威胁,主要包括保护数据隐私和数据完整性验证两个方面。如果传感器节点的数据隐私不能得到保护,用来进行数据收集和处理的无线传感器网络应用系统就不能被部署,而如果所收集数据的完整性得不到验证,将没有用户信任和使用收集到的信息,系统的部署和数据的收集将毫无意义,因此二者缺一不可。

2 相关工作

在典型的无线传感器网络中,传感器节点通常是资源和能量受限的,数据聚集技术是有效降低网络中巨大通信量,从而减少能量消耗的重要手段。设计并使用合理的数据聚集算法,将可以进行高效的数据处理和收集,从而降低通信、计算开销,延长网络的生命周期。关于此方面的研究已经很多,文献[2-4]等皆属此类,但是这些研究成果大都有一个共同的前提,即无线传感器网络的所有传感器节点都是可信任的,并且其间的所有通信都是安全的。

现实环境中我们经常需要在一个不受控制的复杂环境中部署无线传感器网络,加之其无线传输特性,对手很可能窃听

到稿日期:2013-01-18 返修日期:2013-07-17 本文受国家自然科学基金项目:具有可生存能力的安全 DBMS 关键技术研究(60673127),国家高技术研究发展计划(863计划)基金项目:基于网格的数据可靠存储与容侵关键技术(2007AA01Z404),江苏高校优势学科建设工程项目物联网与控制技术,江苏省宿迁学院重点科研基金项目:无线传感器网络中数据隐私保护技术的研究(2012KY14)资助。
石鲁生(1978—),男,硕士,副教授,主要研究方向为安全数据库、无线传感器网络,E-mail:sqcsls@126.com;秦小麟(1953—),男,教授,博士生导师,主要研究方向为安全数据库、空间/时空数据库、分布式环境数据管理与安全等。

节点间的通信，甚至捕获、操纵关键节点。面对实际使用过程中的数据安全问题，基本的数据聚集算法无能为力，因此研究带隐私保护的数据聚集算法成为热点。由于聚集节点在数据聚集过程中处于承上启下的关键位置，目前大多以它受到攻击来展开数据聚集隐私保护算法。目前的研究成果主要有3类实现策略[5]：文献[1,6-8]使用逐跳加密机制，可以得到精确的SUM聚集结果，但通信和计算开销较大；文献[9-13]使用端到端加密机制，利用同态加密算法实现加密数据聚集，能较好地应付各种攻击，并减少了通信、计算开销，但仅支持SUM聚集；文献[14,15]使用非加密策略，通过添加伪装数据或使用数据扰动等技术实现数据聚集中的隐私保护，可以较好地完成非线性聚集操作，通信、计算开销也较小，但在隐私保护能力或聚集精确度上有所欠缺。

除保护数据隐私以外，在无线传感器网络的数据安全性问题中，鉴别数据的完整性也是重要的一环。B. Przydatek等人的SIA[16]算法采用随机抽样机制和互动验证来实现完整性，Yang等人的SDAP[17]算法则使用了分而治之和承诺与认证原则，但是由于它们都是采用基于统计学的抽样检测，很难察觉中间聚集结果的轻微改变，基站获得的最终聚集结果并不一定准确，并且额外增加了很大的通信开销。在使用逐跳加密机制的隐私聚集方案中，He等人提出了iPDA[1]与iCPDA[18]算法，它们分别在文献[6]中提出的SMART与CPDA两种数据聚集隐私保护算法的基础上增加了完整性验证的内容，尝试了在保护数据隐私的同时进行完整性验证，代价同样是通信开销进一步增大。当无线传感器网络中节点数目增大到一定数量时，其为验证完整性而额外增加的通信开销甚至有可能基本抵消了数据聚集所节省的通信开销。

由此可见，如何设计一个带完整性验证和隐私保护的聚集算法而又不增加太多额外的开销是一个值得关注的问题。

3 隐私保护与完整性验证的关键技术

3.1 基于节点度数的加密数据切片技术

在典型的基于数据切片的数据聚集隐私保护算法SMART中，所有节点的隐私数据 d_i 随机切分成 J 片（J 片数据的总和即为 d_i），节点自己保留一片，剩余的 $J-1$ 片都将在加密后进行传输。显然与传统直接加密节点隐私数据的传输方式相比，这种方式增强了隐私保护性能，但同时大大增加了通信开销，无线信道中发生碰撞的几率也增大了。解决这个问题的关键在于减少需要传输的数据切片数量。

为既达到隐私保护的目的，又不增加太多通信开销，我们使用基于节点度数的加密数据切片技术。首先确定一个加密数据切片发送和接收工作结束后每个节点度的最小值 $MinD$，然后让不同度数的节点在相应的时间片内根据自身度数进行数据切片和加密传输。若一个节点的度为 k，则它将在第 k 个时间片内，将自身数据分解为 $MinD-k$ 个数据切片，并进行加密传输。那些度数大于 $MinD$ 的节点将在最后一个时间片将自身的隐私数据作为一个数据切片加密传输。显然这样需要传输的数据切片数量将大大减少，加之我们使相应节点在每发送或接收一个加密数据切片后度数加1，又进一步降低了通信开销。同时不同度数节点的数据切片被限定在不同时间片内加密传输，既保护了数据隐私又降低了数据在无线信道中发生碰撞的几率。

3.2 基于数据冗余的完整性验证技术

数据聚集后的结果是用户决策判断或制定应对方案的直接重要依据，一旦其完整性遭到破坏，会使后续一系列措施的效果将大打折扣，甚至失效。完整性的验证还可以帮助判断关键节点是否被对手捕获、数据是否被篡改、数据通信过程中是否出错等等，因此它在数据聚集过程中同样占有重要地位。然而在数据聚集过程中将隐私保护和完整性验证有效结合起来并非易事。因为保护数据隐私时总是设法通过加密、扰动等技术极力隐藏数据，如将节点数据切片并加密传输等，与此相反，完整性验证则希望通过对数据的公共访问或同行监督来鉴别其完整性。

完整性的验证必定带来额外的开销，关键是这种开销不能太大。通过构造不相交聚集树，基站即可从多棵聚集树获取冗余数据并验证聚集结果的完整性。

在iPDA算法中，不相交聚集树的构造由于受到聚集节点的限制，网络中总会存在一些不能被聚集树覆盖的节点，即它们无法参与聚集，因此聚集的精度和完整性验证并不一定能得到保证。而我们首先构造不受聚集节点限制的多棵不相交聚集树，它们以基站为共同根节点，同时对网络中节点的覆盖率达100%。然后让每个传感器节点利用基于节点度数的加密数据切片技术将其隐私数据切片，并分别加密发送给多棵树的相应节点，使各树得到的总数据相等。加密数据切片的传输过程全部结束后，这些不相交聚集树完成各自的数据聚集，将最终结果上传至基站，由此基站获得了冗余的多个聚集结果。由于没有哪个节点（基站除外）同时属于任意两棵不同聚集树，一旦攻击者篡改或丢弃了某节点信息，各聚集树必将得到不同的聚集结果，而且越靠近聚集树根节点的节点信息被攻击，聚集结果的不一致性越明显。基站通过比较多个聚集结果来验证最终结果的完整性，只有那些通过完整性验证的聚集结果才能被基站所接受。构造了全覆盖不相交聚集树为保证聚集精度和有效的完整性验证打下了坚实的基础，采用基于节点度数的加密数据切片技术向不同聚集树传输数据时，虽然通信开销有所增加，但不至于过大，仍在一个可以接受的范围之内。

4 带完整性验证的数据聚集隐私保护算法

本文利用基于路由树的网内聚集办法，通过不相交聚集树获取的冗余数据来验证聚集结果的完整性，以不同度数节点加密传输不同数量的数据切片来保护数据隐私，设计了一种带完整性验证的数据聚集隐私保护算法IPSMA（An Integrity-Protecting Private Slice-Mix-Aggregate），它能以较低的开销在聚集的同时实现隐私保护和完整性验证。

在带完整性验证的数据聚集隐私保护算法IPSMA中，无线传感器网络使用一个连通图 $G(V,E)$ 来表示。其中 V 是顶点集合，每个 $v(v\in V)$ 代表网络中的一个传感器节点，网络中传感器节点的数量 N 表示为 $N=|V|$；E 是边的集合，每个 $e(e\in E)$ 代表网络中的一个无线通信链路，只要两个传感器节点可以直接通信，它们在图中就被一条边相连。

无线传感器网络中的节点共分3种类型：基站、聚集节点和叶子节点。基站是应答查询的节点，最终的聚集结果将传

送至此。在一般的数据聚集算法中，基站通常被构造为聚集树的根节点。本文只考虑单一基站的情况。除了基站即根节点外，所有的非叶子节点都是聚集节点，它们负责转发基站的查询、聚集数据并上报结果。叶子节点只负责采集数据并上报。网络中所有传感器节点都可以作为聚集节点或叶子节点。

在拥有 N 个节点的无线传感器网络中，定义普通数据聚集函数为：$y(t)=f(d_1(t),d_2(t),\cdots,d_N(t))$，其中 $d_i(t)$ 表示某传感器节点 i 在 t 时刻采集到的数据。由于 *sum* 函数在使用时没有太多的限制条件，而且其他很多聚集函数如：*average*、*max*、*count*、*variance* 等均可化简为 *sum* 函数[1]，故我们只关注 *sum* 函数，即 $y(t)=\sum_{i=1}^{N}d_i(t)$。

IPSMA 算法由 3 部分组成：不相交聚集树的构建、保护数据隐私、数据聚集与完整性验证。

4.1 不相交聚集树的构建

令 m 为不相交聚集树的棵数，显然有 $m\geqslant 2$。为说明方便，令 $m=2$，并将两棵不相交聚集树分别称为聚集树甲和聚集树乙，简称甲树和乙树。$m>2$ 的情况很容易据此扩充。

为达到验证聚集结果完整性的目的，甲树和乙树虽然在网络中相互交织在一起，但却不能相交，即必须保证除根节点外的任一节点不能既属于甲树又属于乙树。因此在传感器网络中，除基站节点作为甲树和乙树共同的根节点外，其他任一节点能且仅能成为甲、乙两树中某一棵树的节点。设某一非基站节点成为甲树节点的概率为 $p_{甲}$ $(0\leqslant p_{甲}\leqslant 1)$，成为乙树节点的概率为 $p_{乙}$ $(0\leqslant p_{乙}\leqslant 1)$，显然 $p_{甲}+p_{乙}=1$。在理想情况下，各节点成为甲树和乙树节点的概率相等，均为 50%，即 $\sum_{i=1}^{N}p_{甲}=\sum_{i=1}^{N}p_{乙}=\frac{N}{2}$，此时甲、乙两树在网络中呈均匀分布。在实际构造过程中我们也尽可能地追求均匀分布。

甲、乙两棵不相交聚集树的构造过程首先由基站向其邻接节点发出一个 HELLO 消息开始。一个节点接收到第一个 HELLO 消息后会等待一段时间以接收足够多的 HELLO 消息，以便决定自己应该成为甲树节点还是乙树节点。如果它从甲树接收到的 HELLO 消息超过乙树，它很可能选择成为乙树节点，反之，则可能成为甲树节点，以此来平衡整个网络中甲、乙两树的节点分布。由此可以分别确定 $p_{甲}$ 和 $p_{乙}$：

$$p_{甲}=\frac{H_{乙}}{H_{甲}+H_{乙}},p_{乙}=\frac{H_{甲}}{H_{甲}+H_{乙}} \tag{1}$$

式中，$H_{甲}$ 和 $H_{乙}$ 分别为节点从甲树和乙树接收到的 HELLO 消息的数量，且 $H_{甲}\neq 0$、$H_{乙}\neq 0$。注意式(1)中 $H_{甲}$ 和 $H_{乙}$ 值同时为 0 的情况是不存在的，因为 $H_{甲}=H_{乙}=0$ 说明节点没有收到任何 HELLO 消息，即该节点没有邻接点，显然正常情况下，网络中不存在这样的节点，即不相交聚集树将对网路中的所有节点全覆盖。但是当网络中节点密度较低时，$H_{甲}$ 或 $H_{乙}$ 单独为 0 的情况是可能存在的，此时说明节点没有收到甲树或乙树的 HELLO 消息，因此它只能成为乙树或甲树节点。如果这种节点太多，网络中甲、乙两树节点的均匀分布就可能遭到破坏，为减少该情况的发生，我们可以适当增加网络中节点的密度。

节点在确定成为甲树节点或乙树节点后，将向其父节点发送消息以确定树形结构，同时父节点收到消息后使自己的度数加 1。基站作为甲、乙两树共同的根节点将分别记录其在甲、乙两树中的度数，其他节点记录其在自己所属聚集树中的度数。

4.2 保护数据隐私

设传感器节点产生的随机数 r 在(−1,1)区间内均匀分布，节点在网络中的实测数据值在区间$[V_{\min},V_{\max}]$范围内。所有数据聚集工作将在基站明确的查询周期(epoch duration)ED 内完成。节点在 ED 周期内分配一个传输时延(transmission duration)TD，用于发送和接收加密数据切片，分配一个聚集时延(aggregation duration)AD，用于沿路由树完成聚集。显然 TD 和 AD 的取值与 $MinD$ 和聚集树的最大高度 MAX_H 值的大小有关。其中 $MinD$ 对 TD 值的影响较大，MAX_H 对 AD 值的影响较大，由此得到：

$$TD=\frac{ED\cdot MinD}{MinD+MAX\_H},AD=\frac{ED\cdot MAX\_H}{MinD+MAX\_H}$$

在传输时延 TD 中主要完成数据切片和加密传输的工作。首先为网络中的每个节点 $i(i=1,2,\cdots,N)$ 在 h 跳内随机从甲、乙两树分别选取节点集 $S_{甲i}$ 和 $S_{乙i}$，且 $|S_{甲i}|=|S_{乙i}|=MinD$，注意 $S_{甲i}$ 中的节点全部为甲树节点，而 $S_{乙i}$ 中的节点全部为乙树节点。将传输时延 TD 平均分配给 $MinD+1$ 个时间片，则有时间片 $t_k=\frac{TD}{MinD+1}(k=0,1,2,\cdots,MinD)$。

算法 1 描述了甲树节点进行数据切片和加密发送的过程，乙树类似。

算法 1 甲树节点进行数据切片和加密发送

1. for t_k($0\leqslant k\leqslant MinD-1$) 中的每个 t_k
2. for 甲树中度为 k 的每个节点 i//$D_i=k$
3. $DC_{i\_backup}=DC_i$
   /* DC_i为节点 i 自身记录的数据(若节点 i 未发送或接收过数据切片，DC_i即为原始读数 d_i，否则不是)；DC_{i_backup}是 DC_i的备份，为向乙树节点发送数据切片做准备 */
4. while($D_i<MinD$)
5. 从$S_{甲i}$中顺序选择一个节点 $j_{甲}$，从 $S_{乙i}$中顺序选择一个节点 $j_{乙}$
6. $DS_{ij_{甲}}=\frac{r\cdot(V_{max}-V_{min})}{MinD}$//从 DC_i上切下一片数据 $DS_{ij甲}$
7. if ($D_i<MinD-1$) $DS_{ij_{乙}}=\frac{r'\cdot(V_{max}-V_{min})}{MinD}$
   /* 从 DC_{i_backup}上切下一片数据 $DS_{ij_{乙}}$，r'为不同于 r 的另一随机数 */
8. else $DS_{ij_{乙}}=DC_{i\_backup}$
   /* 将 DC_{i_backup}直接作为数据切片 $DS_{ij_{乙}}$ */
9. 将 $DS_{ij甲}$ 加密发送给节点 $j_{甲}$，将 $DS_{ij_{乙}}$ 加密发送给节点 $j_{乙}$
10. $D_i=D_i+1$
11. $DC_i=DC_i-DS_{ij甲}$，$DC_{i\_backup}=DC_{i\_backup}-DS_{ij_{乙}}$
12. endwhile
13. endfor
14. endfor
15. 在时间片 t_{MinD}，甲树中所有未进行数据切片的节点 i 将自身数据加密发送至 $S_{乙i}$中任一节点

与加密切片的发送工作相比，加密切片的接收工作简单了很多。算法 2 描述了甲树节点接收加密数据切片的过程，乙树同理。

算法 2 甲树节点接收加密数据切片

1. for t_k($0\leqslant k\leqslant MinD$) 中的每个 t_k

2. for 每个接收到的加密数据切片

3. 利用接收节点 j 和发送节点 i 的共享密钥将其解密得 DS_{ij}

4. $DC_i = DC_i + DS_{ij}$

5. if(节点 i 是甲树节点) $D_i = D_i + 1$

6. endfor

7. endfor

如此,传输时延 TD 结束后,甲、乙两树中所有节点的度均不小于 $MinD$。

4.3 数据聚集与完整性验证

在保护数据隐私阶段完成后,网络中各节点的数据已经改变,IPSMA 算法将采用网内聚集方法沿着已经构造好的甲、乙两棵不相交聚集树完成最后阶段的数据聚集。

设基站从甲、乙两树得到的最终聚集结果分别为 $R_{甲}$ 和 $R_{乙}$。显然在没有数据丢失的理想情况下,应有 $R_{甲}=R_{乙}$。然而在现实环境中,数据丢失是无法避免的,因而 $R_{甲}$ 和 $R_{乙}$ 不可能完全相同,但在数据没有受到恶意攻击和破坏的情况下,两者不应相差过大。实际使用时,根据无线传感器网络部署的具体环境及应用系统需要,我们可以设定一个阈值 Th,如果 $|R_{甲}-R_{乙}|<Th$,基站将接受最终聚集结果,否则,认为数据完整性已遭破坏,拒绝它。

4.4 算法示例

设某无线传感器网络中的节点分布情况如图 1 所示,每个圆标记网络中的一个节点,其中灰色为基站,其余为一般传感器节点。图 2 为构造甲、乙两棵不相交聚集树之后的情况,其中灰色基站节点是甲、乙两树共同的根,白色为甲树节点,黑色为乙树节点。甲、乙两树中每个节点的度数 D(基站在甲、乙两树中的度数分别为 3 和 2)如图 2 所示。

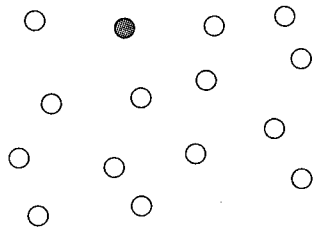

图 1 原始无线传感器网络

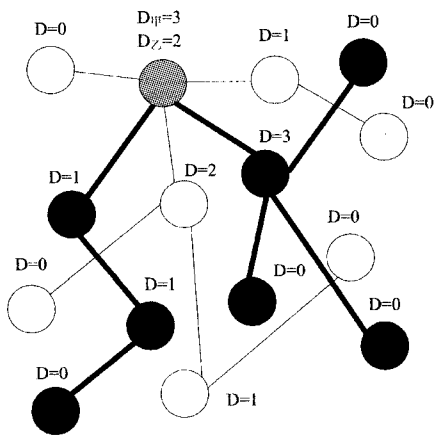

图 2 构建甲、乙两棵不相交聚集树

以图 2 为例,令 $MinD=3$,$h=1$。在 t_0 时间片内,甲树中度为 0 的节点的数据切片发送过程如图 3 所示,其中带箭头的实线和虚线分别表示加密数据切片在甲树和乙树中的传输方向,下同。图 4 则为 t_2 时间片内,甲树中度为 2 的节点的数据切片发送过程。显然在本例中,t_2 时间片后甲树中所有节点的度已均不小于 $MinD$。

当最后一个时间片 t_3 中的加密数据切片和传输工作完成后,整个传输时延 TD 结束。在接下来的聚集时延 AD 中,甲、乙两树将沿各自路由完成最终聚集,将结果 $R_{甲}$ 和 $R_{乙}$ 上传至基站,由基站根据 $|R_{甲}-R_{乙}|$ 与 Th 的比较结果,最终验证聚集结果的完整性。

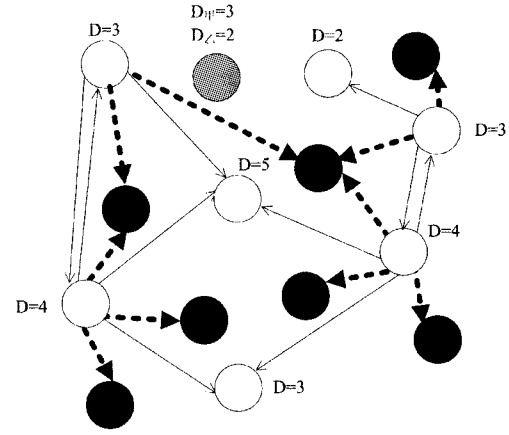

图 3 t_0 时间片内甲树节点的数据切片发送过程

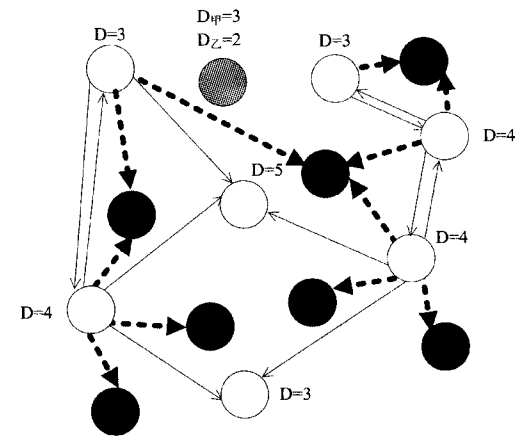

图 4 t_1 时间片内甲树节点的数据切片发送过程

5 性能分析

设计 IPSMA 算法的目的是在保护数据隐私和进行完整性验证的基础上得到准确的聚集结果。由于无线传感器网络资源受限的特点突出,我们还希望尽可能减少能量消耗,尤其是通信开销。本节通过仿真实验对 IPSMA 算法的性能从准确性、隐私保护、完整性和通信开销等 4 个方面加以分析,并与 TAG、SMART、iPDA 等算法进行比较。

我们以 TAG 算法为基础,在 TOSSIM 仿真环境中执行 IPSMA 算法,实验中,500 个传感器节点被随机地部署在一个 400m×400m 的区域内,节点的有效传输距离为 50 米,传输速率为 1Mbps。此时网络中节点度的平均值已达 23.5[1],节点密度较大,因此我们可在 IPSMA 算法中取 $h=1$,即每个节点 $i(i=1,2,\cdots,500)$ 将在一跳范围内随机从甲、乙两树分别选取节点集 $S_{甲i}$ 和 $S_{乙i}$。

5.1 准确性

准确性是一切聚集算法的出发点,失去准确性,算法将失去意义。我们用聚集算法所得结果与节点实测数据之和的比值来衡量准确性,记为 AR。在网络中没有数据丢失的理想

情况下，IPSMA 和 TAG 显然都可得到完全准确的结果，即 $AR=1$。

但在真实场景下，数据丢失是无法避免的。IPSMA 中的数据丢失主要表现在两个方面，一是因时间片用完而导致的加密数据切片无法发送或接受，二是由于无线信道内的碰撞而造成的数据丢失。这些问题显然可以通过增加查询周期 ED 得以解决。

在仿真环境下我们对 TAG 和 IPSMA（$MinD$ 分别取值 1、2、3）进行了 30 次模拟，其均值的统计结果如图 5 所示。可以看出 IPSMA 算法的准确度与 TAG 相仿，基本趋势一致，波动较为明显，且对 $MinD$ 值并不敏感。在 ED 值较小时，两个算法的准确度都不高，随着 ED 值不断增大至 20s 以上，精确度基本可以稳定在 85%～90%。

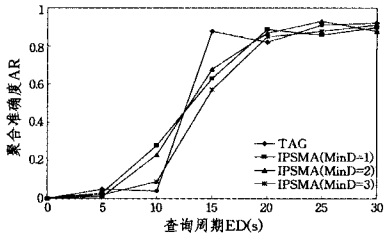

图 5　TAG 和 IPSMA 算法准确度比较

从理论上分析，与同样实现隐私保护和完整性验证的 iPDA 算法相比，IPSMA 算法中不能参加聚集的节点是不存在的，数据切片的发送和接收在指定时间片内进行，避免了一些碰撞，这些都减少了数据的丢失，而数据切片时采用在(－1，1)区间内均匀分布的随机数 r，则进一步降低了数据丢失对精度的影响。因此，在相同的外部条件下，IPSMA 算法的准确性将优于 iPDA 算法。

5.2　隐私保护

在 IPSMA 和 SMART 中隐私保护都是通过节点数据的切片和加密传输来实现的。我们定义 $P(q)$ 为节点数据隐私被泄露的概率，q 为节点间通信链路被破解的概率。在 SMART 算法中，若一个传感器节点将自身数据切成 J 片，近似有，$P(q)=q^{J-1}\sum_{k=0}^{d_{\max}}P(in\_degree=k)q^k$，其中 $d_{\max}$ 表示网络中节点入度的最大值，$P(in\_degree=k)$ 是网络中节点入度为 k 的概率[6]。在 IPSMA 中，显然 $MinD$ 起到了和 J 类似的作用，即只有当节点的所有链路（包括加密数据切片的发送和接收两部分，其总数大于等于 $MinD$）都被破解，节点的真实数据才能被泄露。因此 $P(q)$ 可以近似为：

$$P(q)=\sum_{k=MinD}^{d_{\max}}P(D=k)\cdot q^k \tag{2}$$

式中，$d_{\max}$ 表示树中节点度的最大值，$P(D=k)$ 是树中节点度 $D=k$ 的概率。

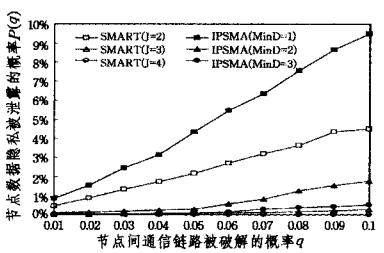

图 6　SMART 和 IPSMA 算法隐私保护性能比较

SMART 算法和 IPSMA 算法的隐私保护性能如图 6 所示。由图 6 可以发现，两种算法节点数据隐私被泄露的概率 $P(q)$ 在 J 和 $MinD$ 取不同值时的趋势完全相同，即随 q 值的增大而增大，但在 $MinD$ 与 $J-1$ 值相等的情况下，由于 SMART 中所有传感器节点的数据被切成 J 片，然后加密传输，隐私保护性能较好；而 IPSMA 中度高于 $MinD$ 的节点并不进行数据切片，进行数据切片的节点，其数据被实际分解成的数据切片数也与节点度 D 有关，为 $MinD-D$ 片，这些在降低 IPSMA 算法通信开销的同时却也使其隐私保护性能略低于 SMART，从图 6 中不难发现这一点。在具体应用场景中，如要求数据隐私被泄露的概率不能高于 2%，两者均可通过选择适当的 $MinD$ 或 J 轻松满足之。

5.3　完整性

聚集结果完整性的主要威胁来自于那些篡改节点信息的攻击行为。这些攻击的目的是让基站接受错误的聚集结果，从而做出不合适甚至错误的决定。如果被攻击的是一个距离聚集树根节点很近的非叶子聚集节点，那么情况将相当严重。

IPSMA 算法是通过构造 m 棵不相交聚集树，利用所获取的冗余数据来进行完整性验证的。由于做到了不相交聚集树对网络中所有节点的全覆盖，使得聚集结果的完整性在结构上得到了保障。如果某个节点的数据被攻击者篡改或丢弃，并使其所在树的最终聚集结果不同于其他树，基站一旦发现这种聚集结果间的差异，遭受攻击的聚集结果将被拒绝。只有那些通过完整性验证的聚集结果才能被基站所接受。

5.4　通信开销

与 TAG 算法不同，在 IPSMA 算法中无论是增加的隐私保护功能还是完整性验证功能都会带来额外的通信开销，关键问题是这种开销不能太大。

我们以传输信息的数量来比较不同算法的通信开销。在 TAG 中，每个节点需要发送两条信息来完成数据聚集，传输信息的总数为 $2N$。在 IPSMA 中，每个节点同样需要发送类似的两条信息，一条是在不相交聚集树的构造阶段，另一条是在最后的数据聚集阶段，同样需传输 $2N$ 条信息。除此之外在数据隐私保护阶段，节点还需要将加密的数据切片分别传输到不同的聚集树，以保护隐私和获取冗余数据验证完整性。仍以两棵不相交聚集树为例，记甲、乙两树中度为 0 的节点个数之和为 N_0，度为 1 的节点个数之和为 N_1，以此类推，得 N_{MinD-1}、N_{MinD} 直至 $N_{d_{\max}}$。由于无论节点发送加密数据切片到同树节点，还是接收同树节点发送过来的加密数据切片，其相应的度数都会加 1，因此对甲树中度 $D=k(0\leqslant k<MinD)$ 的任一节点，其在甲树中实际发送的数据切片数小于等于 $MinD-k$，在乙树中实际发送的数据切片数与甲树一致，则该节点在保护数据隐私阶段发送的总数据切片数小于等于 $2(MinD-k)$。对于 $D\geqslant MinD$ 的节点，在保护数据隐私阶段，其只将自身数据加密发送一次。由此可得 IPSMA 算法中传输信息总数 $ComO$：

$$\begin{aligned}ComO&\leqslant 2N+2(MinD-0)\cdot N_0+2(MinD-1)\cdot N_1+\\&\quad\cdots+2(MinD-(MinD-1))\cdot N_{MinD-1}+(N-\\&\quad(N_0+N_1+\cdots+N_{MinD-1}))\\&=3N+\sum_{k=0}^{MinD-1}(2MinD-2k-1)\cdot N_k\end{aligned} \tag{3}$$

$MinD$ 为不同值时 IPSMA 算法的通信开销与 TAG 和 iPDA 算法的通信开销的比较如图 7 所示。从图 7 中可以看出，各算法的通信开销均与查询周期关系不大。由于增加了隐私保护和完整性验证，IPSMA 算法的通信开销与 TAG 相

比所增加，但与同样具备隐私保护和完整性验证功能的 iPDA 算法相比有明显降低。在 $MinD$ 与 $J-1$ 相等的情况下，由式(3)可知，IPSMA 算法的 $ComO$ 显然小于 iPDA 算法中总的信息传输量 $(2J+1)\cdot N$，反映在图 7 中，IPSMA 对应的通信开销比 iPDA 的减少了 30%～45%不等。较低的通信开销对资源受限的无线传感器网络来说意义重大，IPSMA 算法因为发送了较少的加密数据切片从而获得了较低的通信开销，代价则是隐私保护性能有所下降。

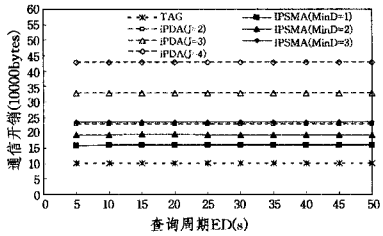

图 7　TAG、iPDA 和 IPSMA 算法通信开销比较

结束语　在无线传感器网络中收集数据时，数据聚集能有效地降低通信量，是减少能量消耗的重要技术手段。随着传感器网络应用领域的不断扩大，在收集数据时，如何保护隐私、验证最终聚集结果的完整性越发显得重要。然而在资源严重受限的无线传感器网络中，如何在既能保护数据隐私又能验证数据完整性的前提下完成数据聚集并非易事。

本文提出的 IPSMA 是一个在 WSN 中实现隐私保护和完整性验证的数据聚集算法。IPSMA 算法通过发送加密数据切片达到隐私保护的目的，利用不相交聚集树得到冗余数据以验证聚集结果的完整性。通过仿真实验，并与相关算法的比较表明，IPSMA 算法在聚集结果的准确性、数据隐私的保护、聚集结果的完整性和通信开销等各方面均有不错的表现。

IPSMA 算法只是我们的初步工作，还存在不少改进的空间，如将其扩展到其他聚集函数、报告篡改节点信息攻击行为的位置、进一步降低通信开销和提高隐私保护效力等等。

参 考 文 献

[1] He W, Nguyen H, Liu X, et al. iPDA: Anintegrity-protecting private data aggregation scheme for wireless sensor networcks [C]//Proceedings of the Military Communications Conference. SanDiego, CA, USA, 2008: 1-7

[2] Madden S, Franklin M J, Hellerstein J M. TAG: A Tiny Aggregation Service for Ad hoc Sensor Networks[C]//Proceedings of the 5th Symposium on Operating Systems Design and Implementation. NewYork, USA, 2002: 131-146

[3] Deshpande A, Nath S, Gibbons P B, et al. Cache-and-query for wide area sensor databases [C]//Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data. SanDiego, CA, USA, 2003: 503-514

[4] Tang X, Xu J. Extending network lifetime forprecision constrained data aggregation in wireless sensor networks[C]//Proceedings of the IEEE INFOCOM. Barcelona, Catalunya, Spain, 2006: 755-766

[5] 范永健，陈红，张晓莹. 无线传感器网络数据隐私保护技术[J]. 计算机学报，2012，35(6)：1131-1146

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

[7] Yao Jian-bo, Wen Guang-jum. Protecting classification priracy data aggregation in wireless sensor networks[C]//Proceedings of the 4th International Conference on Wireless Communication. Networking and Mobile Computing (WiCOM). Dalian, China, 2008: 1-5

[8] 杨庚，王安琪，陈正宇，等. 一种低耗能的数据融合隐私保护算法[J]. 计算机学报，2011，34(5)：792-800

[9] Joao G, Dirk W, Markns S. CDA: Concealed data aggregation for reverse muhicastt raffic in wireleass sensor networks[C]//Proceedings of the IEEE International Conference on Communications(ICC). Seoul, Korea, 2005: 3044-3049

[10] Castelluceia C, Mykletun E, Tsudik G. Efficient aggregation of encrypted data in wireless sensor networks[C]//Proceedings of the 2nd Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitoue). SanDiego, CA, USA, 2005: 109-117

[11] Feng Tai-ming, Wang Chnang, Zhang Wen-shang, et al. Confidentiality protection for distributed sensor data aggregation[C]//Proceedings of the 27th IEEE International Conference on Computer Communications(INFOCOM). Phoenix, USA, 2008: 56-60

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

[13] Papadopoulos S, Kiayias A, Papadias D. Secure and efficient in-network processing of exact SUM queries[C]//Proceedings of the 27th International Conference on Data Engineering(ICDE). Hannover, Germany, 2011: 517-528

[14] Zhang W S, Wang C, Feng T M. GP2S: Generic privacy-preserving solutions for approximate aggregation of sensor data[C]//Proceedings of the 6th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom). Hong Kong, China, 2008: 179-184

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

[16] Przydatek B, Song D, Perrig A. SIA: Secure information aggregation in sensor networks[C]//Proc. of the 1st Int'l Conf. on Embedded Networked Sensor Systems. New York: ACM Press, 2003: 255-265

[17] Yang Yi, Wang Xin-ran, Zhu Sen-cun, et al. SDAP: A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks [C]//ACM Int. l Symp on Mobile Ad Hoc Networking and Computing(MOBIHOC 2006). Florence, Italy, 2006: 356-367

[18] He W, Liu X, Nguyen H, et al. A cluster-based protocol to enforce integrity and preserve privacy in data aggregation[C]//Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops. Montreal, QC, Canada, 2009: 14-19