高效的抗合谋攻击的数据聚集协议

朱友文[1,2]，黄刘生[1,2]，杨威[1,2]，田苗苗[1,2]
（1. 中国科学技术大学 计算机学院，安徽 合肥 230026；2. 中国科学技术大学 苏州研究院，江苏 苏州 215123）

摘 要：考虑到如何评估各个保护隐私数据聚集协议的隐私保护效果，定义了保护隐私数据聚集协议的抗合谋攻击能力的测量模型，并设计了一种高效的抗合谋攻击的保护隐私数据聚集协议。理论分析和模拟实验证实了新协议的抗合谋攻击能力，对比结果显示在同等安全性要求下，新协议的通信量远小于 SMART 协议。

关键词：无线传感网络；保护隐私；合谋；数据聚集
中图分类号：TP309.2 **文献标识码**：A **文章编号**：1000-436X(2010)9A-0223-05

Efficient collusion-resisting data aggregation protocol

ZHU You-wen[1,2], HUANG Liu-sheng[1,2], YANG Wei[1,2], TIAN Miao-miao[1,2]
(1. School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China;
2. Suzhou Institute for Advance Study, University of Science and Technology of China, Suzhou 215123, China)

Abstract: A measurement model was defined to analysis each privacy-preserving data aggregation protocol's ability to resist potential collusion. Then, a new efficient collusion-resisting privacy-preserving data aggregation protocol was proposed. Theoretical analysis and experiment result confirmed the protocol's ability to resist potential collusion and the communication overhead of our scheme is much less than existing protocol's.

Key words: wireless sensor networks; privacy-preserving; collusion; data aggregation

1 引言

无线传感器网络是由大量具有无线通信能力的传感器节点组成，能够部署在广阔的区域中，执行数据收集、环境监测[1]、灾难救助[2]、目标跟踪[3]等任务。在这些应用中，人们经常需要聚集统计信息，而不是每个节点的传感数据。这时，就需要设计出保护隐私的数据聚集协议，使得在收集到统计信息的同时保护每个传感器节点的私有传感数据不被泄露。同时，在无线传感器网络中，多个恶意的节点也可能通过合谋的方式窃取其他节点的私有数据。如何抵抗合谋攻击也是保护隐私数据聚集协议面临的重要问题。

文献[4]首先提出了传感器网络中保护隐私的数据聚集问题，并提出了 2 个保护隐私的数据聚集方案：CPDA 和 SMART。它们用以聚集所有传感数据的总和，同时保证每个私有数据的安全。CPDA 协议首先将所有传感器节点划分为若干个簇，然后每个簇内部的传感数据的部分和聚集到簇头节点，最后簇头再将这些部分和汇集到 Sink 节点。这种方式下，每个簇头会得到一些额外的信息，簇内的合谋攻击就可以得到其他节点私密的传感数据；而且在

收稿日期：2010-07-10
基金项目：国家自然科学基金重大研究计划基金资助项目（90818005）；国家自然科学基金资助项目（60903217, 60773032）；中国博士后科学基金资助项目（20090450701）
Foundation Items: The Major Research Plan of the National Natural Science Foundation of China (90818005); The National Natural Science Foundation of China (60903217, 60773032); The China Postdoctoral Science Foundation Funded Project (20090450701)

协议执行过程中需要进行大量的矩阵运算，计算量过大。由于传感器具有能量受限、计算受限、通信受限和存储能力有限等特点[1]，因此，CPDA 并不适用于无线传感器网络。SMART 协议通过将每个传感数据随机划分为J（SMART 协议中的参数）个数据片，并将各个分片随机发给其他传感器节点，然后节点对收到的分片的和直接进行数据聚集。SMART 协议的计算量小，但是其通信开销过高，在保护隐私的传感数据时，发送了大量的冗余消息。

鉴于此，给出了抗合谋攻击的量化测量模型，并设计了一个新的保护隐私数据聚集协议。本文的主要贡献如下。

1) 首先给出了保护隐私数据聚集协议的抗合谋攻击能力的测量模型，用于分析测量保护隐私数据聚集协议的抵抗合谋攻击能力。

2) 本文提出了一种新的保护隐私的数据聚集协议，用于安全地计算传感数据总和。新协议的计算量较小，模拟实验室显示在同等安全性要求下新方案的通信量大约只有 SMART 协议的一半，是一种高效的抗合谋攻击的保护隐私数据聚集协议。

2 系统模型

2.1 网络模型

本文所考虑的无线传感器网络由一个 Sink 节点和N个传感器节点组成。整个网络可以看作一个连通的无向图 $G(V, E)$，其中 Sink 节点和各个传感器是图的节点，无线链接形成图的各个边。Sink 节点记为 v_0，N个传感器依次为 $v_1,v_2,\cdots,v_N$，即 $V=\{v_1,v_2,\cdots,v_N\}$，$|V|=N+1$。

在这里，记 $N_k(v_i)=\{v_j \mid v_j \in V$ 且 v_i 到 v_j 的最短路径为 $k\}$。那么，$N_1(v_i)$为 v_i 的邻居节点的集合。每个传感器节点测量的传感数据是 0 到 U_d 之间的一个整数，其他类型的数据（比如噪音、湿度等）也可以转换为整数再进行处理。

2.2 设计目标和安全模型

无线传感器网络中，保护隐私数据聚集的目标如下：

1) 每个传感器节点私密的传感数据不会被泄露给 Sink 节点、其他的传感器节点以及网络中可能的窃听者；

2) 聚集结果只能被 Sink 节点得到，对其他的传感器节点和窃听者都是保密的；

3) 协议执行过程中的通信量、计算量和需要存储空间应该尽量的小，以适应无线传感器节点资源有限的特点。

以往的方案[4,5]中假设 Sink 节点是可信的，这里，假设 Sink 节点和每个传感器网络都可能被俘获，Sink 节点也可能成为潜在的恶意攻击者。攻击者可能进行的攻击类型很多[6~8]，为了专注于数据聚集和保护隐私的问题，我们和文献[4, 5]考虑的攻击类型相同，即窃听者和被俘获的节点对传感数据的窃听和合谋，即他们只通过分析自己所收到和拥有的数据，试图得到其他节点的私密数据。

3 抵抗合谋攻击的数据聚集协议

3.1 抵抗合谋攻击的量化测量模型

为了对保护隐私的数据聚集协议的抗合谋攻击能力进行量化测量，给出了如下定义。

定义 1 设f为一个保护隐私数据聚集协议，传感器节点 v_i 持有私密的传感数据 n_i，如果至少 k 个节点（包括 Sink 节点和传感器节点）的合谋可以获取 n_i，那么称节点 v_i 的私有传感器数据 n_i 是 k-collusion 的，记节点 v_i 在协议f中的抗合谋攻击能力 $C(vi)=k$。

定义 2 设 f 为一个保护隐私数据聚集协议，其中持有私有数据的传感器节点集合为$\{v_1,v_2,\cdots, v_N\}$，那么协议f的抗合谋攻击能力 $C(f)$为

$$C(f)=\frac{1}{N}\sum_{i=1}^{N}C(v_i)$$

其中，$C(v_i)$为节点 v_i 在协议 f 中的抗合谋攻击能力。

可以看出，定义 1 反应了每个私有的传感数据在面临合谋攻击时的安全程度；定义 2 则评价了保护隐私数据聚集协议抵抗合谋攻击的平均效果。

通过文献[4]中的一个数据聚集例子来说明这 2 个定义。网络中包含 6 个传感器节点和一个 Sink 节点，执行 SMART 协议[4]进行数据聚集。其数据聚集过程如下：每个传感器节点将其所持有的私有传感数据随机分为 3 片，并将其中的 2 个随机分片分别发给 2 个随机选择的邻居节点（如图 1(a)所示，节点 v_i 指向 v_j 的箭头表示 v_i 向 v_j 发送了分片 d_{ij}，比如 v_1 分别发送了分片 d_{12} 和 d_{15} 到 v_2 和 v_5），自己保留余下的一个数据分片，即设定 SMART 协议中的参数 $J=3$，$h=1$。然后，节点 v_j 计算收到的所有分片 d_{xj}（x 表示其他节点的编号）和自己保留的分片 d_{jj} 的总和 $r_j=d_{jj}+\sum_{x\neq j}d_{xj}$。最后，利用基于树的路由

协议[9]，所有节点将 $r_i(i=1,2,\cdots,7)$聚集到 Sink 节点，使 Sink 节点得到和为

$$S=\sum_{i=1}^{7} r_i$$

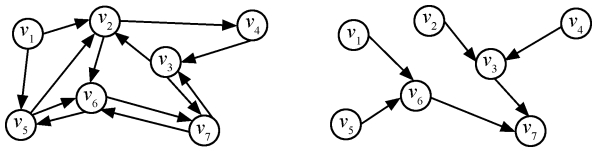


图 1　SMART 协议的一个例子

在图 1 的方案中，节点 v_i 在第一个阶段发送了 2 个数据 d_{ia} 和 d_{ib}（节点 v_a 和 v_b 分别收到了这 2 个数据），在第二个阶段发送了一个数据 r_i。为了得到 n_i，必须得到节点 v_i 所有收到和发出的数据，那么至少需要所有与节点 v_i 交互的节点合谋才能够窃取 n_i。由此，得到了每个节点的抗合谋攻击的能力，如表 1 所示。

表 1 图 1 中各个传感器节点的抗合谋攻击能力

节点	抗合谋攻击的能力 v_6
v_1	3
v_2	5
v_3	3
v_4	2
v_5	3
v_6	4

根据定义 2 可以得到，在图 1 的例子中 SMART 协议的抗合谋攻击能力为

$$C(\text{SMART})=\frac{1}{6}\times\sum_{i=1}^{6}C(v_i)$$
$$=\frac{3+5+3+2+3+4}{6}=3.333\,33$$

进一步，易于看出上述方案是存在冗余消息的。比如，在发送分片阶段，如果节点 v_5 不向节点 v_6 发送分片，而是只发送一个分片给节点 v_2，并不会降低任何节点的抗合谋攻击能力。当然，该实例所采用的 SMART 协议在执行过程中还发送其他多条冗余消息，这里不再详细分析。

为了避免发送冗余的无用消息，设计了一个新的保护隐私数据聚集协议，最大限度地提高每个扰动数据（随机数、随机分片等）在抗合谋攻击中的作用，提高了通信效率。

3.2　协议过程

每发送一个扰动数据，可以提高发送者和接受者双方的抵抗合谋攻击的能力，然而，如果他们之间再相互发送扰动数据，并不会提高任何一个节点的抗合谋攻击能力。因此，为了提高每个扰动数据对整个协议抗合谋攻击能力的贡献，在满足安全性前提下尽量少地发生扰动数据，就需要避免节点之间的重复发送。另一方面，减少每个扰动数据发送的跳数也可以降低系统的通信负载。据此，设计了一种新型的利用扰动数据的数据聚合协议。该方案共计有 4 个阶段：系统预备、发起查询、扰动阶段和聚集阶段，其具体过程如下。

1) 系统预备

文献[10]中的方案，每 2 个节点生成共享的对称加密密钥，用于保证接下来的通信过程中数据发送的安全。根据实际需求，各个节点协商选取节点抗合谋攻击能力下限L（L为一个正整数）。在协议执行完成之后，将保证每个节点的抗合谋攻击能力不低于L。

2) 发起查询

令 $s_0=0$，s_0 是 Sink 节点的私有数据。Sink 节点 v_0 发送到<*Query*, r_{0i}, *X*>每个邻居节点 v_i，并执行 $s_0=s_0+r_{0i}$，其中 r_{0i} 为 Sink 节点生成的随机数，用于数据扰动。

3) 扰动阶段

节点 v_i（$i=1,2,\cdots,N$）执行如下操作。

令 $R_i\neq\Phi$，$s_i=n_i$，若收到<*Query*, r_{ji}, *X*>，则执行 $R_i=R_i\cup\{v_j\}$，$s_i=s_i-r_{ji}$；

生成随机数 r_{ik}，发送<*Query*, r_{ik}, *X*>到 $v_k\in N_1(v_i)-R_i$，并计算 $s_i=s_i+r_{ik}$；

设置 $g=2$，$t=L-|N_1(v_i)\cup R_i|$，其中$|N_1(v_i)\cup R_i|$表示集合 $N_1(v_i)\cup R_i$ 中元素的个数，然后执行如下操作：

```
WHILE t>0 DO
    Q=N_g(v_i)−R_i;
    WHILE  Q≠Φ 且t > 0 DO
        随机选取 v_x ∈ Q;
        Q=Q−{v_x};
        生成随机数 r_ix;
        发送<Query, r_ix, X>  到 v_x;
        s_i=s_i+r_ix;
        t=t−1;
    ENDWHILE
    g=g+1;
ENDWHILE
```

4) 聚集阶段

与 SMART 协议的聚集阶段类似，利用基于树

的路由协议[9]，节点 v_i（i=0,1,2,…,N）聚集 s_i 到 Sink 节点，Sink 节点得到总和 $S=\sum_{i=0}^{N}s_i$。

3.3 协议分析

上述协议执行过程中，对于每个用于数据扰动的随机数 r_{ik}，节点 v_i 和节点 v_k 分别执行了 s_i=s_i+r_{ik} 和 s_k =s_k−r_{ik}。而且 s_i 的初始值为 n_i（s_0 的初始值为 0），因此在扰动阶段结束后，有 $\sum_{i=0}^{N}s_i=\sum_{i=0}^{N}n_i$。也就是说，聚集阶段 Sink 节点得到 $S=\sum_{i=0}^{N}s_i=\sum_{i=0}^{N}n_i$，协议可以得到正确的聚集结果。

通过模拟实验分析了协议的抗合谋攻击能力和通信效率。实验中设定的网络包含了 600 个传感器节点和一个 Sink 节点，所有节点随机分布在一个 400m 宽的正方形区域内，每个传感器节点的信号传输半径设定为 50m。实验结果如下。

1) 抗合谋攻击能力分析

在扰动阶段，若|$N_1(v_i)\cup R_i$|中元素的个数小于 L，那么 t=L−|$N_1(v_i)\cup R_i$| 大于 0，后面的 WHILE 循环将被调用，使得节点 v_i 再生成 t 个随机数，发给其他节点。最终，节点 v_i 的抗合谋攻击能力将达到 L。若|$N_1(v_i)\cup R_i$|中元素的个数不小于 L，那么节点 v_i 的抗合谋攻击能力等于|$N_1(v_i)\cup R_i$|，也不小于 L。因此，协议执行完成之后，每个节点的抗合谋攻击能力都不小于 L。

在上述网络配置下，记录了新协议的抗合谋攻击能力，其结果如图 2 所示。

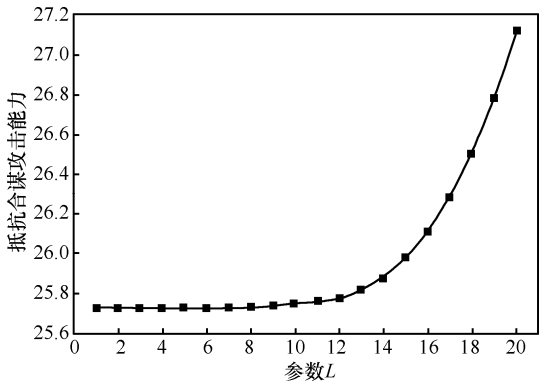


图 2 本文中协议的抗合谋攻击能力

从图 2 可以看出，本文中协议的抗合谋攻击能力随着参数 L（节点抗合谋攻击能力的下限）变大而逐渐增大。当选择参数 L=20 时，协议的抗合谋攻击能力已经超过 27。随着 L 的增加，协议的抗合谋攻击能力还将继续增大。

2) 通信效率分析

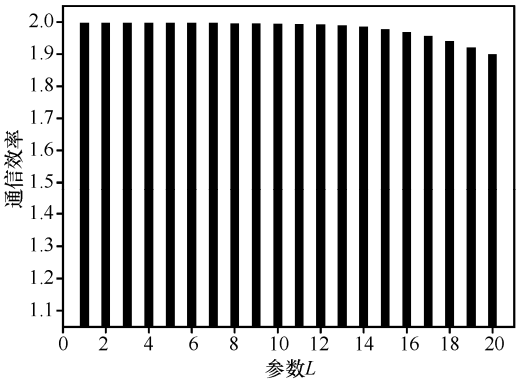
利用随机数和随机分片等数据扰动方法的保护隐私数据聚集协议需要发送一些额外的扰动数据。因此，其通信效率对协议的实用性影响至关重要。接下来，将新协议和 SMART 协议的通信效率进行对比分析。

定义 3 设 f 为一个保护隐私数据聚集协议，$C(f)$ 为其抵抗合谋攻击能力，m 为协议执行过程中所有节点发送的扰动数据（包括随机数和随机分片等）总数。记协议 f 的通信效率为

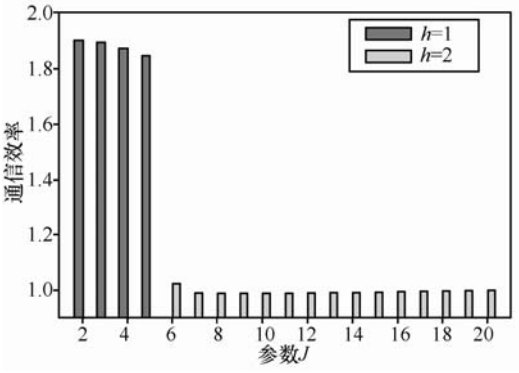
$$E(f)=\frac{NC(f)}{m}$$

其中，N 为无线传感器网络中持有私密传感数据的节点个数。

易于看出，在一定的安全性要求下，即抗合谋攻击能力 $C(f)$确定时，协议所发送的额外的扰动数据越少，协议的通信效率 $E(f)$越高。$E(f)$反映了每个扰动数据对协议的抗合谋攻击能力的贡献程度。图 3 记录了新协议和 SMART 协议的通信效率。



(a) 新协议的通信效率



(b) SMART 协议的通信效率

图 3 新协议和 SMART 协议的通信效率

实验结果显示，新协议的通信效率在 1.9 到 2.0 之间；而 SMART 协议的通信效率要低得多，特别是在其参数 J（每个数据划分的分片个数）大于 5 时，由于分片个数的增加，选择分片发送的目标节点的跳数 h 也必须相应地增加，SMART 协议的通信效率将降低到 1.0 附近，大约是协议的一半。因此，在安全性要求较高时，新协议的优越性更加明显，其所需要发送的扰动数据大约只是 SMART 协议的一半，即新协议是一直高效的抗合谋攻击的数据聚集方案。

4　结束语

如何在完成数据聚集的同时，保护私有传感数据不被泄露是一个重要的研究问题。本文定义了保护隐私数据聚集协议的抗合谋攻击能力的测量模型，并设计了一种高效的抗合谋攻击的保护隐私数据聚集协议。理论分析和模拟实验证实了新协议的抗合谋攻击能力；对比结果显示在同等安全性要求下，新协议的通信量大约只有 SMART 协议的一半。

参考文献：

[1] CULLER D, ESTRIN D, SRIVASTAVA M. Overview of Sensor Networks[J]. IEEE Computer, 2004, 37(8): 41-49.

[2] XU N, RANGWALA S, CHINTALAPUDI K, *et al*. A wireless sensor network for structural monitoring[A]. Proc of the ACM Conference on Embedded Networked Sensor Systems[C]. Baltimore, USA, 2004.

[3] MAINWARING A, POLASTRE J. Wireless sensor networks for habitat monitoring[A]. WSNA'02[C]. Atlanta, Georgia, USA, 2002.

[4] HE W, LIU X. PDA: privacy-preserving data aggregation in wireless sensor networks[A]. Proc of 26th IEEE International Conference on Computer Communications (Infocom)[C]. Anchorage, Alaska, USA, 2007.

[5] ZHANG W, WANG C, FENG T. GP2S: generic privacy-preservation solution for approximation aggregation of sensor data[A]. Proc of 6th Annual IEEE International Conference of Pervasive Computing and Communications[C]. Hong Kong, China, 2008.

[6] CHAN H, PERRIG A, SONG D. Secure hierarchical in-network aggregation in sensor networks[A]. Proc of 13th ACM Conference on Computer and Communications Security (CCS)[C]. New York, NY, USA,2006.

[7] YANG Y, WANG X. SDAP: a securre hop-by-hop data aggregation protocol for sensor networks[A]. ACM Mobihoc[C]. 2006.

[8] BARTOSZ P, SONG D, PERRIG A. SIA: secure information aggregation in sensor networks[A]. ACM SenSys[C]. 2003.

[9] MADDEN S, FRANKLIN M J, HELLERSTEIN J M. TAG: a tiny aggregation service for ad-hoc sensor networks[A]. OSDI[C]. 2002.

[10] JANA S, PREMNATH S N. On the effectiveness of secret key extraction from wireless signal strength in real environments[A]. Proc of Mobicom'09[C]. Beijing, China, 2009.

作者简介：

朱友文（1986-），男，安徽阜阳人，中国科学技术大学博士生，主要研究方向为信息安全和无线传感网络。

黄刘生（1957-），男，安徽安庆人，中国科学技术大学教授、博士生导师，主要研究方向为信息安全、高性能算法、分布式计算等。

杨威（1978-），男，安徽六安人，博士，中国科学技术大学博士后，主要研究方向为信息安全和量子信息。

田苗苗（1987-），男，安徽阜阳人，中国科学技术大学硕士生，主要研究方向为信息安全和无线传感网络。