基于复数的隐私保护数据融合方案

曹晓梅 刘大伟

(南京邮电大学计算机学院 南京 210003)

摘 要 致力于如何实现无线传感器网络的安全数据融合,提出了一种基于复数的隐私保护数据融合方案CNPD(Complex Number-based Privacy Data Aggregation)。CNPD方案中,传感器节点在采集到数据之后,利用自身的私有数据形成复数形式的自定义数据来隐藏和融合敏感数据。即使隐私数据被攻击者窃听或解密,该方案也可以防止它们恢复出敏感信息,从而有效地保护了用户隐私数据的安全。通过理论分析和仿真实验表明:与PDA相比,CNPD在通信开销以及安全方面更加高效。

关键词 无线传感器网络,数据融合,复数,数据安全

Complex Number-based Privacy Data Aggregation

CAO Xiao-mei LIU Da-wei

(School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract The scheme is committed to realize the security of data aggregation, and we proposed an secure data aggregation scheme CNPD based on in wireless sensor networks. CNPD scheme applies its own private data to form a plural custom data to hide sensitive data. Even if the private data being eavesdropped or decrypted by the attacker, this method can prevent them to resuming the sensitive information, so there is an effective protection of user privacy data security. The theoretical analysis and simulation experiments finally show that CNPD is more efficient than PDA in terms of communication overhead and security.

Keywords Wireless sensor networks, Data aggregation, Complex number, Data security

无线传感器网络(Wireless Sensor Netwok,简称WSN)集微机电技术、嵌入式技术、网络和无线通信技术于一身[1],近年来被广泛地应用于各种场合来收集信息,比如追踪关键设施,监视动物的栖息地等[2]。由于被相邻节点采集到的数据在时间和空间方面是相关的,因此WSN数据融合成为近几年来颇受关注的研究领域[3,4]。网络内的数据融合可以减少WSN中传输数据量的大小,以及数据传输和数据收集所涉及到的节点数目,从而高效利用WSN有限的节点和网络资源。

传感器节点之间数据的无线传输,使它们容易遭到拦截和窃听。WSN中的另一个问题是如何保护日常生活中敏感测量数据的安全性,尤其当WSN被应用于商业或者军事场景时,安全性对于WSN来说就显得格外的重要,采集数据、传输过程和对节点物理位置的定位等都不能让他人获知,否则就可能泄露重要的机密信息,最终造成难以估量的损失[5]。数据融合过程中保证数据的安全同等重要[6],安全数据融合的目标就是保证基站最后得到正确的可以接受的融合结果,这就要求无线传感器在采集数据、传递和融合过程中能够有效地应对各种类型的攻击,并且在节点有限的计算、通信能力和存储空间的情况下,能够满足各种安全需求。

许多WSN安全融合协议用来解决数据融合的安全性,从而使攻击者无法通过窃听或捕获节点的方式发起攻击。本文提出了一种高效的WSN安全数据融合方案CNPD。该方案是利用一个复数的代数表达式,在传感器节点向接收器进行传输时使用算术运算进行数据隐藏及数据融合,即使隐私数据被攻击者窃听或解密,该方法也可以防止它们恢复敏感信息。

1 相关工作

在WSN中有两种安全数据融合协议:基于点对点加密传输的数据融合安全方案和基于逐跳加密传输的数据融合安全方案。逐跳加密的数据融合方式主要是采用常规的对称密码算法,如:RC5、RC6。由于逐跳加密方式在均衡开销和安全性上存在很大的安全隐患,点对点加密方式正逐渐引起广泛的重视。

在点对点加密方式的数据融合操作中,只有基站能够解密得到最终的数据融合结果,攻击者即使捕获了数据融合节点,也无法得到有效的数据[7]。要想实现数据融合操作的点对点加密传输,一个直接而有效的办法就是使用秘密同态加密算法[8,9]。J. Girao, D. Westhoff, M. Schneider提出的CDA算法[8]将Domingo-Ferrer同态算法[9]引入到无线传感器的数据融合领域中来。该方案只需要进行简单的模加法运算,但是普通叶节点的传输开销加大,安全弹性较差。C. Castelluccia等人提出了一种基于流密钥的同态加密算法CMT[10],

曹晓梅 副教授,主要研究方向为网络与信息安全、无线网络安全性、入侵检测和容侵性;**刘大伟** 硕士生,主要研究方向为无线传感器网络的数据融合安全。

整个算法简单高效，可抗击选择明文攻击和重放攻击，传输开销也明显小于 CDA 算法。但是这个算法在 ID 传输方面存在很大的问题，方案的开销比较大，而且网络开销分布极不均匀。E. Mykletun, J. Girao, D. Westhoff 提出了一种基于公钥密码学的同态加密算法 ECEG[11]。该算法是基于椭圆曲线的加密算法，同样支持加法数据融合，但是计算开销以及传输开销非常大。

与本文的研究最为密切的方案是 PDA[12]，作者针对不同的网络拓扑结构提出了 CPDA 和 SMART 两种算法。该方案主要采用了数据分割和多路归进的思想，并没有更多地涉及复杂的密码算法。攻击者通过监听或者是捕获节点的方式无法获取有效的数据信息，所以该方案有很好的保密性和弹性，但是由于网络中数据传输产生不必要的流量，造成了较高的计算和通信开销。W. Li 在文献[7]中提出了一种新的数据融合方案 CACR。该方案利用中国剩余定理的数学运算来进行数据分割，在减少网络开销和网络延时方面起到了很好的效果，降低了数据融合结果被攻击者获取的风险。另外，H. Li 提出的 EEHA 算法[13]和 G. Yang 提出的 ESPART 算法[14]从减少数据通信量和提高精确度方面对 SMART 方案进行了改进。

2 CNPD 方案

为了克服 CPDA 和 SMART 方案的局限性，我们提出一个新的安全数据融合方案 CNPD。

CNPD 使用应用复数的代数表达式隐藏融合数据，以防止其它受信任的传感器节点或者攻击者参与进来。更具体地说，我们用复数的附加属性来隐藏和融合敏感数据。因为其它融合函数比如平均值、计数、方差、标准差和其它任何的测量数据都可以转化为附加融合函数 SUM，所以 CNPD 着重附加融合函数 SUM。为了得到所有传感器的数据，查询服务器可以使用高效的树构造算法，例如 TAG[15]和 SRT(语义路由树)[16]传播融合数据的查询。

如图 2 所示，我们假设一个多跳的 WSN 数据融合模型，该模型适用于数据收集，如环境监测或目标追踪。这个模型的具体属性：1. 查询服务器或者基站应是一个根的网络拓扑，并且可以位于网络的任何地方。2. 大量能量有限的传感器节点均匀地部署在网络区域中，并且根据 sink 节点的跳数被布置在不同的层次中。3. 每个传感器节点具有感知、汇总和转发数据的能力，并且可以周期性地将固定长度的数据包发送到数据接收器。4. 在不需要接收或发送数据时，传感器节点可以切换到睡眠模式或低功耗模式，以保持自己的能量。

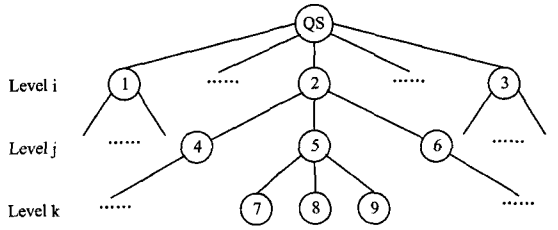

图 1 WSN 多跳数据融合

另外，所有传感器节点共享两种类型的密钥[17]，一种是传感器节点与基站之间的密钥，第二种是数据安全融合树中这些传感器节点之间的对称密钥。方案有关的一些定义如下。

定义 1 复数 X 是实部加上一个虚部得到的扩展数据，虚部表示为 i，i 满足：$i^2=-1$，即 $i=\sqrt{-1}$。

例如，一个复数 $C=30+80i$，30 和 80i 分别是实部和虚部单位。

定义 2 每个传感器节点都有一个私有的数据，传感器节点的采样数据与它们自身私有的数据结合成一个隐藏值。

例如，如果一个传感器节点的温度读数是 25，35 是这个传感器节点自身私有的数据，这个数据只有查询服务器和传感器节点知道，然后读数可以隐藏为 25+35=60。

定义 3 隐藏值加上一个虚部组成自定义数据，这个自定义的数据是复数的形式。

例如，如果 60 是隐藏值，75i 是虚部，这两个值只有查询服务器和传感器节点知道，那么我们可以得到一个自定义数据的复数形式 $C=60+75i$。

由于传感器节点在能源供应、处理器速度和内存方面的资源约束，我们需要设计一个 WSN 的数据融合算法，该算法的执行应尽可能少地消耗资源，使 WSN 的应用程序运行更长的时间。下面是 SUM 融合函数的算法。

输入：一个融合的 WSN 和 SQL 类型的 SUM 融合查询
输出：SUM 融合结果

第 1 步 根据源节点数据创建自定义数据
for n sensor nodes
Sense m // 数据采样
Mask(m,n)// 利用自身的私有数据 n 隐藏 m
a=m+n;// n 是自身的私有数据
GetCpxNum(a,bi) // 虚数 bi
$R_1=a+bi$;// 转化成复数形式
$Enc(K_{x,y},(R_1))$ // $K_{x,y}$是传感器节点 x 和 y 的对称密钥
$C_j=E_K(R_1)$;// 加密自定义数据
$Transmit(C_j)$

第 2 步 应用复数的附加属性求每一个中间融合者的自定义数据的和值(即中间结果)
for all received customized data
$Drc((K_{y,x},(C_j))$ // 解密接收到的数据
Add()
$R'=R_1+R_2$;// 融合两个源节点的自定义数据
$Enc(K_{y,z},(R'))$
$C_1=E_K(R')$;
$Transmit(C_1)$

第 3 步 通过查询服务器计算融合结果
Receive(C) // 加密的中间融合结果
for all C
$Drc((K,(C_1))$
Add()
$SUM_2=R+R''$;// R 是中间融合结果的总和，R″是查询服务器自定义的读数

第 4 步 通过查询服务器提取来自 SUM_2 的传感器节点的实际 SUM 值
$Disjoin(SUM_2)$ // 将 SUM_2 分成实部和虚部
Take real unit SUM_2
$SUM_1=$ Compute(sum of real seeds of all source nodes)
$SUM=SUM_2-SUM_1$;
return SUM

在第 1 步中，当查询服务器接收到一个带有 SUM 的类似于 SQL 的查询时，该请求将会根据请求的范围被传播到整个 WSN 或只是网络的一部分。首先每个传感器节点的采样数据 m 与自身私有的实数 n 隐藏到 a 中。然后 GetCpxNum ()将一个私有的虚部 bi 与 a 结合形成一个复数形式 $R_1=a+bi$。为此在网络部署中，主设备在与 WSN 中所有的传感器节点建立一个成对的密钥后会为它们提供独一无二的实数和虚部。接下来，源节点通过使用对称密钥 K 和传感器节点发送到父节点的密文加密自定义的数据，$C_1=a+bi$。在第 2 步中，当中间传感器节点(融合节点)收到从子节点传来的加密的数据时，融合节点首先通过使用各自的对称密钥对数据进行解密，然后通过使用复数的附加属性加上特定的数据来产生中间结果。接着中间结果进行加密后发送到上层查询服务器。在第 3 步中，查询服务器收集所有的中间结果并通过使用各自的对称密钥对它们进行解密。查询服务器通过中间结果以及自身的私有数据，得出最终的自定义数据融合结果 SUM_2。在算法最后的步骤中，查询服务器将复数形式的自定义数据 SUM_2 分离成实部和虚部。由于传感器节点的源数据隐藏在实部中，因此查询服务器通过计算 SUM_2 值减去 SUM_1(所有传感器节点的私有数据的总和)值来得到最终实际的 SUM 值。图 2 表示了在一个聚合的多跳 WSN 中，传感器的私有数据使用复数的安全集合，其中包括 9 个传感器节点 N_0-N_8 以及顶部的一个查询服务器 N_0。

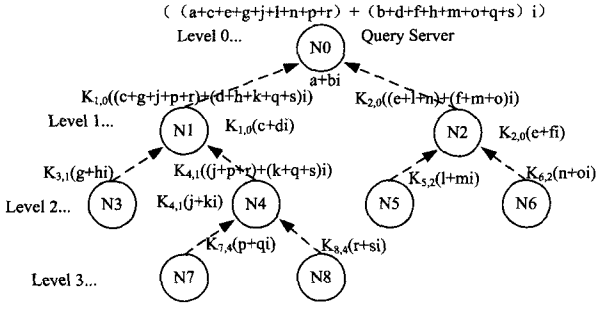

图 2　多跳 *WSN* 自定义传感器读数集合

3　仿真与分析

3.1　通信开销

3.1.1　CPDA 方案

在这个协议中，簇头发送 3 条消息(其中 2 条用于数据隐藏，1 条用于数据融合)，每个簇成员发送 2 条消息(都用于数据隐藏)。假设在 WSN 中，N 是传感器节点的总数，N_c 是涵盖 n 个传感器节点的簇，即 N_c 是簇头的数量。所以，$(N-N_c)$就是簇成员的数量。那么 CPDA 中平均交换的消息数量为：

$$AvMegCPDA=3N_c+2(N-N_c)=2N+N_c$$

所以，CPDA 的通信开销就是 $O(2N+N_c)$。

3.1.2　SMART 方案

该方案中，在空闲状态(分割系数 $J=3$，且网络密集)下，每一个节点的私有数据融合都需要 3 条消息(其中 2 条用于数据隐藏，1 条用于数据融合)。因此，SMART 中有 N 个传感器节点的 WSN 的通信开销就是 $O(3N)$。

3.1.3　CNPD 方案

在我们的方案中，由于每个节点本身可以隐藏自己的数据，并不需要交换消息来隐藏数据，因此数据融合只有一条消息。那么在有 N 个传感器节点的 WSN 中，方案的通信开销为 $O(N)$，类似于没有数据隐私的 TAG。

通过上述分析可以看出，在现有的方案中，CNPD 方案在通信开销方面更加高效。因此，在 WSN 中，我们的方案可以更好地利用资源，通过延长传感器节点运行应用程序的寿命来提供安全的隐私数据融合。

图 3 表示了在 WSN 中产生消息数量方面的通信开销。从图中可以看出随着传感器节点数目的增加，CPDA，SMART 和 CNPD 生成的消息数目也自然增加，但是 CNPD 分别与 CPDA 和 SMART 方案比较只产生了约一半和三分之一的消息数量。同时在 WSN 中每个生成的消息都需要一定的能量才能到达查询服务器，从而 CNPD 分别比 CPDA 和 SMART 方案节约了 230%和 300%的能量。

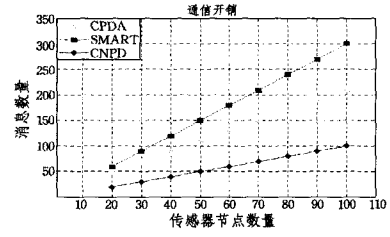

图 3　不同协议产生消息的数量

3.2　安全性

在 WSN 数据隐私性保护方面，当传感器节点被攻击者窃听或捕获时，它们所采集的私有数据有可能会泄露出去。在 CNPD 方案中，传感器节点将采集到的数据与自身的私有数据通过相关的数学运算进行隐藏，并通过查询服务器提供的虚部形成一个复数形式的自定义数据。即使传感器节点被攻击者窃听或破解，由于无法通过查询服务器获知传感器节点自身的隐私数据，其也无法恢复出传感器节点真实的私有数据。在这种情况下，CNPD 方案中任意一个节点的隐私数据被破解的概率 $P_{CNPD}(q)=0$。

图 4 显示了 3 种方案节点间的链接被破解的概率。从图中可以看出，在查询服务器不被窃听或捕获的情况下，CNPD 的安全性比 CPDA($p_c=0.3$)和 SMART($J=3$)更高。虽然在大规模的传感器网络中，随着 CPDA 方案中 p_c 取值的不断减小以及 SMART 方案中 J 取值的不断增加，传感器节点的隐私数据被破解的概率会大大降低甚至趋近于 0，但是较小的 p_c 值和较大的 J 值势必会产生更大的通信开销，这就需要在设计方案时对隐私保护和通信开销进行权衡。因此，在查询服务器不被窃听或捕获的情况下，CNPD 方案可以满足大部分隐私保护的要求。

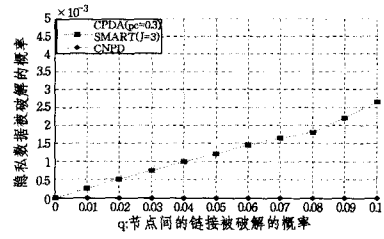

图 4　3 种方案的隐私保护对比

结束语　本文针对容易产生敏感数据的环境提出了一个

高效的安全数据融合方案 CNPD。该方案利用复数的附加属性，将发送到查询服务器之前的感测数据转换成复数形式。即使数据被窃听和解密，攻击者也难以恢复出敏感信息。通过分析可以表明，CNPD 方案在通信开销以及安全方面比现有的方案更加高效。

由于 CNPD 方案保护数据隐私是建立在现有高效的密钥管理方案基础上的，因此只有两者共同携作为用户提供基本的安全属性，如访问控制、消息完整性和消息保密性等，才能更好地保护用户的数据隐私。在今后的工作中，我们会使用 TOSSIM 仿真对方案进行更加广泛的性能评估。

参 考 文 献

[1] 孙利民，等. 无线传感器网络[M]. 北京：清华大学出版社，2005

[2] Culler D, Estrin D, Srivastava M. Overview of Sensor Networks [J]. IEEE Computer Society Press, 2004, 37(8): 41-49

[3] 康健，左宪章，唐力伟，等. 无线传感器网络数据融合技术[J]. 计算机科学，2010，37(4)：31-35

[4] 张鹏，喻建平，刘宏伟. 传感器网络安全数据融合[J]. 计算机科学，2011，34(8)：106-108

[5] Dong Xiao-mei, Li Shan-shan. A Secure Data Aggregation Approach Based on Monitoring in Wireless Sensor Networks[J]. IEEE Computer Society, 2011, 10(9): 122-129

[6] Chen Chien-ming, Lin Yue-Hsun, Lin Ya-Ching, et al. RCDA: Recoverable Concealed Data Aggregation for Data Integrity in Wireless Sensor Networks[J]. IEEE Computer Society, 2012, 23 (4): 727-734

[7] 黎为. 无线传感器网络数据融合安全方案的研究[D]. 长沙：湖南大学，2009

[8] Girao J, Westhoff D, Schneider M. CDA: Concealed Data Aggregation for Reverse Multicast Traffic in Wireless Sensor Networks[C]// Proc of IEEE International Conference on Communications. Washington: IEEE Computer Society Press, 2005: 3044-3049

[9] Ferrer J D. A Provably Secure Additive and Multiplicative Privacy Homomorphism[C]// Proc of the 5th International Conference on Information Security. London: Springer-Verlag Press, 2002: 471-483

[10] Castelluccia C, Mykletun E, Tsudik G. Efficient Aggregation of Encrypted Data in Wireless Sensor Networks[C]// Proc. of the Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services. 2005: 109-117

[11] Mykletun E, Girao J, Westhoff D. Public Key Based Cryptoschemes for Data Concealment in Wireless Sensor Networks[C]// Proc of IEEE International Conference on Communications. New York: IEEE Communications Society Press, 2006: 2288-2295

[12] He W B, Liu X, Nguyen H. Pda: Privacy-preserving Data Aggregation in Wireless Sensor Networks[C]// Proc of 26th IEEE International Conference on Computer Communications. Washington: IEEE Computer Society Press, 2007: 2045-2053

[13] Li H, Lin K, Li K. Energy-efficient and High-Accuracy Secure Data Aggregation in Wireless Sensor Networks[J]. Computer Communication, 2011, 34: 591-597

[14] 杨庚，王安琪，等. 一种低能耗的数据融合隐私保护算法[J]. 计算机学报，2011，34(5)：792-800

[15] Madden, Samuel R, Franklin, et al. TAG: a Tiny Aggregation Service for Ad Hoc Sensor Networks[J]. OSDI, 2002, 36(SI)

[16] Madden, Franklin, Hellerstein, et al. TinyDB: an Acquisitional Query Processing System for Sensor Networks[J]. ACM Transactions on Database Systems, 2004, 30(1): 122-173

[17] Zitterbart M. An Efficient Key Establishment Scheme for Secure Aggregating Sensor Networks[C]// Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security. March 2006: 303-310

(上接第 290 页)

[7] Li H, Yu D. A statistical study of neighbor node properties in Ad hoc network[C]// Proceedings of the International Conference on Parallel Processing Workshops(ICPPW'02). 2002

[8] Bettstetter C. On the minimum node degree and connectivity of a wireless multihop network[C]// Proceeding of ACM International Symposium on Mobile Ad hoc Networking and Computing (MobiHoc). June 2002

[9] Santi P, Blough D M, Vainstein F. A probabilistic analysis for the range assignment problem in ad hoc networks[C]// Proceeding of ACM International Symposium on Mobile Ad hoc Networking and Computing(MobiHoc). Oct 2001

[10] Miller L E. Probability of a two-hop connection in a random mobile network[C]// 2001 Conference on Information Sciences and Systems. March 2001: 21-23

[11] Panchapakesan P, Manjunath D. On the transmission range in dense ad-hoc radio networks[C]// Proc. SPCOMM 2001. Bangalore, India, July 2001: 1-38

[12] Penrose M D. On k-connectivity for a geometric random graph [J]. Wiley Random Structures and Algorithms, 1999, 15(2): 145-164

[13] Xue F, Kumar P R. The number of neighbors needed for connectivity of wireless networks[C]// Wireless Networks. 2004: 169-181

[14] Bollobas B. Random graphs[M]. Cambridge University Press, 2001

[15] PenroseM. Randomgeometricgraphs[M]. Oxforduniversitypress, 2003

[16] Boudec J L. Understanding the simulation of mobility models with palm calculus[EB/OL]. technical report, http://lcawww.epfl.ch/publications/leboudec, 2005

[17] 刘宴涛，安建平，卢继华，等. 无线自组织网络个体移动模型分析[J]. 通信学报，2010，31(2)：36-43

[18] Upton G J G, Fingleton B. Spatial data analysis by example[M]. Jone Wiley & Sons, 1985

[19] Boudec J L, Vojnovic M. The random trip model: stability, stationary regime, and perfect simulation[J]. IEEE/ACM trans. on networking, 2006, 14(6): 1153-1166