

能量均衡的数据融合隐私保护算法研究

鄂 旭^{a,b}, 杨明婧^a, 励建荣^b, 毛玫静^a, 谭 艳^a

(渤海大学 a. 信息科学与技术学院; b. 食品科学研究院, 辽宁 锦州 121000)

摘 要: 针对传感器节点能耗不均的问题进行研究, 为提高数据安全保护强度, 提出一种簇间节点能量保持均衡的数据融合隐私保护算法。通过分簇路由均衡各节点能量, 延长网络生命周期, 利用同态 Hash 函数保护数据完整性, 并引入簇组密钥, 以降低节点数据被破译的概率, 加强对隐私数据的保护。实验结果表明, 与现有的基于向量密钥的数据融合保护算法相比, 该算法具有较高的安全性和较长的网络生命周期。

关键词: 分簇路由; 数据融合; 隐私保护; 簇组密钥; 完整性检测

中文引用格式: 鄂 旭, 杨明婧, 励建荣, 等. 能量均衡的数据融合隐私保护算法研究[J]. 计算机工程, 2016, 42(4): 126-130.

英文引用格式: E Xu, Yang Mingjing, Li Jianrong, et al. Research on Energy-balanced Data Fusion Privacy-preserving Algorithm[J]. Computer Engineering, 2016, 42(4): 126-130.

Research on Energy-balanced Data Fusion Privacy-preserving Algorithm

E Xu^{a,b}, YANG Mingjing^a, LI Jianrong^b, MAO Meijing^a, TAN Yan^a

(a. Department of Information Science and Technology;

b. Food Science Research Institute, Bohai University, Jinzhou, Liaoning 121000, China)

[Abstract] Aiming at the energy consumption inequality of sensor nodes, and to improve data security protection intensity, this paper puts forward a data fusion privacy-preserving algorithm with balancing cluster node energy. Through clustering routing to balance each node energy, it prolongs the network life cycle, uses the homomorphism Hash function to protect data integrity, introduces cluster group key to reduce the probability of node data to be deciphered, and strengthens the protection of private data. Experimental results show that, compared with the existing protection data fusion algorithm based on vector secret key, the algorithm has higher security and longer network lifetime.

[Key words] clustering routing; data fusion; privacy-preserving; cluster group key; integrity detection

DOI: 10.3969/j.issn.1000-3428.2016.04.023

1 概述

物联网(Internet of Things, IoT)技术广泛应用于农牧业、建筑、交通、环境监测等多个方面, 作为物联网的两大关键技术之一, 无线传感器网络(Wireless Sensor Network, WSN)因其传感器节点的代价很低, 且本身具有传感、计算和通信的能力, 被放置在各式各样的环境中用来收集相应环境的信息, 并将信息进行除杂融合, 以此来进行环境特征的描述。如在水产养殖区域, 因为很难在大面积的水面上布线^[1], 所以可运用无线传感器网络实现数据融合。

多传感器融合的数据能够有效地反映水产养殖

地的水质变动, 研究人员用此数据进行水产品的监管和控制^[2]。多传感器融合中节点的能量来自于其内部的电池, 所以, 运行时它的存储、计算能力会受到限制。文献[3]表明, 使用 Micadot 节点在进行 1 bit 的数据传输时所消耗的能量是处理器执行一条指令的 800 倍, 这就需要在实际利用的时候减少通信数量, 采用合适的数据融合算法, 以便有效地去除冗余数据, 降低节点的能耗。

在实际应用时, 因为无线传感器网络自身的特点很容易遭到恶意者的攻击, 如何保护数据在节点传输途中不被窃取和更改成为亟待解决的问题。为了阻止数据被恶意捕获, 许多学者已提出了不同的隐私保

基金项目: 中国博士后科学基金资助项目(2012M520158); 辽宁省自然科学基金资助项目(2014020141); 辽宁省社会科学规划基金资助重点项目(L14AGL001); 辽宁省“百千万人才工程”基金资助项目(2012921058)。

作者简介: 鄂 旭(1971-), 男, 教授、博士, 主研方向为信息安全、数据挖掘; 杨明婧, 硕士研究生; 励建荣, 教授、博士; 毛玫静、谭 艳, 硕士研究生。

收稿日期: 2015-06-02 **修回日期:** 2015-08-11 **E-mail:** 79216084@qq.com

护方案。其中,文献[4]提出了基于簇的隐私数据融合(Cluster-based Private Data Aggregation,CPDA)方案和基于数据切片的隐私保护算法(SMART)2种数据融合算法。但是CPDA的计算量大,SMART的数据通信量大,而且对数据丢失敏感。之后又对这2种算法进行了扩展,提出复杂度比较高的可进行完整性检测的隐私数据融合算法 iPDA (integrity-protecting Private Data Aggregation)和 iCPDA (integrity-protecting Cluster-based Private Data Aggregation)算法。文献[5-8]以微型融合服务(Tiny Aggregation Service for Ad-hoc Sensor Networks,TAG)算法^[9]为基础提出了不同的隐私保护算法。文献[5]利用同态加密算法对数据进行隐私保护,并用中国剩余定理对汇聚节点进行数据完整性检测。文献[6-7]则是对SMART算法精确度或性能进行改进。文献[8]利用同态Hash函数对采集数据进行运算,生成一个Hash函数验证码,且不可逆推,以此来避免节点数据的泄露,并利用组密钥和密钥向量较精确地检测数据丢失情况,提高汇总数据的精确度。文献[10-12]利用同态加密进行隐私保护,即使应用人员在没有解密密钥的情况下也可以对密文直接进行加、减、乘、除运算,并可以减少时延。

本文基于文献[8]提出一种能量均衡的数据融合隐私保护算法(PAEDA)。算法分为4个阶段,即簇的建立、簇内融合、簇间融合和完整性验证。在簇建立时,为了平衡各个簇的内部及簇之间在进行通信时的能耗,使用文献[13]提出EBUCA算法建立簇群,用同态加密的方法进行隐私保护,并引入簇组密钥,使各簇节点在选择密钥时具有自己的选择范围,从而减少出现相同密钥的几率,降低节点数据被破译的概率,保障簇内节点或簇头在进行传输时的数据安全。

2 PAEDA 算法的系统模型

在传感器节点固定的水产养殖区域中,多传感器节点之间形成一个连通图 $G(V,E)$, V,E 分别表示网络中的节点集合与节点间的链路集合。如图1所示,节点被分成若干簇,由簇首节点传递本簇的融合信息。

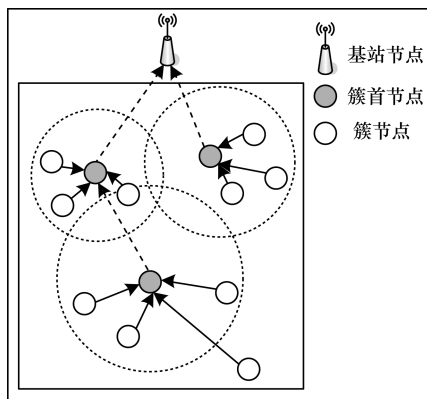


图1 PAEDA 算法建模

PAEDA 算法利用 E-HEED 算法^[14]进行簇首的选

择,该算法是对高效的分簇协议算法 HEED(A Hybrid, Energy-Efficient Distributed Clustering Approach)^[15]的改进,使簇首能够比较均匀地分布在网络区域内,并利用EBUCA算法来建立簇群,保障簇内的能量均衡,延长节点及网络的生命周期。

2.1 簇首选择

无线传感器网络中的节点除了基站节点都要参与簇首的选择,E-HEED算法也继承了LEACH算法中轮的思想,同样包含初始化阶段和稳定阶段。在初始化阶段中,簇首节点被选择出来,度量节点能否成为簇首节点的标准主要由该节点当前的剩余能量和循环次数组成,因此,节点 i 在第 t 轮能够被选择成为簇首节的概率 $P_i(t)$,其计算方法如下:

$$P_i(t) = \max \left\{ P_{\text{prob}} n/x \cdot \frac{E_{i-\text{cur}}}{E_{i-\text{init}}}, P_{\text{min}} \right\} \quad (1)$$

其中, P_{prob} 和 P_{min} 是网络系统中的统一参数; P_{min} 是 $0 \sim 1$ 之间的随机数; $E_{i-\text{cur}}$ 是节点 i 当前的剩余能量; $E_{i-\text{init}}$ 是节点 i 的初始能量,剩余能量与初始能量的百分比为节点能量等级的标准; n 为网络中初始能量相等的节点数量; x 为网络中能量等级较高的节点数量,这些节点可以参加簇首选择。

如图2所示,在每轮的准备阶段基站节点发出“hello”信号给网络中的各节点,节点在收到信号后将自己的地址信息发送回基站节点,并选择出簇首,图2中灰色节点表示簇首节点,其余为簇节点(文中其余各图中的灰色节点皆表示簇首节点)。

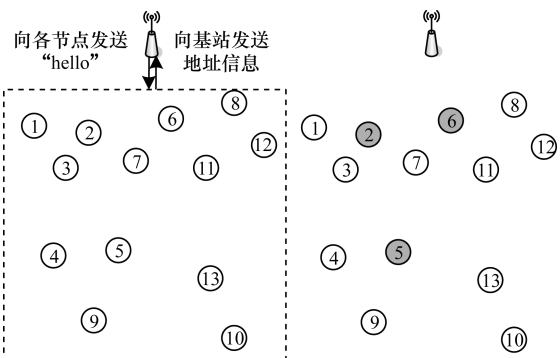


图2 簇首的选择

2.2 簇的建立

在簇首节点选择完成之后,其他的节点处于待定的状态,基站节点开始以相同的功率向基站节点发送“start”信号,各簇首节点查看接收信号的强弱程度,并确定簇首节点到基站节点的距离 d_{hi-s} ,通过下列公式计算该簇的半径 R_{hi} 。

$$R_{hi} = \left[(1 - \rho_{hi}) \frac{d_{\text{max}} - d_{hi-s}}{d_{\text{max}} - d_{\text{min}}} \right] R_{\text{max}} \quad (2)$$

其中, ρ_{hi} 为区域内簇首节点的密度,其取值范围是 $(0, 1)$; $d_{\text{max}}, d_{\text{min}}$ 分别为 d_{hi-s} 的最大值和最小值,即所有簇

首节点中距离基站节点的最远距离和最近距离。

从式(2)可以看出,簇半径由簇首的密度及簇首节点与基站节点的距离决定。簇首密度越大、距离基站节点越近的簇半径越小,距离基站节点近的簇首可以保留一部分能量用于簇首间数据转发。在保障簇首节点多跳传输的同时,减少簇首节点的能量消耗,延长节点寿命。图3所示为簇的形成过程,各节点根据运算加入合适的簇内。

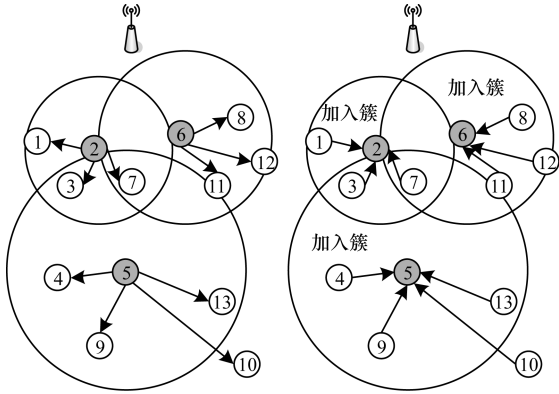


图3 簇的形成

2.3 数据的隐私保护

在此阶段要进行数据的采集、加密、传输、融合等。在算法开始时,先根据网络中节点的数量计算出密钥池的个数 $S_{num} > \lceil \lg N \rceil$, 基站节点向各簇首节点发布包括密钥池、安全素数 m 以及 g 的信息包, 其中, $m = a \times b$ (a, b 为安全素数), g 是节点采集的数据乘群中的生成元; 簇首节点从密钥池中选取 $P > \lceil \lg N / n_i \rceil$ 个密钥组成簇组密钥, 再将信息包中密钥更新以后发布给自己的簇成员节点, 簇节点则从簇组密钥中任选 1 个 ~ P 个密钥作为自己的密钥进行加密, 同时生成长度为 P 的密钥向量 A_i , 用来记录所选的密钥。引入簇组密钥可以更有效地避免节点选择相同的密钥, 从而增强数据安全性。

2.3.1 数据加密

在加密阶段, 节点 i 将采集的数据 d_i 加密得到其同态 Hash 验证码 h_i , 其计算方法如下:

$$h_i = H(d_i) = g^{d_i} \bmod m \quad (3)$$

节点将自己选取的密钥相加得到密钥和 K_{i_sum} , 再将采集的数据与密钥和相加, 形成新的自定义数据 R_i , 其计算方法如下:

$$R_i = d_i + K_{i_sum} \quad (4)$$

在节点 i 的密钥向量、同态 Hash 验证码以及自定义数据都生成之后, 将信息包整体加密, 并传递给簇首节点。

2.3.2 数据融合

在簇首节点接收完簇节点的信息包后, 获取所有的密钥向量、同态 Hash 验证码以及自定义数据,

对密钥向量及自定义数据分别进行加法融合, 得出簇组密钥向量 A_{Hi_sum} 与簇组自定义数据和 R_{Hi_sum} , 亦将所有的验证码进行乘法融合, 得出簇组验证码积 h_{Hi_pro} 。分别表示如下:

$$A_{Hi_sum} = \sum_{i=1}^z A_i = \left(\sum_{i=1}^z a_{i_1}, \sum_{i=1}^z a_{i_2}, \dots, \sum_{i=1}^z a_{i_p} \right) \quad (5)$$

$$R_{Hi_sum} = \sum_{i=1}^z R_i = \sum_{i=1}^z d_i + \sum_{i=1}^z K_{i_sum} \quad (6)$$

$$h_{Hi_pro} = \prod_{i=1}^z H(d_i) = g \sum_{i=1}^z d_i \bmod m \quad (7)$$

在完成以上计算之后, 簇首节点便把簇组密钥向量、簇组自定义数据和簇组验证码加密, 采用多跳的方式传递给基站节点。

图4为数据加密融合与传递的过程, 假设网络中共有 13 个传感器节点, 密钥库中密钥个数 $\geq \lceil \lg 13 \rceil = 4 > \lfloor \lg 13 \rfloor = 4$, 取 5 作为密钥的个数, 且 $\lfloor \lg 13/3 \rfloor = 3 \lfloor \lg 13/3 \rfloor = 3$, P 取值为 4, 即各簇首从密钥库中分别取 4 个密钥作为各自的簇组密钥, 现以簇 2 为例详细描述该过程, 假设簇首节点 2 取 S_1, S_2, S_3, S_4 组成簇组密钥。

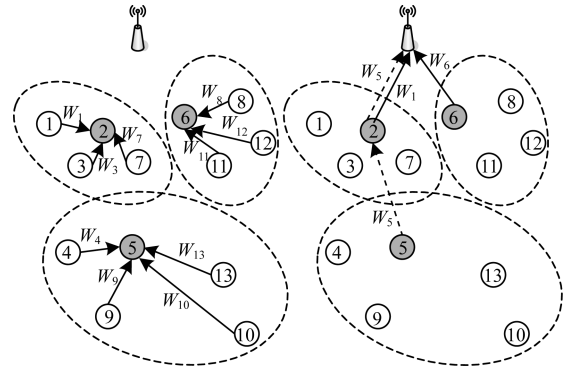


图4 数据的融合与传递

若节点 1, 2, 3, 7 采集的真实数据分别为 d_1, d_2, d_3, d_7 , 选取的密钥所对应的密钥向量分别为 A_1, A_2, A_3, A_7 , 其中, $A_1 = (1, 0, 1, 0)$, $A_2 = (0, 0, 1, 0)$, $A_3 = (0, 1, 1, 0)$, $A_7 = (1, 1, 0, 1)$, 数据 d_1, d_2, d_3, d_7 的同态 Hash 验证码及自定义数据分别为:

$$h_1 = H(d_1) = g^{d_1} \bmod m, R_1 = d_1 + S_1 + S_3$$

$$h_2 = H(d_2) = g^{d_2} \bmod m, R_2 = d_2 + S_3$$

$$h_3 = H(d_3) = g^{d_3} \bmod m, R_3 = d_3 + S_2 + S_3$$

$$h_7 = H(d_7) = g^{d_7} \bmod m, R_7 = d_7 + S_1 + S_2 + S_4$$

在簇首节点 2 接收完 W_1, W_3, W_7 后, 对其进行解密, 将解密后的信息与自身数据融合得到 h'_2, R'_2, A'_2 , 其中, $h'_2 = g^{d_1 + d_2 + d_3 + d_7} \bmod m$; $R'_2 = d_1 + d_2 + d_3 + d_7 + 2S_1 + 2S_2 + 3S_3 + S_4$ 。在融合结束后, 将 h'_2, R'_2, A'_2 加密生成 W_2 , 传递给基站节点, 图中的 W_i 为节点 i 的验证码、密钥向量、自定义数据加密后

的信息包。

2.3.3 数据完整性验证

在基站节点接收到所有簇首节点传输的信息后,先为各融合包解密,将解密后各簇首的3种信息再进一步融合,以获得网络中全部节点的密钥向量 A_{sum} 、自定义数据和 R_{sum} 、验证码积 H_{pro} ,再从自定义数据和中将各节点采集的数据和 SUM_d 分离出来,计算出 SUM_d 的同态 Hash 验证码 $H(SUM_d)$,比较 $H(SUM_d)$ 与 H_{pro} 是否相等,以此进行数据的完整性验证,判断是否有数据被恶意篡改或伪造。若两者相等,说明在数据传递、融合途中数据没有被泄露,融合的数据均可信;若两者不等则可能是有恶意者捕获了节点,将节点采集的数据进行了篡改,数据不可信,应放弃此次融合的数据。其中采集数据和的计算方法如下:

$$SUM_d = R_{sum} - A_{sum} = R_{sum} - \sum_{j=1}^{S_{num}} S_j \cdot A_{sum} \quad (8)$$

3 PAEDA 算法设计与分析

3.1 算法设计

输入 节点个数 n , 节点 ID, 初始能量 E_{i-init} 以及参数 P_{prob}, P_{min}

输出 融合数据

(1) 簇首选举

1) 计算各节点的能量比 $P = E_{i-cur}/E_{i-init}$, P 值大的成为簇首预备节点,计算预备节点的概率 $P_i(t)$, E_{i-cur} 为节点的剩余能量, E_{i-init} 为节点初始能量。

2) 节点各自选取随机数 y , 若 $y < P_i(t)$, 则该节点成为簇首节点,随机数选取范围是 $0 \sim 1$ 。

(2) 建立簇组

1) 簇首节点向网络中的节点公布自己的位置及簇半径 R_{Hi} 。

2) 节点计算自己与收到信号的簇首节点的距离 d_{i-Hi} , 如果 $\min(d_{i-Hi}) < R_{Hi}$, i 加入 Hi 簇内; 否则, 找下一个 $\min(d_{i-Hi})$, 当有 $\min(d_{i-Hi}) < R_{Hi}$ 时, 加入该簇, 否则, 找 $\min(d_{i-j})$, 加入 j 所在的簇。

(3) 数据安全性保护

1) 基站节点选取 S_{num} 个密钥作为整个网络的密钥库, 将密钥库公布给簇首节点。

2) 簇首节点 Hi 从密钥库中选取 P 个密钥生成簇密钥组, 将密钥组公布给自己的簇成员节点。

3) 簇成员节点 i 从簇组密钥中选取 1 个 $\sim P$ 个密钥形成自己的密钥, 并生成密钥向量 A_i , 向量中 0 为未选取此密钥; 1 为选取此密钥。

4) 节点 i 采集的真实数据为 d_i , 计算 $H(d_i), R_i$ 。

5) 将 h_i, R_i, A_i 打包成一个信息包, 对信息包加密生成 $W_i, W_i = Enc \langle h_i, R_i, A_i \rangle$, 将 W_i 传递给簇首

节点 Hi 。

6) Hi 等待所有的簇成员节点传递完信息后, $Dnc(W_i), Agg(h_i, R_i, A_i)$, 计算出 $A_{Hi-sum}, R_{Hi-sum}, h_{Hi-pro}, W_{Hi}$ 。

7) 如果簇首节点距离基站节点的距离 $d_{Hi-sink}$ 小, 则 $Transmit(W_{Hi})$; 否则, 找到距离自己最近的簇首节点进行多跳传递, $Transmit(W_{Hi})$ 。

(4) 验证数据完整性

1) 基站对所有的簇首节点信息进行解密, 计算出 $SUM_d, H(SUM_d), H_{pro}$ 。

2) 如果 $H(SUM_d) = H_{pro}$, 则将数据加入数据库中; 否则, 证明数据是不完整的, $Dele(SUM_d)$ 。

3.2 算法分析

从数据的安全性、节点能耗 2 个方面对算法进行分析, 采用 Matlab R2015a 进行仿真实验, 网络环境的设置为 $200 \text{ m} \times 100 \text{ m}$ 的水产养殖地, 有 400 个传感器节点随机分布在环境中, 节点的初始能量为 2 J , 节点的传输速率为 1 Mb/s , 传输距离为 40 m 。

3.2.1 数据安全性分析

在本文算法中, 当密钥池中密钥的个数大于或等于 $\lfloor \lg 13 \rfloor$ 时, 节点 i 的数据可能被暴露的概率即为恶意者找到该节点密钥的概率, 即:

$$Q(i) = 1/C_{S_{num}}^P \cdot (2^P - 1) \quad (9)$$

图 5 为 VHIPA 算法^[8]与本文算法数据安全性的对比。VHIPA 算法中数据的安全性随着所选密钥池的大小变化, 由于本文算法引入了簇组密钥, 因此簇组密钥的数量 P 的大小也会引起安全性的变化, 图中分别列举了 $P=6, P=7, P=8$ 情况下数据的暴露概率, 从图中可以看出, 当密钥池中密钥的个数达到 $S_{num} = 6$ 后, 节点数据的暴露概率明显下降, 当 $S_{num} = 10$ 时, 数据暴露概率基本趋近于 0, 此时, 恶意者要破解密钥得到其真实数据的几率渺茫, 由此可以看出本文算法数据安全性更高。

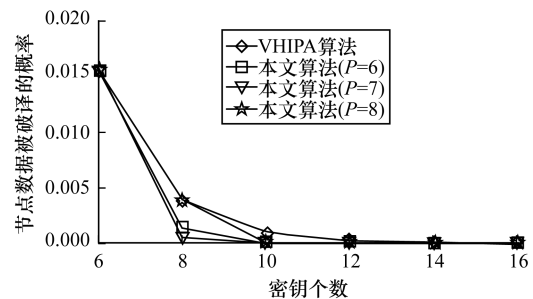


图 5 隐私保护性对比

3.2.2 节点寿命分析

PAEDA 算法采用的簇首选择及建簇方法可以保持网络节点能耗均衡, 同态加密算法也避免了将节点数据分片加密, 在网络运行时簇节点与簇首节点都只需向上级发送一个加密信息, 即网络中的节

点在整个运算中只需发送一个信息包,通信开销也减少为 $O(n)$,使得每个节点寿命有所延长,图6为 PAEDA 算法与其他 2 个算法通信量的对比,其中, VHIPA 算法在建立融合树时同 TAG 算法相同,需将节点自己的父节点设置成其接收到的 Hello 信号的源节点,并将此信号发送出去。在融合阶段,各节点向上级发送一个数据包,因此, VHIPA 算法的通信量为 $O(2n)$; SMART 算法的数据通信量与其数据的分片数量有关,图6中只显示其分片数为4时的通信量为 $O(3n)$ 。

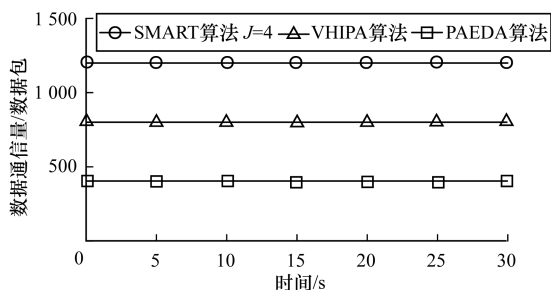


图6 3个算法的通信量对比

图7为本文算法与 LEACH 算法在进行相同运算次数后节点的剩余情况对比,从图中可以看出,本文算法中节点的寿命(网络运行轮数)明显延长,从600轮之后才开始有节点死亡,而且节点的死亡速率较缓,由此可以证明本文算法增加了网络的生命周期。

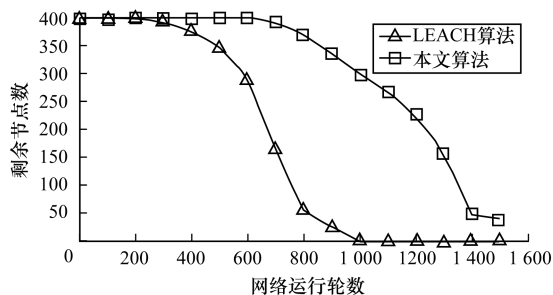


图7 网络生命周期对比

4 结束语

本文提出一种基于能量均衡的数据融合隐私保护算法。采取分簇协议降低因碰撞导致数据丢失的概率以及网络拥挤的情况,在簇首的选择上增加了可选择的机会,并通过节点密度及距离均衡节点能量的消耗,促使节点寿命的延长。在加密算法中加入了簇组密钥,为各簇节点提供足够密钥的同时,降低了各节点选择相同密钥的几率和节点被捕获后数据暴露的概率,加强了数据隐私的安全。但本文没有过多考虑比较适合当前网络的簇首数量问题,以及水产养殖环境中众多自然因素的影响,在下一步工作中将针对这些问题展开研究。

参考文献

- [1] 李道亮,傅泽田.集成化水产养殖数字化集成系统[M].北京:电子工业出版社,2010.
- [2] 刘海燕,刘云.基于数据融合的无线传感器网络退感知研究[J].计算机工程,2015,41(4):1-6.
- [3] Szweczyk R, Ferencz A. Energy Implications of Network Sensor Designs[EB/OL]. (2010-11-21). http://www.google.com/url?url=http://bwrcs.eecs.berkeley.edu/Classes/CS252/Projects/Reports/robert_szweczyk.pdf&rct=j&frm=1&q=&src=s&sa=U&ved=0ahUK EwiA67S3l43LahUDVRQKHS8pBIoQFggZMAA&sg=A FQjCNFCzYiWrXuEYMTtDWNapiJG4lhIQ.
- [4] He Wenbo, Liu Xue, Nguyen H, et al. PDA: Privacy-preserving Data Aggregation in Wireless Sensor Networks[C]//Proceedings of the 26th IEEE International Conference on Computer Communications. Washington D. C., USA: IEEE Press, 2007: 2045-2053.
- [5] 周强,杨庚. WANS 中可保护数据完整性和隐私的数据融合算法[J]. 计算机应用研究, 2013, 30(7): 2100-2104.
- [6] 杨庚,王安琪,陈正宇,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.
- [7] 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[J]. 计算机学报, 2013, 36(1): 189-200.
- [8] 张丹. 适用于 WSN 的数据融合完整性保护算法研究[D]. 南京: 南京邮电大学, 2013.
- [9] 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. New York, USA: ACM Press, 2002: 131-146.
- [10] Castelluccia 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: Networking and Services. Washington D. C., USA: IEEE Press, 2009: 109-117.
- [11] Albath J, Madria S. Secure Hierarchical Data Aggregation in Wireless Sensor Networks[C]//Proceedings of WCNC'09. Washington D. C., USA: IEEE Press, 2009: 1-6.
- [12] Agrawal S, Boneh D. Homomorphic MACs: MAC-based Integrity for Network Coding[J]. Applied Cryptography and Network Security, 2009, 5536: 292-305.
- [13] 卢先领,王莹莹,王洪斌,等.无线传感器网络能量均衡的非均匀分簇算法[J]. 计算机科学, 2013, 40(5): 78-81.
- [14] 王刚,张红伟,李晓军.无线传感器网络中簇首选择算法研究[J]. 通信技术, 2010, 43(8): 35-40.
- [15] Younis O, Fahmy S. HeeD: A Hybrid Energy-efficient Distributed Clustering Approach for Ad-hoc Sensor Networks[J]. IEEE Transactions on Mobile Computing, 2004, 3(4): 660-669.