软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2014,25(10):2385−2396 [doi: 10.13328/j.cnki.jos.004495]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

传感器网络中一种基于单向哈希链的过滤方案*

刘志雄[1], 王江涛[1], 王伟平[2], 刘华富[1], 王建新[2], 张士庚[2]

[1](长沙学院 计算机科学与技术系,湖南 长沙 410022)

[2](中南大学 信息科学与工程学院,湖南 长沙 410083)

通讯作者: 王江涛, E-mail: wjt77@163.com

摘 要: 已有传感器网络中,过滤机制只能在转发过程中过滤虚假数据而无法过滤重复数据,且无法防范协同攻击.提出了一种基于单向哈希链的过滤方案HFS.在HFS中,节点在部署后将密钥和初始哈希值预分发给部分中间节点存储,每个数据包附带 t 个 MAC 和新鲜哈希值,转发节点同时对数据包中检测节点之间相对位置关系的合法性、MAC 和哈希值的正确性以及哈希值的新鲜性进行验证.理论分析及仿真实验结果表明,HFS 可同时过滤传感器网络中的虚假数据和重复数据,并能有效对抗协同攻击.

关键词: 无线传感器网络;虚假数据;重复数据;单向哈希链;协同攻击

中图法分类号: TP393

中文引用格式: 刘志雄,王江涛,王伟平,刘华富,王建新,张士庚.传感器网络中一种基于单向哈希链的过滤方案.软件学报, 2014,25(10):2385−2396. http://www.jos.org.cn/1000-9825/4495.htm

英文引用格式: Liu ZX, Wang JT, Wang WP, Liu HF, Wang JX, Zhang SG. One-Way hash chain based filtering scheme in wireless sensor networks. Ruan Jian Xue Bao/Journal of Software, 2014,25(10):2385−2396 (in Chinese). http://www.jos.org.cn/1000-9825/4495. htm

One-Way Hash Chain Based Filtering Scheme in Wireless Sensor Networks

LIU Zhi-Xiong[1], WANG Jiang-Tao[1], WANG Wei-Ping[2], LIU Hua-Fu[1], WANG Jian-Xin[2], ZHANG Shi-Geng[2]

[1](Department of Computer Science and Technology, Changsha University, Changsha 410022, China)

[2](School of Information Science and Engineering, Central South University, Changsha 410083, China)

Corresponding author: WANG Jiang-Tao, E-mail: wjt77y@163.com

Abstract: Existing filtering schemes in wireless sensor networks can only filter out false reports but not the replayed reports during forwarding. Furthermore, they can not resist cooperative attacks. In this article, a one-way hash chain based filtering scheme (HFS) is presented. In HFS, each node distributes its key and initial hash value to some other nodes after deployment. When a report is generated for an observed event, it carries the MACs and fresh hash values from t detecting nodes. Each forwarding node validates the legitimacy of the relative position of the detecting nodes carried in the report, the correctness of the MACs and hash values, and the freshness of these hash values. Analysis and simulation results show that HFS can not only filter out false reports and replayed reports simultaneously, but also resist collaborative attacks efficiently.

Key words: wireless sensor networks; false reports; replayed reports; one-way hash chain; collaborative attacks

随着通信技术、嵌入式计算技术和传感器技术的飞速发展和日益成熟,无线传感器网络(wireless sensor networks,简称 WSNs)在军事和民用领域获得了越来越广泛的应用[1].传感器网络通常部署在野外或者是敌方区域,攻击者可以通过俘获节点并利用存储在节点内的秘密信息捏造事实上不存在的虚假事件,发动虚假数据

* 基金项目: 国家自然科学基金(60873265, 61379117); 教育部新世纪优秀人才计划(NCET-10-0798); 湖南省教育厅科学研究重点项目(13A114)

收稿时间: 2012-04-28; 定稿时间: 2013-09-02

注入攻击[2],或者将缓存的合法数据重复注入到网络中[3].这些非法数据将会引发错误警报、干扰用户决策并消耗宝贵的网络资源[4].

鉴于虚假数据和重复数据的威胁,不少学者提出了一些解决办法[3−13].它们的共同特点是:在数据包后附带 t 个消息验证码(message authentication code,简称 MAC)或时间戳等额外信息,并在数据转发过程中对 MAC 实施认证,这里,t 是系统参数.这些方案可以有效过滤虚假数据,但若攻击者利用妥协节点将缓存的合法数据重复发送到网络中,则中间节点将无法进行过滤,导致重复包最终都传输到 sink,造成网络能量的浪费.此外,它们无法检测和过滤由不同地理区域的妥协节点协同伪造的虚假数据.如图 1 所示,假设 t=5,R_0 为针对突发事件 e 所产生的合法数据报告,S_1,…,S_5 为妥协节点.若攻击者利用妥协节点 S_5 将缓存的数据包 R_0 重复发送到网络中,或者利用不同地理区域的妥协节点 S_1,…,S_5 协同伪造假包 R_1 并发送到网络中,则转发节点和 sink 都无法进行过滤.

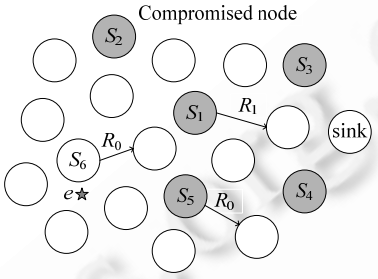

Fig.1 Replayed and false reports injection attack in WSN

图 1 攻击者利用妥协节点注入重复数据和虚假数据

本文提出一种基于单向哈希链的过滤方案 HFS,将节点邻居信息(包括邻居节点 ID、密钥及初始哈希值)预分发给部分中间节点存储,并在数据报告中同时附带 t 个检测节点的 ID、所产生的 MAC 以及新鲜的哈希值,然后由中间节点在转发过程中对数据包中包含的检测节点之间相对位置关系、MAC 以及哈希值进行验证,从而同时将虚假数据和重复数据过滤掉,并有效抵抗协同攻击.

1 相关工作

Ye 等人率先对传感器网络中虚假数据过滤的问题进行了讨论,提出了 SEF 机制[3].SEF 将一个全局密钥池分成 $n(n>t)$个密钥分区,每个分区包含 m 个密钥.每个节点在部署前随机地从全局密钥池中选取一个密钥分区,并从中任选 k 个密钥进行存储.检测到突发事件后,多个检测节点联合产生一个包含 t 个互不相同的 MAC 的数据报告.在数据包转发过程中,与检测节点拥有相同密钥分区的中间节点能够以概率 k/m 对数据包中的一个 MAC 进行验证.所有漏过转发过滤的假包将最终由 sink 进行过滤.SEF 存在如下两个问题:首先,若攻击者利用妥协节点将缓存的合法数据重复发送到网络中,则其邻居节点都无法进行检测,从而迅速耗尽节点有限的能量;其次,若多个妥协节点利用所获取的密钥协同伪造虚假数据,转发节点也无法对其进行过滤.

Zhu 等人提出了在路由时进行交叉、逐步认证的机制 IHA[4].IHA 将节点组织成簇,每个簇头(cluster head,简称 CH)建立一条到 sink 的路径;接下来,路径中相距 t+1 跳的节点通过建立对偶密钥的方式成为协作节点.检测到突发事件后,每个检测节点利用与 sink 共享的私钥以及与下游协作节点共享的对偶密钥产生两个 MAC.簇头收集 t+1 个检测节点的 MAC 信息生成数据报告.在转发过程中,每个节点对数据包中由其上游协作节点所产生的 MAC 进行验证,一旦验证成功,再利用与下游协作节点共享的对偶密钥重新计算一个 MAC,并替换上游协作节点所产生的 MAC.如此交叉、逐步验证,以过滤假数据.IHA 采用一种确定性的方式进行密钥分发和数据验证,限制了网络中不同区域的妥协节点协同地伪造假数据包.然而,一旦路由发生变化,这种确定性的数据验证方式也随之失效,重新建立路由并分发密钥需要较大的维护开销.

Yu 等人提出了一种基于成组的转发过滤方案 GRSEF[5].GRSEF 将节点分成 t 组,每组中有多个节点以较大

概率同时感知到突发事件;接下来,采用多坐标系对网络区域进行划分;然后,基于区域进行认证密钥分发.检测到突发事件后,来自不同组的节点联合产生一个数据报告.与 SEF,IHA 相比,GRSEF 能够通过减少数据包中冗余信息而有效提高过滤假包的概率,但多轴划分和密钥分发消耗了较多的网络能量,不利于节省有限的传感器网络资源.

Yang 等人认为对称密钥技术的安全性较低,提出了一种基于密码交换(commutative cipher)的路由过滤机制 CCEF[6].该方案假设各节点与基站共享一个会话密钥(session key),路由中,报文转发节点不知道报文源节点的会话密钥,但可以通过交换密码对数据包进行认证,从而提高了安全性.Wang 等人利用椭圆曲线密钥技术来进一步提高 CCEF 的安全性,提出了 PDF 机制[7].Ren 等人将网络划分为多个 cells,通过门限共享技术和公钥机制来保证数据的安全性,提出了 LEDS 机制[8].

Perrig 等人提出了检测重复数据的 SPINS 机制[9],其基本思想是:在发送方和目标方各维护一个计数器,且计数器每次更新,发送方利用哈希函数[10]对计数器加密后附在数据包中一起发送,目标方解出计数器并实施认证. Zigbee 协议(即 IEEE 802.15.4)也采用计数器作为时变参数来检测重复数据[11].Yu 等人提出基于随机数检测重复数据的方案 SRAR[12]:发送方利用哈希函数对随机数加密后附在数据包中发送,目标方解出随机数并实施认证.Chen 等人提出基于时间戳的检测方案 TSPC[13],该方案要求所有节点保持时间同步,发送方在数据包中嵌入时间戳,目标节点对时间戳实施认证.

SEF,IHA,GRSEF 等方案只能过滤虚假数据而无法过滤重复数据,且无法防范协同攻击;CCEF 和 LEDS 等方案采用效率低且开销大的公开密钥技术,无法顺利应用在性能有限的传感器网络中;SPINS 和 Zigbee 等方案只能由 sink 检测重复数据而无法在转发过程中过滤;TSPC 方案要求所有节点保持时间同步,也无法顺利应用在能量有限的传感器网络中.本文研究如何在转发过程中同时过滤传感器网络中的重复数据和虚假数据,并有效对抗协同攻击.

2 基于单向哈希链的过滤方案 HFS

2.1 系统模型及相关假设

假设 sink 节点拥有全局密钥池信息、单向函数以及所有节点用来产生单向哈希链的随机数,且能量充足,具备强大的计算、存储和自我保护能力,无法被妥协,并能够过滤所有最终到达的重复数据和虚假数据.

假设传感器节点部署密度足够大,每个突发事件 e 都可被多于 t 个节点同时检测到.各个检测节点共同选举感知信号最强(即距离 e 最近)的节点作为中心节点(central of stimulas,简称 CoS).与 SEF[3]一样,本文假设所有检测到突发事件 e 的节点都能直接与 CoS 通信.CoS 收集其他检测节点产生的 MAC,并从中任选 t 个以生成数据报告.

假设网络在部署后一段较短的时间内是安全的,没有节点被俘获.节点交换邻居信息、分发邻居信息均在这一时间段完成.网络部署后,攻击者可以俘获网络中的多个节点,利用存储在这些节点中的秘密信息伪造虚假数据,并发送到网络中或者重复注入合法数据包.此外,攻击者还可以篡改、污染合法数据包等[3,19],但本文仅针对攻击者注入虚假数据和重复数据的情形提出一种解决方案.

2.2 节点部署与初始化

部署前,给每个节点分配唯一的 ID 标识.存在一个全局密钥池 $G=\{K_i:0\leqslant i\leqslant W-1\}$,每个节点 S_i 从中随机选择一个互不相同的密钥进行存储.此外,还给每个节点 S_i 预置一个随机数 x_i 和单向函数 F,该函数具备“单向、不可逆”特性,即:给定输入参数 a,很容易计算 $F(a)=b$;但反过来,无法由 b 推出 a[14].

接下来,各节点 S_i 按如下步骤生成一条长度为 u 的单向哈希链:

- 先计算 $F(x_i)=y_1,F^2(x_i)=F(y_1)=y_2,\ldots,F^u(x_i)=F(y_{u-1})=y_u$;
- 接下来,令 $h_i^u=y_1,h_i^{u-1}=y_2,\ldots,h_i^1=y_u$,即得到单向哈希链 $H_i=h_i^1,h_i^2,\ldots,h_i^u$,其中,$h_i^{q_i}$ $(1\leqslant q_i\leqslant u)$表示节点 v_i 的哈希链中第 q_i 个哈希值.

根据函数 F 的性质:由 $h_i^{q_i}$ 可以计算得到索引值小于 q_i 的哈希值,如公式(1)所示,但无法计算索引值大于 q_i 的哈希值.

$$h_i^j = F^{q_i-j}(h_i^{q_i}), j = 1,2,...,q_i - 1 \quad (1)$$

部署后,各节点 S_i 生成包含节点 ID、密钥 K_i 以及哈希链中索引值最小的哈希值 h_i^1 的短消息$\{S_i,K_i, h_i^1\}$,并广播该消息.接收到邻居节点发送的短消息后,节点 S_i 生成一个 hello 包:

$$\{S_i,S_{a_1},...,S_{a_j};K_i,K_{a_1},...,K_{a_j};h_i^1,h_{a_1}^1,...,h_{a_j}^1\},$$

其中,S_{a_j} 为 S_i 的邻居节点的 ID,K_{a_j} 为节点 S_{a_j} 所存储的密钥,$h_{a_j}^1$ 为节点 S_{a_j} 的索引值最小的哈希值.接下来,S_i 建立一条到 sink 的路径 $Path(S_i)$={S_i,S_1,…,S_d,sink},并将 hello 消息沿 $Path(S_i)$进行传输.此外,S_i 将使用过的哈希值 h_i^1 从哈希链中删除.

接收到 hello 包后,中间节点 $S_k(1\leqslant k\leqslant d)$存储节点 S_i 的邻居信息: $N(S_i)=\{S_{a_1},...,S_{a_j}\}$,并以概率 c_k/c_0 成为 S_i 的验证节点,其中,c_k 和 c_0 分别表示节点 S_k 和 S_i 距离 sink 的跳数.若成功当选为 S_i 的验证节点,S_k 从 hello 包中任选一个节点 S_{a_x} $(1\leqslant x\leqslant j)$,并以格式 $\{S_{a_x},K_{a_x},h_{a_x}^1\}$ 存储 S_{a_x} 的信息,然后将 S_{a_x} 的密钥和哈希值从 hello 包中删除(保留节点号 S_{a_x}),接下来转发 hello 包;反之,若没有当选为 S_i 的验证节点,S_k 直接转发 hello 包.hello 包传输到 sink 后不再转发.图 2 给出了密钥和哈希值分发的过程.

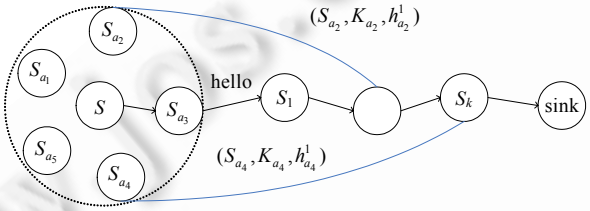

Fig.2 Keys and hash value pre-distribution

图 2 密钥和哈希值预分发

2.3 数据报告生成

事件发生后,各个检测节点共同选举出感知信号最强的节点作为中心节点,中心节点将感知数值 e(包括事件内容、事件位置以及发生时间)发送给各检测节点.检测节点 S_i 收到数据 e 后,将自己的感知数值与 e 进行比较,若误差在规定的阈值范围内,则利用密钥 K_i 对 e 进行加密,生成消息认证码 MAC:M_i=$K_i(e)$.接下来,各检测节点将节点 ID、MAC、哈希链中索引值最小的哈希值以及哈希值索引发送给 CoS,然后从哈希链中删除刚使用过的哈希值.CoS 收集其他 t–1 个检测节点的信息,并将 CoS 以及其他 t–1 个节点的 ID、MAC、哈希值及索引附带在感知数值 e 后面,生成数据报告,且必须将 CoS 的信息置于其他检测节点的信息之前.例如,假设 S_1,S_2,…,S_t 共同检测到突发事件 e,其中,S_1 为 CoS,则生成的数据报告为 R:{e;S_1,…,S_t;q_1,…,q_t; $h_1^{q_1}$,...,$h_t^{q_t}$;M_1,…,M_t}.其中,S_j 和 $q_j(1\leqslant j\leqslant t)$分别表示节点号和哈希值索引.接下来,CoS 将数据报告 R 转发给下一跳.与 SEF 一样,HFS 采用布鲁姆过滤器(Bloom filters)[15]将数据包中附带的 MAC 和哈希值分别映射成长度为 L_s 的两个字符串,以减小数据包长度以及节点的存储开销.

2.4 转发过滤

由于预先存储了上游源节点的邻居信息以及部分上游节点的密钥和初始哈希值,中间节点能以一定概率对数据包中各检测节点之间相对位置关系、MAC 以及哈希值进行验证.

当接收到转发数据包 R 时,中间节点 S_i 首先对数据包中附带的节点 ID、哈希值及索引、MAC 的数量进行检查,然后根据预存储的邻居信息检查各个检测节点之间相对位置关系的合法性,接下来验证 MAC 的正确性,最后验证哈希值的新鲜性和正确性.中间节点对数据包 R 的具体验证步骤如下:

(1) 检查数据包 R 中包含的节点 ID、哈希值、哈希值索引以及 MAC 是否各为 t 个.若任何一项的数量不符合要求,则丢弃 R;

(2) 检索存储的邻居信息表,若没有存储中心节点 S_1 的邻居信息,则丢弃 R;

(3) 检查 R 中各个检测节点 $S_2,\ldots,S_t$ 是否都是 S_1 的邻居节点.若其中任意一个节点不符合要求,则丢弃 R;

(4) 若存储了 R 中某个检测节点 $S_v(1\leqslant v\leqslant t)$的密钥 K_v,则利用 K_v 对 E 重新计算一个 M 并与 R 中附带的 M_v 比较:若二者相等,则说明 M_v 正确;否则,丢弃 R;

(5) 若存储了 R 中某个检测节点 $S_v(1\leqslant v\leqslant t)$的哈希值 $h_v^{q_d}$,则先将 q_d 与 R 中附带的 S_v 的哈希值索引 q_v 进行比较:

- 若 $q_v\leqslant q_d$,则说明 R 中的哈希值 $h_v^{q_v}$ 不新鲜,故丢弃 R;
- 反之,则接下来判断 $h_v^{q_d}=F^{q_v q_d}(h_v^{q_v})$ 是否成立:若成立,则说明哈希值 $h_v^{q_v}$ 正确,并将所存储的哈希值 $h_v^{q_d}$ 更新为 $h_v^{q_v}$;否则,丢弃 R;

(6) 若以上验证都通过,则将 R 转发给下一跳节点.

图 3 为转发过滤算法伪代码.

/***on receiving report** R*/

1. Check that t $\{S_v, M_v, h_v^{q_v}\}$ tuples exist in R; drop R otherwise.
2. Check from the pre-stored neighbor information table, if it has not stored the neighbor information for S_1, drop R otherwise.
3. Check the t−1 node IDs $\{S_v, 2\leqslant v\leqslant t\}$ all belong to the neighbor set of S_1: $N(S_1)$; drop R otherwise.
4. If it has one key $K\in\{K_v, 1\leqslant v\leqslant t\}$, it computes $M=K(e)$ and see if the corresponding M_v is the same as M; drop R otherwise.
5. If it has stored one hash value $h_v^{q_d}$, drop R if $q_v\leqslant q_d$ is true. Otherwise it then checks if $h_v^{q_d}=F^{q_v q_d}(h_v^{q_v})$ is true, if true, then it updates the stored hash value to $h_v^{q_d}$, drop R otherwise.
6. Send R to the next hop.

Fig.3 Psuedo-Code in en-route filtering

图 3 转发过滤算法伪代码

3 性能分析与仿真结果

由于已有工作大都基于 SEF 框架,且都无法在转发过程中对重复数据进行过滤以及防范协同攻击,本文仅选取经典的 SEF 方案与 HFS 进行性能比较分析.

3.1 防范协同攻击的能力

在已有方案中,例如 SEF,中间节点仅验证数据包中附带的 MAC 的正确性,故,不同地理区域的妥协节点(例如图 4 所示中的 $S_1,\ldots,S_5$ 等)可被攻击者用来协同伪造虚假数据.

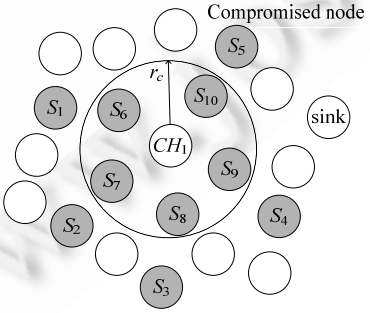

Fig.4 Collaborative attacks in SEF and HFS

图 4 SEF 和 HFS 中协同攻击比较

而 HFS 将源节点邻居信息预分发给下游节点存储,并由转发节点对数据包中各检测节点之间相对位置关系的合法性进行验证(即,判断各检测节点是否都是中心节点 CoS 的邻居),从而可以有效地对抗协同攻击.例如,假设攻击者俘获了分布于不同地理区域的节点 $S_1,\ldots,S_5$,然后以 S_1 为 CoS 伪造假包 R:{E;S_1,S_2,…,S_5;M_1,M_2,…,M_5}并发送到网络中.当接收到 R 后,若存储了 S_1 的邻居信息,则转发节点可以判断出 S_2 不是 S_1 的邻居,从而将 R 丢弃;反之,若没有存储 S_1 的邻居信息,说明 R 不是正确的源节点所产生的包,故也将 R 丢弃.

3.2 妥协容忍能力

根据过滤规则,攻击者在俘获至少 t 个不同的密钥分区之后,便可彻底攻破 SEF 安全机制.例如,当 t=5 时,攻击者在俘获了图 4 中拥有不同密钥分区的节点 $S_1,\ldots,S_5$ 之后,即可伪造出 SEF 无法过滤的假包.而为了攻破 HFS,攻击者必须俘获至少 t 个节点,且其中存在节点 S,使得其他妥协节点都是 S 的邻居.例如,当 t=5 时,攻击者在俘获了图 4 中节点 S_6 及其邻居 $S_7,\ldots,S_{10}$ 之后,便可伪造以 S_6 为 CoS 的假包,并利用 $S_6,\ldots,S_{10}$ 构造 5 个正确的 MAC 以及合法的节点 ID,使得转发节点和 sink 都无法对假包进行过滤.

定理 1. 在 SEF 中,假设存在 n 个密钥分区,网络中每个节点随机选择一个密钥分区进行存储.假设攻击者从网络中随机俘获了 N_c 个节点($N_c \geqslant t$),则其中至少存在 t 个节点的密钥分区不同的概率为

$$p_{\mathrm{SEF}}=\frac{\sum_{i=t}^{n}\left[C(n,i)\cdot\sum_{j=0}^{i}((-1)^j\cdot C(i,i-j)\cdot(i-j)^{N_c})\right]}{n^{N_c}} \tag{2}$$

证明:首先计算随机俘获 N_c 个节点后($N_c \geqslant t$),攻击者获得刚好 t 个密钥分区的概率.记 N_c 个妥协节点所构成的集合为 Q_1;从 n 个密钥分区中任取 t 个的方法数量为 $C(n,t)$,并记选取的 t 个密钥分区所构成的集合为 Q_2,则 Q_1 中的每个元素各从 Q_2 中任取一个元素进行映射,使得 Q_2 中每个元素都至少被映射一次的方法数量为

$$\begin{aligned}\varphi&=C(t,t)\cdot t^{N_c}-C(t,t-1)\cdot(t-1)^{N_c}+C(t,t-2)\cdot(t-2)^{N_c}-\ldots+(-1)^{t-2}\cdot C(t,2)\cdot 2^{N_c}+(-1)^{t-1}\cdot C(t,1)\cdot 1^{N_c}\\&=\sum_{j=0}^{t-1}((-1)^j\cdot C(t,t-j)\cdot(t-j)^{N_c})\\&=\sum_{j=0}^{t}((-1)^j\cdot C(t,t-j)\cdot(t-j)^{N_c})\end{aligned} \tag{3}$$

因此,从网络中随机俘获 N_c 个节点后,攻击者刚好获得 t 个密钥分区的方法数量为 $C(n,t)\times\varphi$.同理可计算出攻击者刚好获得 t+1 个、t+2 个、…、n 个密钥分区的方法数量.于是,攻击者获得至少 t 个密钥分区的方法总数为

$$\sum_{i=t}^{n}\left[C(n,i)\cdot\sum_{j=0}^{i}((-1)^j\cdot C(i,i-j)\cdot(i-j)^{N_c})\right] \tag{4}$$

最后,Q_1 中的每个元素各从 Q_2 中任取一个元素进行映射的方法数量为 n^{N_c}.因此,当攻击者从网络中随机俘获了 N_c 个节点时,其中至少存在 t 个节点的密钥分区不同的概率为 p_{SEF}.得证. □

定理 2. 在 HFS 中,假设攻击者随机俘获了网络中的 N_c($N_c \geqslant t$)个节点,则其中至少存在这样的 t 个节点:“其中存在节点 S,使得其他 t–1 个节点都是 S 的邻居(即与 S 的距离不大于节点通信半径 r_c)”的概率为

$$p\left(\frac{r_c}{2}\right)<p_{\mathrm{HFS}}<p(r_c) \tag{5}$$

其中,$p(\mu)$为

$$\sum_{i=t}^{N_c}\left[C(N_c,i)\cdot\left(\frac{\pi\mu^2}{Z}\right)^i\cdot\left(1-\frac{\pi\mu^2}{Z}\right)^{N_c-i}\right] \tag{6}$$

证明:记网络区域面积为 Z,半径为 μ 的区域为 M,则每个节点以概率 $\pi\mu^2/Z$ 分布在 M 中.故,M 中刚好有 t 个妥协节点的概率为 $p_b=C(N_c,t)\times(\pi\mu^2/Z)^t\times(1-\pi\mu^2/Z)^{N_c-t}$.同理可求出 M 中刚好有 t+1 个、t+2 个、…、N_c 个妥协节点的概率.因此,攻击者获得至少 t 个同处于某个半径为 μ 的区域中的妥协节点的概率为 $p(\mu)$.

另外,记事件"$y(t\leqslant y\leqslant N_c)$个节点中,存在节点 S,使得其他 $y-1$ 个节点与 S 的距离都不大于节点通信半径 r_c"为 a;记事件"y 个节点同处于某个半径为 $r_c/2$ 的区域中"为 a_0;记事件"y 个节点同处于某个半径为 r_c 的区域中"为 a_1.显然有 $a_0\rightarrow a, a\rightarrow a_1$.因此,$p(r_c/2)<p_{HFS}<p(r_c)$.得证. □

图 5 给出了当 $t=5, r_c=2.5$m, $n=15$,网络半径为 25m 时,p_{SEF} 和 p_{HFS} 的理论分析曲线和仿真实验结果.其中,p_{HFS} 的理论值介于 $p(r_c/2)$和 $p(r_c)$之间,仿真结果是在相同参数条件下随机测试 10 000 次的平均值.如图 5 所示,攻击者在俘获少量节点后,便能以较高概率攻破 SEF,但需俘获较多节点才能攻破 HFS.例如,当 $N_c=20$ 时,攻击者攻破 SEF 和 HFS 的概率分别为 99%,0.01%.因此,由理论分析和实验结果分析可知,HFS 的妥协容忍能力远强于 SEF.

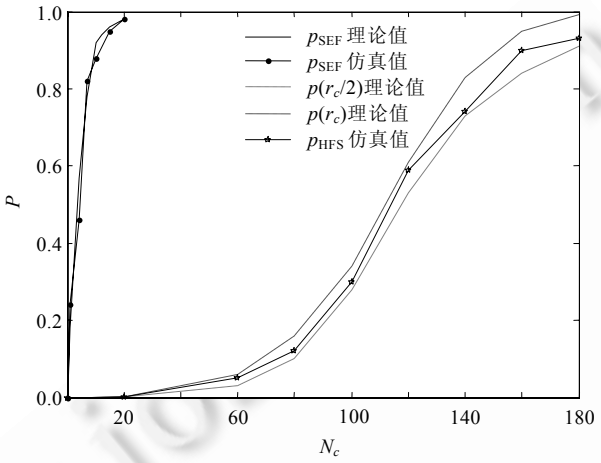

Fig.5 Theoretical and simulation results of p_{SEF}, p_{HFS}, $p(r_c/2)<p_{HFS}$ (theoretical value)$<p(r_c)$

图 5 p_{SEF},p_{HFS} 的理论值与仿真结果,$p(r_c/2)<p_{HFS}$(理论值)$<p(r_c)$

3.3 过滤概率

首先分析 HFS 过滤重复数据的能力.由于每个转发节点以一定概率存储上游节点的验证哈希值,故,可以通过验证数据包中哈希值的新鲜性而过滤重复数据.重复数据在网络中传输的跳数越大,被过滤的概率也越大.

假设攻击者利用妥协节点 S 将缓存的合法数据包 R_a 重复发送到网络中.令 S 到 sink 的转发路径为 $Path(S)=\{S,S_1,\ldots,S_d,\text{sink}\}$,$S$ 的邻居节点数量为 $num(S)$,S 到 sink 的跳数为 c_0,S_i 到 sink 的跳数为 $c_i(1\leqslant i\leqslant d)$,显然有 $c_i=c_0-i$.由于中间节点 S_i 以概率 c_i/c_0 存储源 S 的一个邻居节点的验证哈希值,故,碰巧拥有 R_a 中包含的 t 个节点中的一个的哈希值的概率为

$$p_{a\_i}=\frac{t}{num(S)+1}\cdot\frac{c_0-i}{c_0} \tag{7}$$

因为 R_a 中包含的 t 个哈希值都是不新鲜的,故,每个转发节点 S_i 能以概率 p_{a_i} 对 R_a 进行过滤.因此,重复包 R_a 在 H 跳内被过滤的概率为

$$p_a(H)=1-\prod_{j=1}^{H}(1-p_{a\_j})=1-\prod_{j=1}^{H}\left(1-\frac{t}{num(S)+1}\cdot\frac{c_0-j}{c_0}\right) \tag{8}$$

接下来分析 HFS 过滤虚假数据的能力.由于每个转发节点预存储了各个上游节点的邻居信息,并验证数据包中各检测节点是否都是中心节点的邻居,故,攻击者只能利用某个妥协节点 S 及其邻居来伪造假包.假设攻击者俘获了节点 S 以及 S 的 N_d-1 个邻居.当 $N_d\geqslant t$ 时,攻击者可以伪造出中间节点无法过滤的假包;当 $N_d<t$ 时,攻击者为了伪造出以 S 为 CoS 且"看起来合法"的假包 R_b,必须捏造 S 的 $t-N_d$ 个邻居节点的 MAC 和哈希值.由于中间节点 S_i 以概率 c_i/c_0 存储源 S 的一个邻居节点的密钥和哈希值,故,碰巧拥有这 $t-N_d$ 个伪造的节点中的一个的密钥和哈希值的概率为

$$p_{b\_i}=\frac{t-N_d}{num(S)+1}\cdot\frac{c_0-i}{c_0} \tag{9}$$

因此,假包 R_b 在 H 跳内被过滤的概率为

$$p_b(H)=1-\prod_{j=1}^{H}\left(1-\frac{t-N_d}{num(S)+1}\cdot\frac{c_0-j}{c_0}\right) \tag{10}$$

图 6 所示为重复包和假包过滤概率随传输跳数 H 变化的曲线,其中,N_d=3,c_0=20,$num(S)$=8,t=5.从图 6 中可以看出,HFS 能以较高概率同时对重复包和假包进行过滤.例如,前 5 跳过滤重复包和假包的比例分别达到 96.4%和 65%;随着传输跳数的增加,过滤比例越来越大,其中,所有重复包能在前 8 跳被过滤掉,约 95%的假包在前 20 跳能被过滤掉.

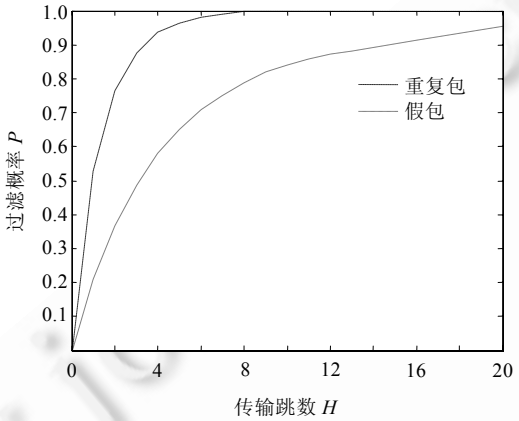

Fig.6 Filtering probability of replayed and false reports

图 6 重复包和假包过滤概率

3.4 能量开销

HFS 消耗的能量来自 4 个方面:(1) 在初始化阶段,节点获取邻居信息、建立到 sink 的路径以及将邻居信息分发给中间节点的开销;(2) 生成数据包时,中心节点 CoS 与其他检测节点之间的通信开销;(3) 中间节点进行 MAC 验证和哈希值验证的计算开销;(4) 中间节点转发数据包的通信开销.在阶段(1)和阶段(2)两个阶段中,节点之间交互的数据包较小,且持续时间较短;此外,阶段(3)中,MAC 及哈希值的计算开销比传输数据包的能耗低了几个数量级[4,16−18].因此,我们仅考虑转发数据包的能耗.

令 L_r,L_n,L_i,L_k 以及 L_s 分别表示不采用安全机制时的纯数据包长度、节点号长度、密钥索引长度、哈希值索引长度以及布鲁姆过滤器长度,则 HFS 和 SEF 中数据包长度可分别表示为 $L_{r_0}=L_r+2L_s+(L_n+L_k)t$, $L_{r_1}=L_r+L_s+tL_i$.例如,当 L_r=24bytes,L_n=10bits,L_i=10bits,L_k=10bits,L_s=64bits 时,L_{r_0} 和 L_{r_1} 分别为 52.5bytes 和 38.25bytes.显然,与 SEF 相比,HFS 在数据包中增加的额外负荷将导致传输能量的增大,但考虑到 HFS 具备防范协同攻击及过滤重复数据的能力,这些额外开销是可以承受的.此外,若攻击者利用不同地理区域的妥协节点协同注入假包,或者注入重复包,则 HFS 可以通过尽快将假包和重复包过滤掉而比 SEF 要节省能量,我们将在仿真实验部分对此进行验证.公式(11)给出“1 虚假数据+β重复数据”在 HFS 中传输 H 跳所消耗的能量.

$$E_0=L_{r_0}\cdot\left[\left(1+\sum_{i=2}^{H}\left(i\cdot p_{b\_i}\cdot\prod_{j=1}^{i-1}(1-p_{b\_j})\right)\right)+\beta\cdot\sum_{i=2}^{H}\left(i\cdot p_{a\_i}\cdot\prod_{j=1}^{i-1}(1-p_{a\_j})\right)\right] \tag{11}$$

3.5 存储开销

在 HFS 中,每个节点需存储一个预分配的密钥、一条长度为 u 的单向哈希链、所有上游源节点的邻居信息以及部分上游节点的密钥和哈希值.假设网络区域为 50×50m^2,其中随机部署了 400 个通信半径为 2.5m 的节

点,则每个节点的平均邻居节点数量为 8,所在平均路径数量为 40.当密钥、节点 ID、哈希值及哈希值索引的长度分别为 64bits,10bits,64bits 和 10bits 时,存储一个预分配的密钥、一条长度为 u 的单向哈希链、40 个上游源节点的邻居信息以及 100 个上游节点的密钥和哈希值约需 2.3KB.当前,主流节点(如 UCB 研制的 MICA2 节点)配置 3KB 以上的 SRAM 和 128KB 以上的 ROM[2],显然能够满足需求.此外,哈希链长度 u 的取值必须足够大,为了减少存储开销,节点可以仅存储哈希链中部分哈希值,例如第 k 个、第 $2k$ 个、…,并由这些存储的值来计算其他哈希值.

3.6 仿真实验

为了进一步验证 HFS 的性能,本文利用 C++语言建立了模拟仿真平台.仿真环境如下:传感器节点随机分布在一个方形网络区域中,一个静态源节点和一个静态 sink 节点分别位于区域两侧.节点发送和接收一个 SEF 数据包的功耗分别为 60mW 和 12mW[3],发送和接收一个 HFS 数据包的功耗分别为 81mW 和 16mW,发送一个数据包耗时 10ms.其中,源节点产生假包和重复包各 100 个.其他仿真参数的设置见表 1.实验考察的性能指标主要为妥协容忍能力、过滤概率以及能量消耗,考察的参数主要为妥协节点数量 N_c 和传输跳数 H.取 10 次仿真实验的平均值作为实验结果.

Table 1 Simulation parameters

表 1 仿真参数

Parameters	Value
Network area	50×50m^2
Number of nodes	400
Interval of generating a report	2s
Transmission radius of sensor nodes	2.5m
Sensing radius of sensor nodes	5m
Number of MACs each report carries (t)	5
Size of global key pool in SEF (N)	150
Number of keys partition in SEF (n)	15
Number of keys each node stores (k)	5

为了有效评价相关机制的过滤能力,本文提出利用单跳可过滤假包或重复包的累积值 f 进行性能评价,如公式(12)所示:

$$f=\sum_{H=1}^{\infty}(Q_H/H) \tag{12}$$

其中,Q_H 为第 H 跳过滤的假包或重复包个数.若某机制的过滤性能较好,则假包或重复包在网络中传输较少跳数即可被过滤,从而 f(0≤f≤100)较大;若 f=0,则说明该机制不具备过滤假包或重复包的能力.

图 7 所示为 f 随妥协节点数量 N_c 的变化情况.从图 7 中可以看出:

(1) HFS 能容忍的妥协节点数量为 110 个,远多于 SEF 能容忍的 12 个.因为 SEF 无法防范协同攻击,故攻击者在俘获少量节点后即可攻破 SEF;而 HFS 能通过相对位置关系验证防范协同攻击,故攻击者需俘获大量节点才能攻破 HFS;

(2) SEF 的假包过滤能力随 N_c 的增大而迅速降低,当 N_c 为 12 时 f 便降为 0;而 HFS 的假包过滤能力随 N_c 的增大缓慢降低,仅当 N_c 为 110 时,f 才减为 0;

(3) HFS 具备较好的重复包过滤能力(f=76.5),而 SEF 无法过滤重复包.因为 HFS 在数据包中附带哈希值,故能通过哈希值的新鲜性验证将重复包过滤掉;而 SEF 没有在数据包中附带任何“时变参数”,故无法检测和过滤重复包.

图 8 给出了当 N_c=10 时,过滤概率 P 随传输跳数 H 的变化情况.从图 8 中可以看出:

(1) HFS 能同时以较高概率对假包和重复包进行过滤,例如,前 5 跳过滤假包和重复包的比例分别达到 80%和 97%;随着 H 的增大,其过滤假包和重复包的概率都迅速增大,可分别在前 10 跳、前 6 跳将所有假包、重复包过滤掉;

(2) SEF 在前 5 跳仅能过滤掉约 25%的假包,且其假包过滤概率随 H 的增大而缓慢增大.例如,在前 20 跳仅能过滤掉约 72%的假包.此外,SEF 无法过滤重复包.

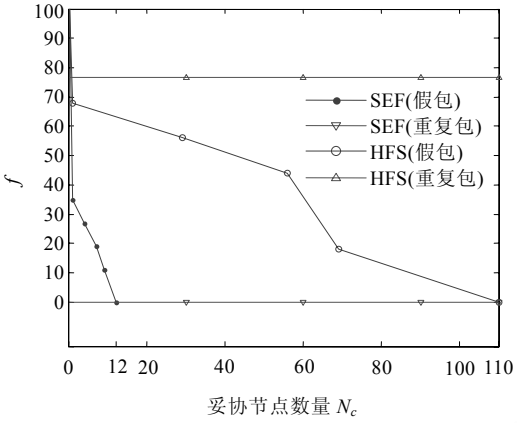

Fig.7 Filtering probability changes according to the number of compromised nodes N_c

图 7 过滤概率随妥协节点数量 N_c 变化

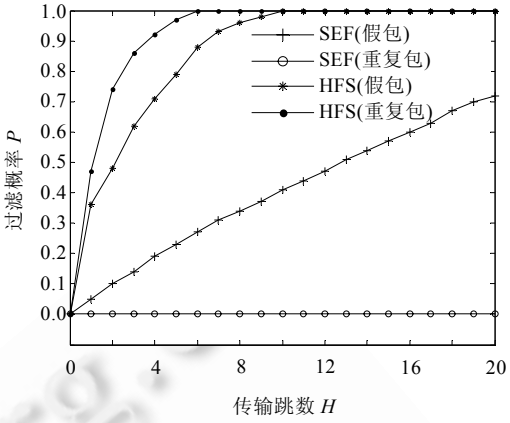

Fig.8 Filtering probability P changes according to the transmitted hops H

图 8 过滤概率 P 随传输跳数 H 变化

图 9 所示为能耗 E 随妥协节点数量 N_c 的变化情况.从图 9 中可以看出:

(1) 仅当 N_c=0 或 N_c≥110 时,HFS 传输假包的能耗大于 SEF.因为此时二者的假包过滤概率相同,而 HFS 数据包长于 SEF,故 HFS 能耗更大.在其他情况下,HFS 可通过尽早过滤假包而比 SEF 更节省能量;

(2) 随着 N_c 的增加,HFS 传输假包的能耗增幅明显小于 SEF.例如:当 N_c 由 0 增到 100 时,HFS 能耗仅由 0.97Joules 增到 19.4Joules;而仅当 N_c 由 0 增到 12 时,SEF 能耗便由 0.72Joules 增到 14.4Joules;

(3) HFS 传输重复包的能耗远低于 SEF,分别为 1.06Joules 和 14.4Joules.

图 10 给出了当 N_c=10 时,能耗 E 随传输跳数 H 的变化情况.

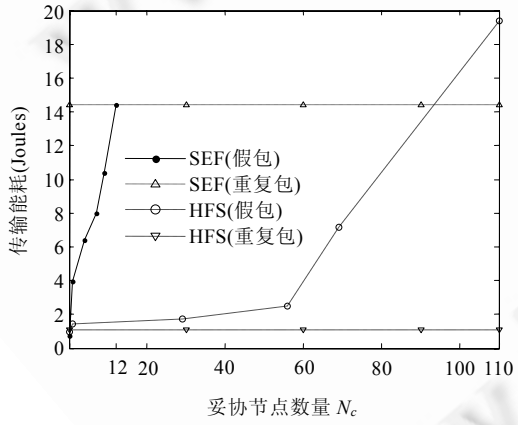

Fig.9 Energy consumption E changes according to the number of compromised nodes N_c

图 9 能耗 E 随妥协节点数量 N_c 变化

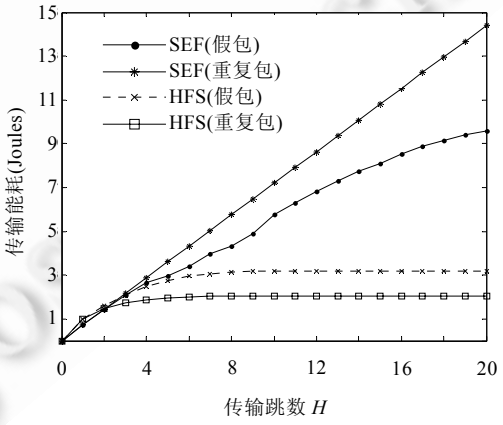

Fig.10 Energy consumption E changes according to the transmitted hops H

图 10 能耗 E 随传输跳数 H 变化

从图 10 中可以看出:

(1) SEF 传输 100 个假包的能耗随 H 的增加而逐渐增大,例如,传输 6 跳和 20 跳的能耗分别为 3.4Joules 和 9.6Joules;而等量假包在 HFS 中传输的能耗较少,例如,传输 6 跳时,E 仅为 1.5Joules,且 6 跳后 E 不

再增大;

(2) SEF 传输 100 个重复包的能耗随 H 的增加而逐渐增大,例如,传输 20 跳时 E 为 14.4Joules;而等量重复包在 HFS 中传输 6 跳以上时 E 仅为 2.02Joules.

4 结束语

本文提出一种基于单向哈希链的过滤方案 HFS,将感知节点产生的 MAC 以及新鲜的哈希值附带在数据包中发送,由转发节点同时对数据包中各检测节点之间相对位置关系的合法性、MAC 的正确性以及哈希值的正确性和新鲜性进行验证,从而可同时过滤传感器网络中的虚假数据和重复数据,并能有效对抗协同攻击.然而中间节点需存储较多信息,故,减少节点存储开销将成为进一步的工作.

References:

[1] Jian Q, Gong ZH, Zhu PD, Gui CM. Overview of MAC protocols in wireless sensor networks. Ruan Jian Xue Bao/Journal of Software, 2008,19(2):389−403 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/19/389.htm [doi: 10.3724/SP.J.1001.2008.00389]

[2] Su Z, Lin C, Feng FJ, Ren FY. Key management schemes and protocols for wireless sensor networks. Ruan Jian Xue Bao/Journal of Software, 2007,18(5):1218−1231 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/18/1218.htm [doi: 10.1360/jos181218]

[3] Ye F, Luo HY, Lu SW, Zhang LX. Statistical en-route filtering of injected false data in sensor networks. In: Proc. of the 23th IEEE Conf. on Computer Communications (INFOCOM 2004). Hong Kong: IEEE Press, 2004. 2446−2457. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1354666 [doi: 10.1109/INFCOM.2004.1354666]

[4] Zhu SC, Setia SJ, Jajodia S, Ning P. An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks. In: Proc. of the IEEE Symp. on Security and Privacy (S&P 2004). Berkeley: IEEE Press, 2004. 259−271. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1301328 [doi: 10.1109/SECPRI.2004.1301328]

[5] Yu L, Li JZ. Grouping-Based resilient statistical en-route filtering for sensor networks. In: Proc. of the 28th IEEE Conf. on Computer Communications (INFOCOM 2009). Rio de Janeiro: IEEE Press, 2009. 1782−1790. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp= &arnumber=5062098 [doi: 10.1109/INFCOM.2009.5062098]

[6] Yang H, Lu SW. Commutative cipher based en-route filtering in wireless sensor networks. In: Proc. of the 60th IEEE Vehicular Technology Conf. (VTC 2004). Los Angeles: IEEE Press, 2004. 1223−1227. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1400217 [doi: 10.1109/VETECF.2004.1400217]

[7] Wang HD, Li Q. PDF: A public-key based false data filtering scheme in sensor networks. In: Proc. of Int'l Conf. on Wlreless Algorithms, Systems and Applications (WASA 2007). Chicago: IEEE Press, 2007. 129−138. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4288224 [doi: 10.1109/WASA.2007.27]

[8] Ren K, Lou WJ, Zhang YC. Providing location-aware end-to-end data security in wireless sensor networks. In: Proc. of the 25th IEEE Conf. on Computer Communications (INFOCOM 2006). Barcelona: IEEE Press, 2006. 585−598. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4358997 [doi: 10.1109/TMC.2007.70753]

[9] Perrig A, Szewczyk R, Wen V, Culler D, Tygar JD. Spins: Security protocols for sensor networks. In: Proc. of the 7th Annual Int'l Conf. on Computing and Networking (MobiCom 2001). Rome: ACM Press, 2001. 521−534. http://dl.acm.org/citation.cfm?id=582464 [doi: 10.1023/A:1016598314198]

[10] Rivest R. The MD5 Message-Digest Algorithm. RFC, 1992. http://tools.ietf.org/html/rfc1321

[11] LAN/MAN Standards Committee. IEEE std. 802.15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (LR-WPAN). New York: IEEE Press, 2003. 111−120. http://ieeexplore.ieee.org/xpl/tocresult.jsp?reload=true&isnumber=35824

[12] Zhang WS, Cao GH. Group rekeying for filtering false data in sensor networks. In: Proc. of the 24th IEEE Conf. on Computer Communications (INFOCOM 2005). Miami: IEEE Press, 2005. 503−514. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1497918 [doi: 10.1109/INFCOM.2005.1497918]

[13] Chen SJ, Dunkels A, Osterlind F, Voigt T, Johansson M. Time synchronization for predictable and secure data collection in wireless sensor networks. In: Proc. of the 6th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net 2007). Corfu: IEEE Press, 2007. 165−172. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-79751

[14] Lamport L. Password authentication with insecure communication. Communications of the ACM, 1981,24(11):770−772. http://dl.acm.org/citation.cfm?id=358797 [doi: 10.1145/358790.358797]

[15] Bloom BH. Space/Time trade-offs in hash coding with allowable errors. Communications of the ACM, 1970,13(7):422−426. http://dl.acm.org/citation.cfm?id=362692 [doi: 10.1145/362686.362692]

[16] Ma M. Resilience of sink filtering scheme in wireless sensor networks. Computer Communications, 2006,30(1):55−65. http://dl.acm.org/citation.cfm?id=1222519 [doi: 10.1016/j.comcom.2006.07.015]

[17] Gopu JR, Shekar TP, Sagar D. An active en-route filtering scheme for information reporting in wireless sensor networks. Int'l Journal of Advanced Research in Computer Science and Software Engineering, 2012,2(4):349−356. http://www.ijarcsse.com/docs/papers/April2012/Volume_2_issue_4/V2I400130.pdf

[18] Peng SL, Li SS, Liao XK, Peng YX, Xiao N. Estimation of a population size in large-scale wireless sensor networks. Journal of Computer Science and Technology, 2009,24(5):987−996. http://dl.acm.org/citation.cfm?id=1737727 [doi: 10.1007/s11390-009-9273-9]

附中文参考文献:

[1] 蹇强,龚正虎,朱培栋,桂春梅,无线传感器网络 MAC 协议研究进展.软件学报,2008,19(2):389−403. http://www.jos.org.cn/1000-9825/19/389.htm [doi: 10.3724/SP.J.1001.2008.00389]

[2] 苏忠,林闯,封富君,任丰原.无线传感器网络密钥管理的方案和协议.软件学报,2007,18(5):1218−1231. http://www.jos.org.cn/1000-9825/18/1218.htm [doi: 10.1360/jos181218]

刘志雄(1982－),男,湖南娄底人,博士,讲师,主要研究领域为无线传感器网络.
E-mail: lzxterry@163.com

刘华富(1964－),男,教授,CCF 会员,主要研究领域为无线传感器网络.
E-mail: hfliu9063@163.com

王江涛(1977－),男,博士,副教授,主要研究领域为网络通信,可信计算.
E-mail: wjt77@163.com

王建新(1969－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为网络优化理论.
E-mail: wjianxin@mail.csu.edu.cn

王伟平(1969－),女,博士,教授,博士生导师,主要研究领域为 Web 安全技术.
E-mail: weipin@mail.csu.edu.cn

张士庚(1981－),男,博士,讲师,CCF 会员,主要研究领域为无线传感器网络.
E-mail: shigen@mail.csu.edu.cn