doi:10.11959/j.issn.1000-436x.2015143

ONSA：传感网中基于优化非均匀统计特性的源匿名协议

牛晓光[1,2]，魏川博[1,2]，冯为江[1,3]，彭国军[1,2]，张焕国[1,2]

（1. 武汉大学 计算机学院，湖北 武汉 430072；
2. 武汉大学 空天信息安全与可信计算教育部重点实验室，湖北 武汉 430072；
3. 国防科学技术大学 计算机学院，湖南 长沙 410073）

摘 要：针对无线传感网中分级信源位置隐私保护协议的特点，对不同尺寸簇及虚假信息注入统计特性在隐私性、延时和网络负载等方面进行了分析，在此基础上提出了基于优化非均匀统计特性的源匿名协议(ONSA)，以优化非均匀分簇策略及相应符合非均匀统计特性的分级优化虚假信息注入方式来高效地保护信源节点位置隐私。仿真实验结果表明，与现有方法相比，ONSA 能有效减少并均衡网络能量消耗，同时满足源位置隐私保护任务的实时性要求。

关键词：无线传感网；源位置隐私保护；虚假数据分组注入；非均匀统计特征；全局攻击

中图分类号：TP393 **文献标识码**：A

ONSA: optimal non-uniformly statistic-source anonymity protocol in WSN

NIU Xiao-guang[1,2], WEI Chuan-bo[1,2], FENG Wei-jiang[1,3], PENG Guo-jun[1,2], ZHANG Huan-guo[1,2]

(1. School of Computer Science, Wuhan University, Wuhan 430072, China;
2. Key Laboratory of Aerospace Information Security and Trusted Computing MOE, Wuhan University, Wuhan 430072, China;
3. School of Computer Science, National University of Defense Technology, Changsha 410073, China)

Abstract: According to the characteristics of hierarchical event source-location privacy (SLP) in wireless sensor networks (WSN). Firstly, the effect of the cluster size and the statistical property of fake packet injection on the privacy, latency and network traffic were analyzed. Then, an optimal non-uniformly statistic-source anonymity protocol (ONSA)was proposed to achieve source anonymity and reduce the network traffic. ONSA is designed to achieve a trade-off between network traffic and real event report latency through adjusting the transmission rate and the radius of non-uniform clusters. The simulation results demonstrate that ONSA significantly reduces the network traffic while providing source anonymity and meeting the system requirement of the delay.

Key words: wireless sensor networks; source location privacy protection; fake packet injection; non-uniformly statistic; global attacker

1 引言

无线传感网由于其无线通信方式及自身资源受限等原因而容易遭受各种安全威胁。C. Ozturk 等首次提出了无线传感网中信源节点的位置隐私(SLP, source location privacy)暴露问题[1]：在信源节点将监测到的事件信息以逐跳转发方式向基站传输的过程中，敌手很容易利用无线通信方式的广播

收稿日期：2014-07-03；**修回日期**：2014-12-03

基金项目：国家重点基础研究发展计划(“973”计划)基金资助项目(2011CB707106)；国家高技术研究发展计划(“863”计划)基金资助项目 (2013AA122301)；国家自然科学基金资助项目(41127901-06, 61373169)；长江学者和创新团队发展基金资助项目(IRT1278)；湖北省自然科学基金资助项目(2014CFB191)

Foundation Items: The National Key Basic Research Program of China (973 Program) (2011CB707106); The National High Technology Research and Development Program of China (863 Program) (2013AA122301); The National Natural Science Foundation of China (41127901-06, 61373169); The Program for Changjiang Scholars and Innovative Research Team in University (IRT1278); The Natural Science Foundation of Hubei Province (2014CFB191)

特性定位出正在发送数据分组的转发节点。因此，敌手无需破译经过加密处理数据分组的内容，仅通过被动侦听数据分组传输信号来定位转发节点的方式，就可以轻易地逐跳回溯追踪到产生数据分组的信源节点。此外，敌手还能利用传感器节点资源受限、无人照看等特点，通过发动全局或局部、主动或被动等多种形式的共谋攻击，以获取更多的信源位置信息[2]。信源位置隐私暴露严重威胁到系统监测目标的安全性，这已经成为了困扰无线传感网得以广泛应用的主要障碍。

已有的研究工作根据作用环节及保护手段的不同主要分为 2 类[3]：数据分组转发环节的基于传输扰动(transmission perturbation)的源位置隐私保护协议和数据分组产生环节的基于信源泛化(source generalization) 的源位置隐私保护协议。

基于传输扰动的源位置隐私保护协议的基本思想是通过不确定路由技术来模糊传输路径和报文到达时间分别在空间位置和时间顺序上与信源节点的关系[4]：通过路由策略使同一信源节点到达基站的传输路径动态随机变化，相应地，每个数据分组的转发路径传输时间都不相同，这样大大增加了敌手通过局部无线信道侦听逐跳回溯的方式发现信源节点位置的难度，从而实现保护信源节点位置隐私的目的。这类协议具有传输延迟小、传输可靠性高的优点，但是无法抵御由多个敌手联合发起的具备全网侦听能力的流量分析攻击。

为了抵御具备全网侦听流量分析能力敌手发动的源位置隐私攻击，研究者提出了针对数据分组产生阶段的基于信源泛化的源位置隐私保护协议：网络中所有节点均注入虚假数据分组以伪造数据源，通过调整真实数据分组的发送时机和虚假数据分组的注入时机，使所有节点的数据发送特性在统计意义上一致。即使具备全网侦听能力的敌手也无法识别数据分组是否真实，相应地，无法判断真实事件是否发生，从而在根本上保障了信源节点位置隐私。这类协议能有效抵御所有类别的源位置隐私攻击，具备隐私程度高的优点，但是为了保证节点具有相同的数据发送统计特性而推迟真实数据分组的发送时间以及额外注入的虚假数据分组会引起网络负载/传输延迟增大、网络负载/能耗分布失衡的“漏斗效应”[5]（越靠近基站，节点的网络负载/能耗越大；反之，越远离基站，节点的网络负载/能耗越小）进一步加剧、网络生存周期大幅缩短等问题。针对这个问题，文献[6]提出了改进的分级信源泛化思想：在网络中选取部分代理节点，网络所有普通节点及代理节点产生的数据分组均符合相同的统计特性，代理节点负责汇集其邻近区域内普通节点产生的数据分组以及来自上游代理节点的数据分组，对虚假数据分组进行过滤和消减，这样可以大幅度减少网络流量，很大程度上延长了网络生存周期。然而这些方案以对真实数据分组引入过多额外的事件报告延迟为代价：为保证节点数据分组的发送时间间隔服从特定的概率分布，真实数据分组在源节点和代理节点都需要延迟发送。不难看出，在改进的分级信源泛化方案中存在着与网络负载/分布呈相反趋势的“漏斗效应”：源节点距离基站越远，其到基站路径上的代理节点越多，相应地，代理节点对真实数据分组引入的总报告延迟越大；反之，总报告延迟越小。

为了缓解基于信源泛化的源位置隐私保护协议中存在的事件报告延迟和网络生存周期之间的矛盾，本文根据事件报告延迟和网络负载在网络中的非均衡分布特征，首次提出了基于优化非均匀统计特性的源匿名协议(ONSA, optimal non-uniformly statistic-source anonymity)。ONSA 采用了优化分级信源泛化方式来保护源位置隐私：首先，根据不同区域内节点的平均网络负载和能耗特征将网络划分为大小不同的簇，每个簇的簇首节点负责对来自簇内普通节点和上游簇首节点的数据分组进行泛化处理以过滤大量不必要的虚假数据分组；接着，根据应用系统的实时性要求以及节点产生的真实事件数据分组在传输过程中经过泛化处理节点（簇首节点）带来的平均延迟，对不同区域内的节点数据发送统计特性（平均数据分组发送速率）进行优化，从而实现在满足系统报告延迟要求的前提下最大程度地延长网络生存周期的目的。仿真实验和理论分析表明：ONSA 一方面能有效抵御具备全网侦听能力的源位置隐私攻击，另一方面还能显著减少并均衡网络能量消耗，延长网络生存周期，同时满足信源位置隐私保护的实时性要求。

本文的主要贡献如下。

1) 根据传感网在资源消耗、服务质量等方面的非均衡分布特性，首次提出了基于优化非均衡统计特性的分级信源泛化协议 ONSA 来实现真实事件源匿名。ONSA 力图以远离基站区域内节点的剩余能量和靠近基站区域内节点的剩余报告延迟为代

价，通过优化不同区域内的节点数据发送速率和簇尺寸来高效地保护信源节点位置隐私。

2) 将 ONSA 协议中传输延迟和网络负载性能确保的数据速率非均匀统计特征分布优化问题归纳为线性规划理论模型，该模型能够求解网络负载总量保持不变的前提下数据分组传输延迟最大最小值。证明了该模型为等价于约束满足问题(CSP)的 NP 完全问题，并提出了得到该问题近似最优解的启发式搜索算法。

2 相关工作

2.1 基于数据传输扰动的源位置隐私保护协议

C Ozturk 等[1]提出了基于随机漫步思想的“幽灵路由”协议：数据分组首先被随机转发 h 跳后到达伪信源节点，然后，伪信源节点将数据分组通过洪泛路由或最短路径路由方式到达基站。P Kamat 等[2]研究纯粹随机漫步策略对幽灵路由机制的影响，提出了改进的基于扇区的定向漫步和基于跳步数的定向漫步技术。陈娟等[7]提出的基于源节点有限洪泛方案和可视区概念，有效地保证伪源节点既远离真实源节点同时具有更高的地理位置多样性。基于随机漫步思想的信源位置隐私保护研究工作还有文献[8, 9]等，这类方法大都存在传输延迟大、维护开销较高、传输可靠性较低的问题。针对以上问题，文献[10,11]提出了基于混淆环思想的源位置隐私保护方法：选取部分节点构成混淆环，信源节点首先把监测数据分组发送到环上任意节点，然后该节点会把接收到数据分组在环上转发以便与来自其他信源节点的数据分组和环内已有虚假数据分组进行混淆，混淆达到一定程度后发送到基站。Rios 等[12]利用传感节点对其邻近区域内移动敌手位置的感知能力，动态选择能有效避开敌手侦听的到达基站的近似最短路径，从而实现信源位置隐私保护的目的。

总体而言，基于数据传输扰动的信源位置隐私保护技术主要抵御非入侵恶意敌手在局部侦听逐跳回溯模型下的位置隐私攻击。但是由多个敌手联合发起的具备全网侦听能力的流量分析攻击则能够突破此类方法防护，获取信源节点位置。

2.2 基于信源泛化的源位置隐私保护协议

针对基于数据传输扰动的源位置隐私保护机制无法抵御全网侦听流量分析攻击的问题，K Mehta 等[13]提出了基于周期性采集的源位置隐私保护方案：无论是否监测到事件发生，网络中节点均以固定速率发送数据分组。这种方式能提供最大程度上的源位置隐私保护，但引入的过多虚假数据分组以及恒定数据发送速率会导致延迟和能耗急剧增大。为获得隐私和性能的更好折衷，M Shao[4]提出了统计强源匿名性的概念，其核心在于动态调整监测数据发送时机和虚假数据分组的注入时机，使网络中所有节点的数据发送特性在统计意义上相同，即使具备全网侦听能力的敌手也无法识别出信源节点的位置。B Alomair[14]也采用基于统计特性的虚假数据分组注入方法保护信源位置隐私。文献[15, 16]对全网所有节点参与虚假数据分组注入方案改进，通过基站与网络节点协同工作的方式选出一定数量的具备地理多样性特征节点作为虚假信源节点，只有被选定的虚假信源节点和实际信源节点能够向网络注入数据。文献[6, 17~20]针对虚假数据分组注入方案中网络流量和能量开销过大的问题，提出通过在网络中选取部分代理节点负责汇集邻近区域内的数据分组并对虚假数据分组进行过滤和消减。

总体而言，基于信源泛化的源位置隐私保护技术能够有效抵御所有类别的源位置隐私攻击，但额外注入的虚假数据分组会引起网络负载增大、网络生存周期与事件报告延迟之间矛盾加剧、传输可靠性降低等问题。

3 系统模型和设计目标

3.1 系统模型

1) 节点数据发送模型

为抵御全网侦听流量分析攻击，节点通过调节虚假数据分组的注入时机和真实数据分组的发送时机来保证其数据发送时间间隔服从指数分布[4]：节点根据自身的数据分组发送统计特征(λ)和已发送数据分组的实际时间间隔序列($x_1, x_2, \cdots, x_{n-1}$)来确定下一次发送数据分组的时间间隔 x；若监测到事件需要报告真实数据分组，节点根据统计特征容忍系数(α, ε)计算出能够保证敌手无法觉察统计特征变化的最小时间间隔，以缩短等待发送延迟；若在预定间隔内未监测到事件，则在预定时间间隔结束时发送一个虚假数据分组；在通过缩短发送间隔发送真实数据分组之后，节点需要根据统计特征容忍系数(α, ε)计算出尽可能大的时间间隔来恢复其统计特性以备下次真实数据发送时发送延迟尽可能小。

为保证发送时间间隔序列服从指数分布，时间间隔 x 需要满足 Anderson–Darling 非参数检验

$$A^2=-n-\sum_{k=1}^{n}\frac{2k-1}{n}\left[\ln(\mathrm{F}(x_{(k)})+\ln(\mathrm{F}(x_{(n+1-k)}))\right]<c \quad (1)$$

其中，$(x_{(1)}, x_{(2)}, \cdots, x_{(n)})$为实际时间间隔序列$(x_1, x_2, \cdots, x_{n-1}, x)$按照由小到大重新排序后的序列，$n$ 为待检验序列的样本数量，F 为指数分布的概率分布函数，c 为判断序列是否符合指数分布的临界值。

为保证发送时间间隔序列的统计特征为 λ，新确定的时间间隔 x 需要满足

$$(1-\varepsilon)\frac{1}{\lambda}<\frac{1}{n}\sum_{k=2}^{n+1}x_k<(1+\varepsilon)\frac{1}{\lambda} \quad (2)$$

由式(2)可知，为保证真实数据分组的延迟尽可能小，一般情况下，发送时间间隔最小为

$$x_{\min}=(1-\varepsilon)\frac{n}{\lambda}-\sum_{k=2}^{n}x_k \quad (3)$$

显然，发送延迟与系统统计特性 λ 成反比。但如果一段时间内有多个真实事件发生，为了保持其统计特性，真实事件数据分组的发送时间间隔会增大，甚至超过系统对真实事件的最大报告延迟要求。在这种情况下，为满足系统的延迟要求，需以加重网络负载为代价增大指数分布特征参数 λ。

2) 网络模型

为简化问题且不失一般性，本文假设节点均匀部署，基站位于网络中心，网络半径为 n 跳，节点通信半径为 r，网络节点密度为 θ。当特定事件发生时，距离该事件最近的节点会将观测数据在系统允许时间内报告给基站。为减少网络负载和能耗，本文采用分簇网络结构：每个簇由簇首节点和簇内成员节点组成，成员节点将产生的数据分组经由中间节点转发到簇首节点；簇首节点过滤接收到的虚假数据分组，根据自身数据发送统计特性将接收到的真实数据分组向下游簇首节点进行转发，直至到达基站。

引理 1　分簇网络中簇半径越大，簇内节点的平均流量强度越大。

证明　设所有节点的平均数据分组发送间隔均值为 μ，节点通信半径为 r，网络的节点分布密度为 θ，簇的半径为 $radi$ 跳。

单位时间内簇的总流量 $T_{CL\text{-}radi}$ 为

$$T_{CL-radi}=\frac{1}{\mu}\sum_{j=1}^{radi}j(2j-1)\pi r^2\theta \quad (4)$$

簇内节点数量 $N_{CL\text{-}radi}$ 为

$$N_{CL-radi}=\pi(radi\cdot r)^2\theta \quad (5)$$

故半径为 $radi$ 跳的簇内节点的平均流量强度 $TS_{CL\text{-}radi}$ 为

$$TS_{CL-radi}=\frac{\frac{1}{\mu}\sum_{j=1}^{radi}j(2j-1)\pi r^2\theta}{\pi(radi\cdot r)^2\theta}=\frac{1}{6\mu}\left(4radi-\frac{1}{radi}+3\right) \quad (6)$$

显然，簇内平均流量强度随着簇半径的增大而逐渐增大。　　证毕

3.2　敌手攻击模型

本文敌手攻击模型与文献[4]相同：假设敌手不捕获或者控制任何节点，且无法区分数据分组真伪。敌手仅通过监听网络数据分组的方式，获取网络中所有节点的位置以及节点发送数据分组的时间间隔分布。更进一步，敌手可进行统计分析来比较已知的时间间隔分布与自己观测的时间间隔分布是否一致，从而获得事件发生的时间、位置等隐私信息。

3.3　设计目标

现有源位置隐私保护机制存在隐私和性能无法兼顾的问题会导致网络服务质量及资源消耗不均衡程度加剧、隐私性减弱、网络生存时间缩短等后果。本文拟研究兼顾隐私和网络性能的信源位置隐私保护机制，将传输延迟和网络负载性能确保的数据速率非均匀统计特征分布优化问题归纳为分簇网络负载总量保持不变前提下的数据分组传输延迟最大最小值的线性规划问题，进而通过对问题求解得出面向优化分簇网络结构的基于优化非均匀统计特性的源匿名方案。

4　ONSA：基于优化非均匀统计特性的源匿名协议

本文提出了基于优化非均匀统计特性的源匿名协议。ONSA 采用了分级信源泛化方式来保护源位置隐私：首先，根据不同区域内节点的平均网络负载和能耗特征将网络划分成一系列大小不同的簇组成的以基站为中心的簇环，每个簇的簇首节点负责对来自簇内普通节点和上游簇首节点的数据分组进行泛化处理以过滤不必要的虚假数据分组；接着，根据应用系统的实时性要求以及真实事件数据分组在传输过程中经过泛化处理节点（簇首节点）带来的平均延迟，对不同区域内

的节点数据发送统计特性（平均数据分组发送速率）进行优化。

4.1 非均匀分簇网络拓扑结构形成

节点根据到基站距离不同而形成大小不同的簇组成的以基站为中心的簇环，如图 1 所示。若网络中簇环数目定义为 *RING_ NUM*，那么簇环编号从里向外依次为 1, 2, …, *RING_ NUM*。距离基站的距离分别是 R_2, R_3, …,$R_{RING\text{-}NUM}$ 的环形区域内的节点为初始候选簇首节点，相应的簇半径 $radi_j$ 分别是

$$radi_2=R_2-R_1,$$
$$radi_3=R_3-2R_2+R_1 \tag{7}$$
$$radi_{RING\_NUM}=R_{RING\_NUM}-R_{RING\_NUM-1}-radi_{RING\_NUM-1}$$

此外，基站半径为 R_1 的簇的簇首节点，该簇内的成员节点直接发送数据分组到基站，而且以基站为中心的簇所在的簇环编号为 1。初始簇首候选节点广播 *R*-跳 Beacon 报文，节点根据自身到汇聚节点的距离在其所在环形区域中选择簇内通信代价最小的簇首节点，发送 Beacon 响应报文通知该簇首节点，从而确定最终的簇首节点。这样，传感网中的节点组成了以基站为中心的簇环，如图 1 所示。编号分别为 1, 2, …, *RING_NUM* 的簇环内的簇首节点数目 $NC_1, NC_2, \cdots, NC_{RING\_NUM}$ 分别为

$$NC_1=1$$
$$NC_2=\left\lceil \frac{\pi R_2}{R_2-R_1} \right\rceil$$
$$\cdots$$
$$NC_{RING\_NUM}=\left\lceil \frac{\pi R_{RING\_NUM}}{R_{RING\_NUM}-R_{RING\_NUM-1}-radi_{RING\_NUM-1}} \right\rceil \tag{8}$$

4.2 基于非均匀统计特性的分级信源泛化

非均匀分簇网络形成以后，簇首节点接收它所在簇内的成员节点和上游簇首节点发来的数据分组，过滤掉其中的虚假数据分组并转发真实的数据分组到基站。本文假定簇内成员节点与其簇首节点建立成对的密钥，相邻簇环之间的簇首节点共享一个密钥。当网络运行时，每个簇内成员节点根据多跳路由协议向其簇首节点发送加密的

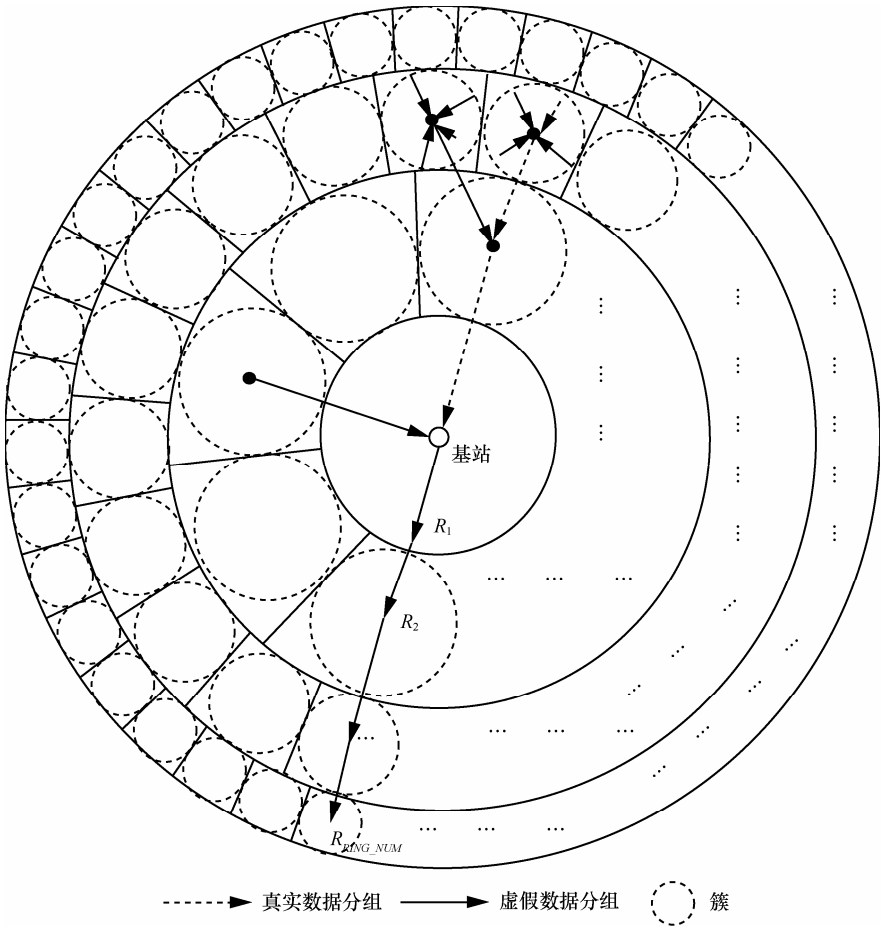


图 1 ONSA 的非均匀分簇网络拓扑结构

数据分组。为了满足事件源匿名性的要求，根据 3.1 节中的方案计算出服从指数分布的数据分组发送时间间隔。当簇内成员节点监测到事件发生时，该节点推迟发送加密的真实消息直到下一个发送时刻。

簇首节点收到簇内成员节点或者上游簇首节点发来的数据分组后首先进行解密，然后检测指示数据分组真假的控制信息：若为虚假消息，簇首节点丢弃该虚假消息；若为真实消息时，簇首节点利用与下游簇首共享的密钥重新加密该消息，并向下游簇首节点或者基站转发该消息。如果簇首节点没有收到真实的消息，簇首节点按规则将发送加密的虚假消息以保证在数据分组的发送服从统计规律。假定真实事件的报告延迟不能大于时延，由于源节点到基站路径上的簇首节点会给数据分组引入额外的延迟，因此，为了满足系统的延时要求，ONSA 需要保证数据分组在源节点与这些簇首节点上的延时之和不大于时延。很明显，源节点距离基站越远，源节点到基站路径上包含的簇首节点越多，相应的报告延时就越大。因此，这些源节点发送数据分组的时间间隔的均值 μ 需要越小才能满足系统延时要求。通过调整不同簇的节点发送速率使每个簇的真实事件报告延迟满足系统要求，并尽量取满足系统延迟要求条件下的最小发送速率以减少网络流量。

4.3　关键问题—非均匀统计特性优化

为了保证系统源匿名性和报告延迟要求，提高网络生存周期，从距离基站不同距离处的簇的尺寸和统计特性（平均数据发送速率）上对 ONSA 协议进行了考虑，将网络总流量最小化作为系统优化目标。网络总流量定义为：流量速率×消息大小×跳数(单位为 byte·hop/second)。由于所有消息的大小相同，本文只考虑流量速率和消息传输的跳数。

ONSA 待解决的优化问题可以按照下面的方式形式化描述。假设网络的半径为 *NWK_RADI*，节点的传输半径为 r，网络中节点分布的面密度为 θ，真实事件服从泊松分布的参数为 λ_{event}，网络拓扑定义见 4.1 节。此外，本文还进行如下定义。

1) 位于簇环 j ($1\leqslant j\leqslant RING\_NUM$)内所有簇的簇首节点到基站的跳数为 $dist_j$，其中 $dist_0=0$。

2) 位于簇环 j ($1\leqslant j\leqslant RING\_NUM$)内所有簇的成员节点发送数据分组(含真实和虚假)的速率为 $rate_j^{\text{mem}}$，簇首节点向外发送数据分组的速率为 $rate_j^{\text{head}}$。

3) 位于簇环 j ($1\leqslant j\leqslant RING\_NUM$)内所有簇的成员节点发送数据分组的时间间隔服从指数分布的均值为 μ_j^{mem}，簇首节点发送数据分组的时间间隔服从指数分布的均值为 μ_j^{head}，簇首节点发送真实数据分组的发送延迟为 $\alpha_j\mu_j^{\text{head}}$，其中 α_j 为簇环 j 内节点的延时压缩比，可根据文献[4]确定。

4) 位于簇环 j ($1\leqslant j\leqslant RING\_NUM$)内所有簇的簇首节点与簇内成员节点之间的最大距离等于簇的半径，为 $radi_j$。

因为每个簇都是以簇首节点为中心的圆形区域，这与整个网络以基站为中心类似。根据引理 1，位于簇环 j ($1\leqslant j\leqslant RING\_NUM$)内所有簇中距离簇首节点跳数为 i($1\leqslant i\leqslant radi_j$)的节点数目为$(2i-1)\pi r^2\theta$。因此簇环 j ($1\leqslant j\leqslant RING\_NUM$)内任一簇的开销计算

$$cost_j = rate_j^{\text{mem}} \sum_{i=1}^{radi_j} i\left(2i-1\right)\pi r^2\theta + rate_j^{\text{head}} dist_j \quad (9)$$

其中，$radi_j$ 由式(7)计算得到。

因此最小化开销为

$$cost = \sum_{j=1}^{RING\_NUM} NC_j\ cost_j \quad (10)$$

其中，NC_j 由式(8)计算。而且由式(7)和式(8)可知，$radi_j$ 与 NC_j 由 $dist_j$ 唯一确定。

此外，

$$rate_j^{\text{mem}} = \frac{1}{\mu_j^{\text{mem}}}, rate_j^{\text{head}} = \frac{1}{\mu_j^{\text{head}}} \quad (11)$$

因此，

$$cost = \sum_{j=1}^{RING\_NUM} NC_j \left(\frac{1}{\mu_j^{\text{mem}}} \sum_{i=1}^{radi_j} i\left(2i-1\right)\pi r^2\theta + \frac{1}{\mu_j^{\text{head}}} dist_j \right) \quad (12)$$

假设位于簇环 j ($1\leqslant j\leqslant RING\_NUM$)内所有簇的成员节点发送延时为 $delay_j^{\text{mem}}$，簇首节点的发送延时为 $delay_j^{\text{head}}$。如果源节点是簇环 $source$ ($1\leqslant source\leqslant RING\_NUM$)内某一簇内的节点，从该源节点发出的数据分组发送到基站的路径上要经过 k 个簇首节点的转发，这 k 个簇首节点分别位于

编号为 $forw_1, forw_2, \cdots, forw_k$ 的簇环内，为了满足系统延时要求，需要下式成立

$$delay_{\text{source}}^{\text{mem}} + delay_{forw_1}^{\text{head}} + delay_{forw_2}^{\text{head}} + \cdots + delay_{forw_k}^{\text{head}} < DELAY \tag{13}$$

其中，$DELAY$ 是一个系统参数，即

$$\alpha_{\text{source}}\mu_{\text{source}}^{\text{mem}} + \alpha_{forw_1}\mu_{forw_1}^{\text{head}} + \cdots + \alpha_{forw_k}\mu_{forw_k}^{\text{head}} < DELAY \tag{14}$$

每个簇中有 3 种数据流。第 1 个是由簇内成员节点向簇首节点发送消息，第 2 个是簇首节点向簇外发送消息，第 3 个是转发上游簇的消息。用 $traffic_j^{\text{mem}}$ 表示位于簇环 j ($1 \leqslant j \leqslant RING\_NUM$)内的任一簇中成员节点产生的流量，用 $traffic_j^{\text{head}}$ 表示簇首节点产生的流量，用 $traffic_j^{forw}$ 表示转发的上游簇的流量，位于簇环 j ($1 \leqslant j \leqslant RING\_NUM$)内的任一簇的面积记为 $area_j$，单位时间内的总流量记为 $traffic_j$。为了使传感网中的能量消耗均衡，则有下式成立

$$\frac{traffic_1}{area_1} = \frac{traffic_2}{area_2} = \cdots = \frac{traffic_j}{area_j} = \cdots = \frac{traffic_{RING\_NUM}}{area_{RING\_NUM}} \tag{15}$$

其中，

$$\begin{aligned} traffic_j &= traffic_j^{\text{mem}} + traffic_j^{\text{head}} + traffic_j^{forw} \\ &= \frac{1}{\mu_j^{\text{mem}}}\sum_{i=1}^{radi_j} i(2i-1)\pi r^2\theta + \frac{radi_j}{\mu_j^{\text{head}}} + \left(\frac{1}{NC_j}\sum_{i=j}^{RING\_NUM}\frac{NC_i}{\mu_i^{\text{head}}}\right) radi_j \end{aligned} \tag{16}$$

$$area_j = \pi\left(radi_j \cdot r\right)^2 \tag{17}$$

注意，最外层簇环内的簇没有来自上游簇的流量。

网络中单位面积内单位时间下发生真实事件的概率服从以 λ_{event} 为参数的泊松分布，因此事件的汇报速率要大于真实事件的发生速率，即

$$\pi(NWK\_RADIr)^2\frac{1}{\lambda_{\text{event}}} < NC_2(\mu_{RING\_NUM}^{\text{mem}} + \sum_{i=2}^{RING\_NUM}\mu_i^{\text{head}}) \tag{18}$$

综上，如何在不降低事件源匿名性，同时保证延时满足系统要求的情况下最小化网络流量的问题被转化为一个数学规划问题

$$\min \sum_{j=1}^{RING\_NUM} NC_j\left(\frac{1}{\mu_j^{\text{mem}}}\sum_{i=1}^{radi_j} i(2i-1)\pi r^2\theta + \frac{1}{\mu_j^{\text{head}}} dist_j\right)$$

$$\text{s.t.}\begin{cases} \alpha_1\mu_1^{\text{mem}} < DELAY \quad ① \\ \alpha_j\mu_j^{\text{mem}} + \alpha_i\sum_{i=2}^{j}\mu_i^{\text{head}} < DELAY \\ \left(2 \leqslant j \leqslant RING\_NUM, j \in N^*\right) \quad ② \\ \frac{traffic_1}{area_1} = \frac{traffic_2}{area_2} = \cdots = \frac{traffic_j}{area_j} = \cdots \\ = \frac{traffic_{RING\_NUM}}{area_{RING\_NUM}} \quad ③ \\ radi_1 + 2\sum_{i=2}^{RING\_NUM} radi_i = NWK\_RADI \quad ④ \\ radi_j \geqslant radi_{j+1} \geqslant radi_j - 1, \\ \left(1 \leqslant j \leqslant RING\_NUM - 1, j \in N^*\right) \quad ⑤ \\ \pi(NWK\_RADI \cdot r)^2\frac{1}{\lambda_{\text{event}}} < \\ NC_2(\mu_{RING\_NUM}^{\text{mem}} + \sum_{i=2}^{RING\_NUM}\mu_i^{\text{head}}) \quad ⑥ \\ 2 \leqslant RING\_NUM \leqslant RING\_MAX \quad ⑦ \end{cases} \tag{19}$$

第①个约束条件和第②个约束条件为延时要求，涉及到调节发送速率；第③个约束条件使网络中流量均衡，涉及到调节发送速率和簇大小；第④个约束条件为所有簇环半径之和应该满足的条件；第⑤个约束条件表示随着到基站距离变大，簇环半径逐渐减小，相邻簇环之间的半径最多只能相差一个节点；第⑥个约束条件表示真实事件的发生速率小于事件汇报速率。

定理 1 ONSA 的非均匀统计特性优化是 NPC 问题中的约束满足问题(CSP，constraint satisfaction problem)。

证明 $NC_j, \mu_j^{\text{mem}}, \mu_j^{\text{head}}, dist_j, radi_j (1 \leqslant j \leqslant RING\_NUM)$ 构成该问题的变量集合 X，由于 $radi_j$, NC_j 由 $dist_j$ 唯一确定，故可将其作为一组变量，记为 $(dist_j, radi_j, NC_j)$。

变量的值域 D 为

$$NC_j \in N^*, 0 < \mu_j^{\text{mem}}, 0 < \mu_j^{\text{head}}$$

$$dist_j \in N^* \wedge \left(dist_j \leqslant NWK\_RADI\right)$$

$$radi_j \in N^* \wedge \left(radi_j \leqslant NWK\_RADI\right)$$

模型中 5 个约束构成该问题的约束集合 C。由

CSP 问题的定义可知，该优化问题为 CSP 问题。

证毕

算法 1 通过使用快速回溯搜索算法找出满足约束条件的变量的值，并进行最优化求解。该算法利用最多约束变量启发式地选择出具有最少剩余值域的变量作为下一个要赋值的变量。如果存在多个具有同样最小剩余值域的变量，则选取其中一个受其他未赋值变量约束个数最多的变量，这样可以将分支因子最小化。然后使用最少约束值对选定的变量进行赋值，即对选定的变量赋值后，该值将最少地移去当前赋值之外的变量值域的值。由于值域的约束数量与非法赋值的可能性成负相关，最终仅会把具有最少约束的值赋给选定变量。即对于 ONSA 中的问题，首先对($dist_1$,$radi_1$,NC_1), ($dist_2$, $radi_2$, NC_2),…,($dist_{RING\_NUM}$, $radi_{RING\_NUM}$, NC_{RING_NUM})赋值，然后对 $\mu_2^{\text{head}},\mu_3^{\text{head}},\cdots,\mu_{RING\_NUM}^{\text{head}}$ 赋值，最后对 μ_1^{mem}, $\mu_2^{\text{mem}},\cdots,\mu_{RING\_NUM}^{\text{mem}}$ 赋值。该算法在对一个变量赋值后，会使用前向检验算法消除未赋值变量值域中与此次赋值相关的不满足约束条件的值。

算法 1　ONSA 模型的快速回溯搜索算法

Input：网络规模: *NWK_RADI*, 系统延迟需求: *DELAY*, 真实事件发生服从的泊松分布参数: λ

Output：变量集 X 中的最优可行分配集

Procedure ONSA-BACKTRACKING(A, var-domains):

1) 初始化变量集 X:($dist_j$,$radi_j$,NC_j), μ_j^{head}, $\mu_j^{\text{mem}}(1 \leqslant j \leqslant RING\_NUM)$, 变量的值域: D, 约束集：C，分配集：A，未赋值变量的值域: var-domains
2) if (分配集 A 满足约束 C) then
3) 比较并记录最小能耗 $cost$ 和 A
4) return
5) else
6) 用最多约束变量启发式算法从 X 中选择一个不在 A 中的变量 x
7) 用最少剩余值域启发式算法从 D 中选出 x 的一个序列 d
8) foreach value v in d do
9) add (x←v) to A
10) if (A 满足条件) then
11) var-domains←FORCHECKING (var-domains, x, v, A)
12) if (变量值域为空) then
13) return failure
14) else
15) result←ONSA-BACKTRACKING (A, var-domains)
16) if(result≠failure)then
17) return result
18) end if
19) end if
20) end if
21) end foreach
22) end if

算法 1 在对一个变量赋值后，会使用前向检验算法消除未赋值变量值域中与此次赋值相关的不满足约束条件的值，前向检验算法如算法 2 所示。

算法 2　前向检验算法

Input：未赋值变量的值域: var-domains,分配变量 x 及其值 v,分配集 A

Output：var-domains 的赋值

Procedure FORCHECKING (var-domains, x, v, A):

1) foreach 从 X 中选出不在 A 中的变量 y do
2) foreach C 中与 y 有关的约束 c 以及 A 中的变量 do
3) 从 y 中删除不满足约束 c 的值
4) end foreach
5) end foreach
6) return var-domains

5　性能仿真和安全性分析

5.1　性能仿真与分析

本节通过仿真实验比较 ONSA、FitProbRate[4]、TFS[6]的性能。在实验中，网络通信半径为 r 个单位长度，网络节点分布的面密度为 θ=2，即单位面积区域中节点数目为 2，系统真实事件报告延迟要求为 1 s，真实事件服从的泊松分布的参数 λ_{event} 为 1/20 或 1/100，网络规模从 5 到 100 跳不等。FitProbRate 和 TFS 的节点发送虚假数据分组的时间间隔均服从指数分布，均值 μ 都为 20 s。系统参数 *DELAY* 被设置为 1 s。本节将从延时、负载均衡性、网络流量以及网络寿命等方面进行考察。

图 2 给出了在不同网络规模下（hop 为 6 跳和

20 跳），FitProbRate、TFS 和 ONSA3 种方案下真实数据分组平均延时的比较。如图 2 所示，FitProbRate 方案和 ONSA 中的延时能够满足系统的延时要求，然而 TFS 方案中的延时不能满足系统的延时要求，尤其当网络规模较大时，TFS 方案中的延时无法满足应用的需求。因此，在下面的分析中本文仅仅考虑能满足需求的 FitProbRate 方案和 ONSA。在图 2(b)中，FitProbRate 方案随着跳数增加，延迟会超过系统要求，这是由于在 FitProbRate 中所有数据分组都发回基站，在基站附近会发生拥塞、分组丢失情况。

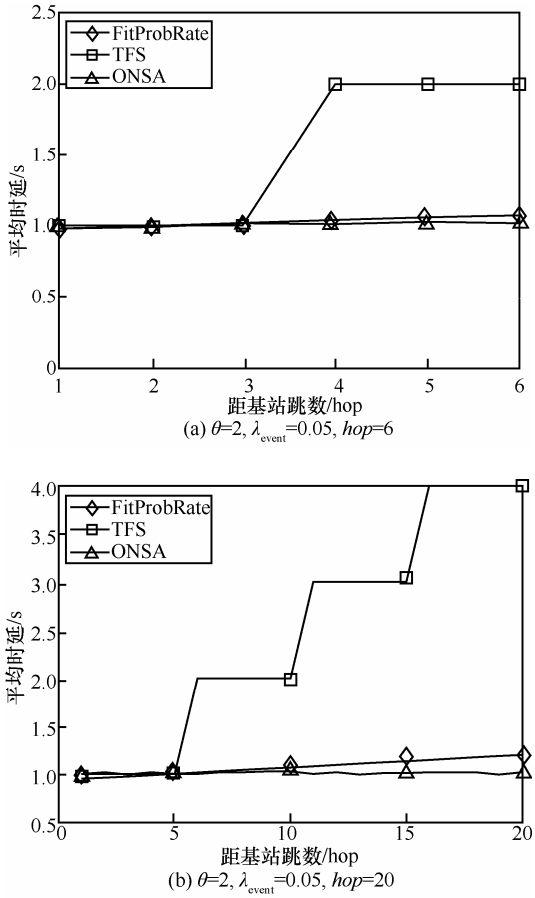


图 2 在不同的网络规模下的平均延时比较

ONSA 中位于不同簇环内的簇首节点和簇内成员节点对真实数据分组的报告延时如图 3 所示。距离基站越远的簇首节点其数据分组发送速率越慢、相应延时越大，而簇内成员节点发送数据分组的速率越快、延时越小。这样能够平衡簇首和簇成员节点的延时，使各簇内真实数据延时保持均衡。由图 3 可知，网络中任何节点作为源节点的真实事件报告延迟均未超过系统规定延时。

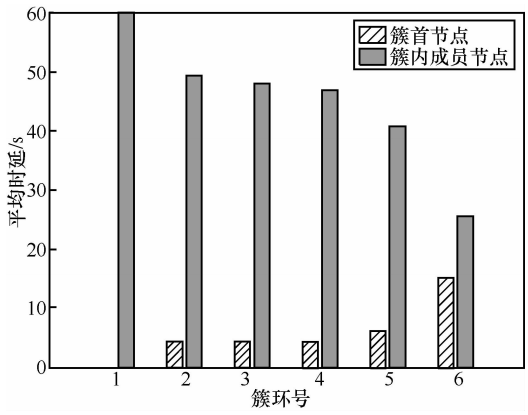
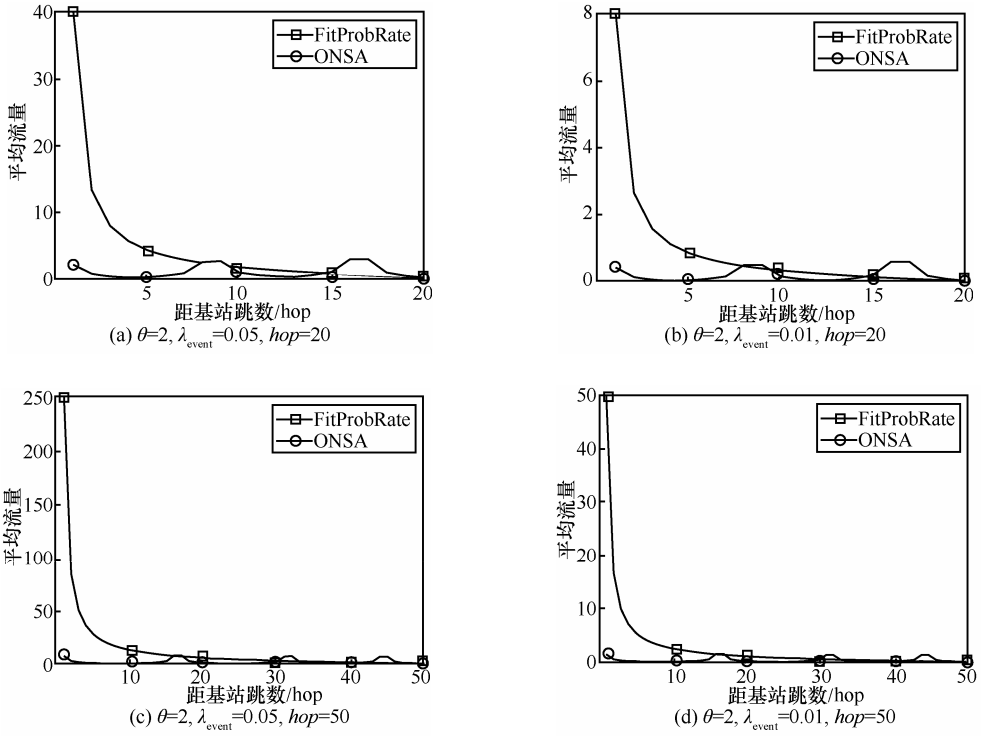


图 3 位于不同簇环的簇内节点的平均延时
（*hop*=100, *DELAY*=1 s, θ=2, λ_{event}=0.5）

图 4 描述了不同网络规模和不同事件发生分布下，FitProbRate 方案和 ONSA 中距离基站不同跳数节点的平均流量。实验结果表明，FitProbRate 方案中平均网络流量随着到达基站距离的减小而急剧增长，基站附近节点的平均流量远大于远离基站区域节点的平均流量。然而，ONSA 中的平均网络流量是基本均衡的，只是在某些区域附近有轻微波动，这是由于簇内成员节点向簇首节点发送数据分组，而导致簇首节点附近流量稍高于其他区域。通过比较发现：网络规模越大，FitProbRate 方案的“漏斗效应”越明显，而 ONSA 的性能基本保持稳定。而且在不同真实事件发生概率下，ONSA 都具有良好的均衡性，网络流量的均衡性能够延长网络寿命。

图 5 表明与 FitProbRate 方案相比，ONSA 可以显著地减少网络流量，而且随着网络的规模增大，网络流量减少的程度也变大。ONSA 与 FitProbRate 方案的网络寿命如图 6 所示。一方面，ONSA 可以极大地延长网络寿命；另一方面，ONSA 对网络寿命提升的程度比对网络流量的减少程度更大，这是因为 ONSA 中网络流量是均衡的。

图 7 和图 8 考察了真实事件服从泊松分布的期望 λ_{event} 以及系统延时对协议性能的影响。由图 7 可知，FitProbRate 和 ONSA 方案的总流量均会随着真实事件发生概率的减小而减少。与 FitProbRate 相比，ONSA 在不同 λ_{event} 下均能够显著地减少网络流量。图 8 给出网络流量随系统延时要求和真实事件发生概率的变化情况。显然，系统可以容忍的延时越大，节点发送数据分组的时间间隔就可以越大，网络流量也越小。



(a) θ=2, λ_{event}=0.05, *hop*=20　(b) θ=2, λ_{event}=0.01, *hop*=20

(c) θ=2, λ_{event}=0.05, *hop*=50　(d) θ=2, λ_{event}=0.01, *hop*=50

图 4　不同网络规模下的平均流量比较

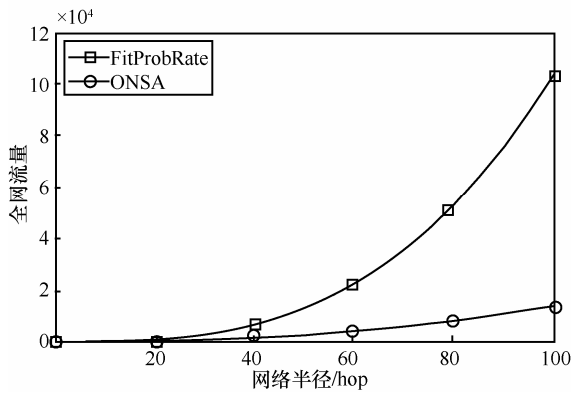


图 5　不同的网络规模下的网络总流量
（θ =2, λ_{event}=1/20, *DELAY*=1 s）

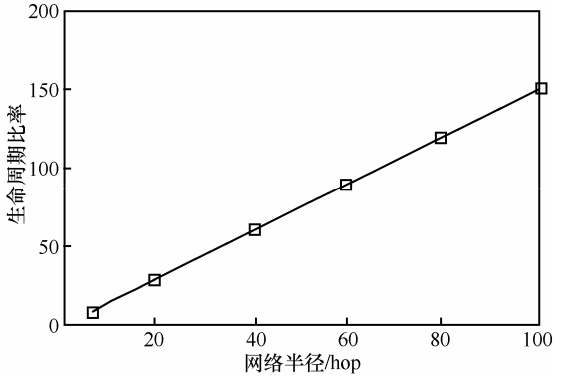


图 6　在不同网络规模下 ONSA 与 FitProbRate 的网络生命周期比率
（θ=2, λ_{event}=1/20, *DELAY*=1 s）

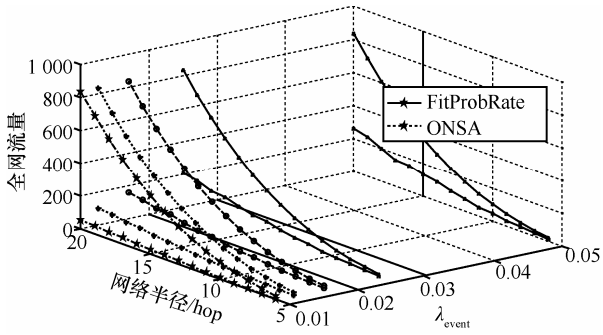


图 7　在不同 λ_{event} 下的网络流量比较（θ=2, *DELAY*=1 s）

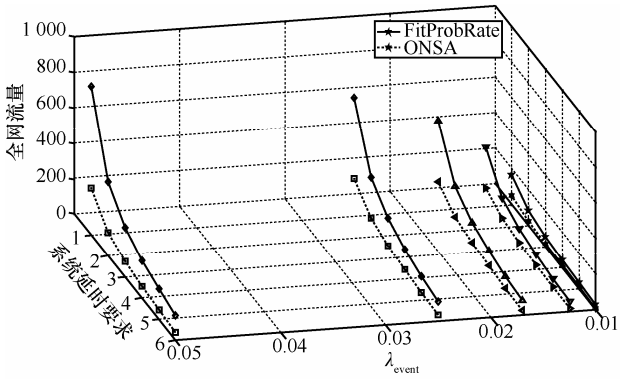


图 8　在不同延时和 λ_{event} 下的网络流量比较（θ=2, *hop*=20）

当系统延时确定后，数据分组在簇首节点的平均延迟不会随着真实事件发生概率的变化而显著

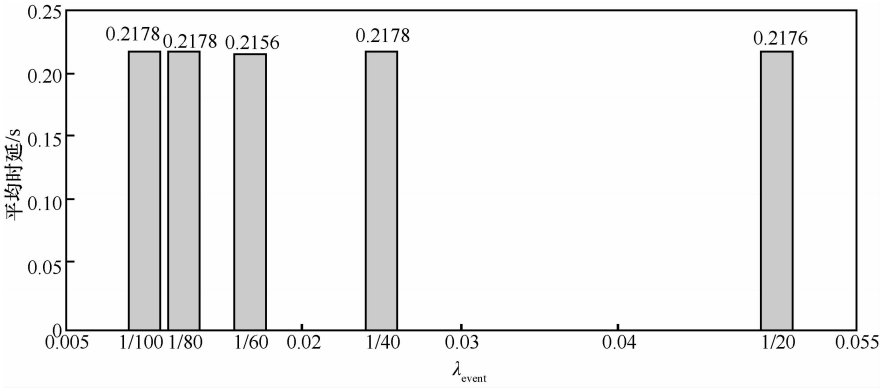


图 9 不同的 λ_{event} 下簇首节点的平均延迟（*DELAY*=1 s, θ=2）

变化，如图 9 所示。这是因为，若发生真实事件的概率减小，虽然节点发送数据分组的时间间隔会变大，但同时延时压缩比会减少。

ONSA 模型的快速回溯搜索算法在对 $(dist_i,radi_i,NC_i)$ $(1\leqslant i\leqslant RING_NUM)$内赋值时，限定簇环数 *RING_NUM* 最大为 *RING_MAX*。这是因为过多的簇环数使路径上的簇首节点数过多，为了满足延迟要求，簇首节点的发送速率需要加快，反而导致能耗更高。在对 $\mu_2^{\text{head}},\mu_3^{\text{head}},\cdots,\mu_{RING\_NUM}^{\text{head}}$ 和 $\mu_1^{\text{mem}},\mu_2^{\text{mem}},\cdots,\mu_{RING\_NUM}^{\text{mem}}$ 搜索赋值时，设置搜索的步长 *step*，这样能快速计算出近似最优解。当簇环数目为 k ($k \leqslant RING\_MAX$)时，搜索算法在最坏情况下时间复杂度为 $O(R2^k l^k m^k)$，R 表示网络半径，l 表示对每一个簇首节点的发送间隔均值 μ_i^{head} $(2\leqslant i\leqslant k)$搜索赋值时的执行次数，$m$ 表示对每个簇环内成员节点发送间隔均值 μ_i^{mem} $(1\leqslant i\leqslant k)$搜索赋值时的执行次数，$l$ 和 m 的最大值为 *DELAY*/*step*。因此，ONSA 模型的快速回溯搜索算法在最坏情况下的时间复杂度为 $O(R(2lm)^k)$。

5.2 安全性分析

ONSA 与 FitProbRate 中数据分组发送统计特性均服从相同的指数分布，因此文献[4]中的安全性分析适合于 ONSA。

文献[6]指出采用虚假流量注入的匿名机制具有事件源不可观测性，其定义如下。

定义 1 如果对于攻击者 A 可能执行的任何一个观测 O，事件 E 发生的概率等于已知观测 O 的条件下事件 E 发生的概率，即 $\forall O$, $P(E)=P(E|O)$，那么事件 E 是不可观测的。

定义 2 如果一个系统中任何可能发生的事件都是不可观测的，即 $\forall$ E, $\forall O, P(E)=P(E|O)$，那么称该系统具有事件源不可观测的性质。

定理 2 ONSA 具有事件源不可观测的性质。

证明 定义 1 中的 $P(E)=P(E|O)$意味着事件 E 和观测 O 是独立的，$P(E\cap O)=P(O)P(E|O)= P(O)P(\text{E})$。因此，根据上述定义可以证明：任一事件 E 和观测 O 是独立地证明 ONSA 均具有事件源不可观测的性质。

攻击者可以侦听分析出每个簇的簇内成员节点发送数据分组的时间间隔服从某一概率分布。然而，由于消息是加密的，而且所有消息长度相同，攻击者无法区分消息真实性。这些消息经多跳路径转发到簇首节点。从簇首节点发出的消息总是按照相同的方式被转发到基站，攻击者无法知道哪些消息被丢弃，哪些消息被重加密后转发出去。因此，即使可以执行任何全局侦听分析，攻击者无法获得关于真实事件的额外信息。真实事件 E 的发生是和攻击者执行的观测 O 独立的。所以，真实事件对于攻击者来说是不可观测的。根据定义 2，ONSA 具有事件源不可观测的性质。

证毕。

6 结束语

本文根据传感器网络中事件报告延迟和网络负载的非均衡分布特征，提出了基于优化非均匀统计特性的分级泛化源匿名协议。ONSA 以远离基站区域内节点的剩余能量和靠近基站区域内节点的剩余报告延迟为代价，通过优化不同区域内的节点数据发送速率和簇尺寸来高效地保护信源节点位置隐私。仿真实验结果和理论分析表明：ONSA 一方面能有效抵御具备全网侦听能力的源位置隐私攻击，另一方面还能显著减少并均衡网络能量消耗，延长网络生存周期，同时满足信源位置隐私保护的实时性要求。

参考文献：

[1] OZTURK C, ZHANG Y, TRAPPE W. Source-location privacy in energy-constrained sensor network routing[A]. Proc of ACM SASN, Washington[C]. DC, USA, 2004. 88-93.

[2] KAMAT P, ZHANG Y, TRAPPE W, *et al*. Enhancing source-location privacy in sensor network routing[A]. Proc of IEEE ICDCS[C]. Columbus, OH, USA, 2005.599-608.

[3] CONTI M, WILLEMSEN J, CRISPO B. Providing source location privacy in wireless sensor networks: a survey[J]. IEEE Communications Surveys & Tutorials, 2013,15(3):1238-1280.

[4] SHAO M, YANG Y, ZHU S, *et al*. Towards statistically strong source anonymity for sensor networks[A]. Proc of IEEE INFOCOM[C]. Phoenix, AZ, USA, 2008.51-55.

[5] AHN G.S., HONG S.G., MILUZZO E. Funneling-MAC: a localized, sink-oriented MAC for boosting fidelity in sensor networks[A]. Proc of ACM SenSys[C]. Boulder, Colorado, USA, 2006.293-306.

[6] YANG Y, SHAO M, ZHU S, *et al*. Towards event source unobservability with minimum network traffic in sensor networks[A]. Proc of ACM WiSec[C]. Alexandria, Virginia, USA, 2008.77-88.

[7] 陈娟, 方滨兴, 殷丽华等. 传感器网络中基于源节点有限洪泛的源位置隐私保护协议[J]. 计算机学报, 2010, 33 (9): 1736-1747.
CHEN J, FANG B X, YIN L H, *et al*. A source-location privacy preservation protocol in wireless sensor networks using source-based restricted flooding[J]. Chinese Journal of Computers, 2010, 33(9): 1736-1747.

[8] XI Y, SCHWIEBERT L, SHI W. Preserving source location privacy in monitoring-based wireless sensor networks[A]. Proc of IEEE IPDPS[C]. Rhodes Island, Greece, 2006.1-8.

[9] LI Y, REN J. Source-location privacy through dynamic routing in wireless sensor networks[A]. Proc of IEEE INFOCOM[C]. San Diego, California, USA, 2010.1-9.

[10] YAO L, KANG L, DENG F, *et al*. Protecting source–location privacy based on multirings in wireless sensor networks[J]. Concurrency Computat.: Pract. Exper, 2013. DOI: 10.1002/cpe.3075.

[11] OUYANG Y, LE Z, CHEN G, *et al*. Entrapping adversaries for source protection in sensor networks[A]. Proc of IEEE WoWMoM[C]. NewYork, USA, 2006.23-34.

[12] RIOS R, LOPEZ J. Exploiting context-awareness to enhance source-location privacy in wireless sensor networks[J]. Computer Journal, 2011, 54(10): 1603-1615.

[13] MEHTA K, LIU D, WRIGHT M. Location privacy in sensor networks against a global eavesdropper[A]. Proc of IEEE ICNP[C]. Beijing, China, 2007.314-323.

[14] ALOMAIR B, CLARK A, CUELLAR J, *et al*. Towards a statistical framework for source anonymity in sensor networks[J]. IEEE Trans on Mobile Computing , 2013,12 (2): 248-260.

[15] CUELLAR J, POOVENDRAN R. Toward a statistical framework for source anonymity in sensor networks[J]. IEEE Trans on Mobile Computing, 12(2):248-260, 2013.

[16] MEHTA K, LIU D, WRIGHT M. Protecting location privacy in sensor networks against a global eavesdropper[J]. IEEE Trans on Mobile Computing, 2012, 11(2): 320-336.

[17] MAHMOUD M M, SHEN X S. Secure and efficient source location privacy preserving scheme for wireless sensor networks[A]. Proc of IEEE ICC[C]. Ottawa, CANADA, 2012.1-5.

[18] LIGHTFOOT L, LI Y, REN J, Preserving source-location privacy in wireless sensor network using Star routing[A]. Proc of IEEE Globecom[C]. Miami, Florida, USA, 2010.1-5.

[19] MAHMOUD M M, SHEN X S. A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks[A]. IEEE Trans on Parallel and Distributed Systems, 2012, 23(10): 1805-1818.

[20] LI Y, REN J, WU J. Quantitative measurement and design of source-location privacy schemes for wireless sensor networks[J]. IEEE Trans on Parallel and Distributed Systems, 2012,23(7):1302-1311.

作者简介：

牛晓光（1979-），男，河北保定人，武汉大学副教授、硕士生导师，主要研究方向为移动计算、无线传感网和信息安全。

魏川博（1990-），男，贵州贵阳人，武汉大学硕士生，主要研究方向为无线传感网和网络安全。

冯为江（1991-），男，河南濮阳人，国防科学技术大学硕士生，主要研究方向为无线传感网和机器学习。

彭国军[通信作者]（1979-），男，湖北荆州人，武汉大学副教授、博士生导师，主要研究方向为软件安全、恶意代码检测、电子证据等。E-mail:guojpeng@whu.edu.cn。

张焕国（1945-），男，河北元氏人，武汉大学教授、博士生导师，主要研究方向为信息安全、密码学、可信计算等。