无线传感器网络恶意节点溯源追踪方法研究

杨 峰[1], 周学海[1,2], 张起元[1], 谢 婧[1], 章曙光[1,3]
(1. 中国科学技术大学计算机系, 安徽合肥 230027; 2. 中国科学技术大学苏州研究院, 江苏苏州 215123;
3. 安徽建筑工业学院电子与信息工程学院, 安徽合肥 230022)

摘 要: 传感器节点可能被攻击者俘获用来发送大量虚假数据, 从而耗尽整个网络的资源. 本文提出一种实用的溯源追踪解决方案: 基于概率包标记算法, 每个节点按照一定概率标记其转发的包, 标记信息填写于包头中的确定域, 通过收集到足够多的数据包, 汇聚节点能够重建一条到源节点的路径. 本文证明了此方案能够应对对所有类型的攻击, 并针对基本标记方法的不足提出了两种改进标记方法. 实验结果表明该溯源追踪解决方案是高效以及实用的.

关键词: 无线传感器网络; 安全性; 溯源追踪
中图分类号: TN915.08 文献标识码: A 文章编号: 0372-2112 (2009) 01-0202-05

A Practical Traceback Mechanism in Wireless Sensor Networks

YANG Feng[1], ZHOU Xue-hai[1,2], ZHANG Qi-yuan[1], XIE Jing[1], ZHANG Shu-guang[1,3]
(1. *Department of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, China;*
2. *Suzhou Institute for Advanced Study, USTC, Suzhou, Jiangsu 215123, China;*
3. *Institute of Electronics and Information Engineering, Anhui Institute of Architecture and Industry, Hefei, Anhui 230022, China*)

Abstract: Sensor nodes can be compromised by attackers and inject large amounts of bogus data to exhaust network resources. A practical traceback mechanism is proposed, in which each forwarding node marks packets with a certain probability. By collecting enough packets, sink node will construct a path back to the source node. It is proved that the mechanism is secure against all forms of attacks and other two marking methods are proposed to improve the performance of the mechanism. Simulation results show that the mechanism is efficient and practical.

Key words: wireless sensor networks; security; traceback

1 引言

无线传感器网络的应用越来越广泛, 同时它所面对的安全威胁也越来越大, 因为无线传感器网络经常需要部署在危险乃至敌对的环境中[1]. 如果一个节点被俘获, 则敌对方就掌握了它所存储的所有信息, 并且能够利用它发送大量虚假信息[2,3], 迅速耗尽网络资源, 例如能量、带宽、计算能力、存储空间等等. 在互联网中, 溯源追踪技术主要被用于处理 DDoS(分布式拒绝服务)攻击[4], 但是无线传感器网络中的溯源追踪与互联网中有很大不同, 在互联网中一般认为路由器得到很好的保护, 而无线传感器网络中并没有单一的路由设备, 路由是由普通网络节点完成的, 每个节点都有可能被俘获, 而被俘获节点可能相互配合, 发动各种类型的攻击. 现存的互联网溯源追踪方法[5~8], 还不能应对多个节点协同攻击的情况.

因为无线传感器网络节点的计算能力有限, 所以之前的大部分研究[2,9,10]都基于对称密钥, 即网络节点与汇聚节点共享密钥. 该领域的研究主要集中于途中过滤[2,9], 即每个节点都可以部分地过滤掉到达本节点的虚假数据, 但是并不能确认发动攻击的节点, 所以只是被动地应对, 而没有主动地反击. 文献[10]采用概率包标记的方法定位攻击节点, 但却需要假定可以向包头写入任意多数据, 即每个节点如果决定标记本数据包, 则将标记数据添加到包头, 因此数据包可能越来越长, 最终不得不进行分片, 处理起来极为困难, 因此其只具有理论上的意义. 文献[11]采用基于公钥体系的解决方案将攻击节点定位于两步的范围内, 但是因为传感器节点计算能力有限, 基于公钥体系的解决方案并不适用于大多数情形.

基于以上情况, 本文提出一种实用于无线传感器网络的溯源追踪解决方案. 该方案基于对称密钥体系, 利

收稿日期: 2007-09-19; 修回日期: 2008-06-15
基金项目: 安徽省自然科学基金(No. 070412030); 高等教育博士学科点专项基金(No. 20050358040); 安徽省教育厅自然科学重点项目(No. KJ2008A103); 国家自然科学基金(No. 60873221)

用包头中的四个域实现溯源追踪, 每一个决定标记的节点标记其中的两个域, 分别为节点域和数据域. 为了确保安全性, 我们没有增加标志位来表示本数据包的标记状态, 而是由节点通过直接判断域中数据来确定, 结果显示, 该方法既能够保证安全性, 又节省了标记需要的空间.

2　基本解决方案

2.1　术语定义与方案综述

为了描述本方案, 首先定义如下术语:

源节点: 发送大量虚假数据包的节点.

被俘获节点: 所有被俘获的节点, 包括源节点以及与源节点配合发动攻击的其他节点等.

定位节点: 溯源追踪方案定位到的节点.

定位成功: 定位节点即被俘获的节点, 或者被俘获节点在定位节点的上游一步邻居之内.

如此定义定位成功是因为: 源节点可以完全不标记包, 这样任何基于标记的溯源追踪解决方案都只能定位到源节点的一步邻居范围内.

该解决方案的基本思路为: 每一个数据包最多可以被两个节点标记, 节点收到数据包后, 将按照一定概率决定是否标记此包. 每个被标记过的数据包都保存了部分路径信息, 因此汇聚节点只需要收集到足够数量的包, 就可以溯源追踪到源节点. 接下来的两小节将分别介绍标记与溯源追踪算法.

2.2　基本标记方法(BPPM)

如图 1 所示, 每个节点标记需要占用包头中的两个域, 每个包最多可被两个节点标记. 任意节点 u 拥有两个与汇聚节点共享的密钥以及对应的加密函数– h_u 与 – h_u' 分别用来加密本节点地址以及收到的数据包. 设原始数据包为 M, 所有用于实现此溯源追踪算法的域初值都为 0. 当节点 i 收到包时, 首先以概率 p 决定是否标记本包. 如果决定标记, 则检查 id_1 域与 $data_1$ 域是否全 0, 若为全 0, 则将 $h_i(i)$ 添加到 id_1 域, 将 $h_i'(M|i)$ 添加到 $data_1$ 域, 如果不全为 0, 再检查 id_2 域以及 $data_2$ 域, 如果为全 0, 则将 $h_i(i)$ 添加到 id_2 域, 将 $h_i'(M|i)$ 添加到 $data_2$ 域内; 若仍然不全为 0, 则节点覆盖标记 id_1 以及 $data_1$ 域, 同时将 id_2 以及 $data_2$ 域重置为 0. 节点 i 标记后的数据包称为 M_i.

2.3　溯源追踪

汇聚节点维护两张表, 第一张表项为 $\{u, u'\}$, 其中 u 为普通节点的节点号, u' 是 u 经过加密后的值; 第二张表项为 (u, k_u), u 为节点号, k_u 为此节点对应的解密密钥. 同时汇聚节点维护两个集合, 第一个为 S, 即所有可能的定位节点; 第二个为 T, 即所有标记过数据包的节点. 当发生虚假信息注入攻击时, 汇聚节点收到一个

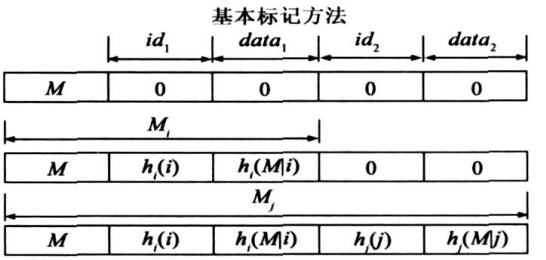

图1　基本标记方法

包, 包结构如图 1 所示, 首先检测包头中的 id_2 与 $data_2$ 域, 如果 id_2 与 $data_2$ 域非全 0, 通过查表得到此标记节点 j, 然后用此节点对应的解密密钥 k_j 解密 $data_2$ 中的数据, 得到的数据与 M_i 进行比较, 如果不相同则验证不通过, 不继续处理此包, 如果通过验证则记此节点为 t, 继续检验前一个节点标记, 如果通过验证, 则记此节点为 s, 不通过则不处理. 如果 id_2 与 $data_2$ 为全 0, 则检验 id_1 与 $data_1$ 域, 如果也为全 0, 则不做任何操作, 否则检验此标记节点, 如果检验通过, 则仍记此节点为 s. 如果 s 不属于 T, 则将 s 加入 S 与 T. 如果 t 不属于 T, 则将此节点加入 T 中, 如果 t 属于 S, 则将 t 从 S 中删除. 图 2 显示了当数据包依次被 s 与 t 节点标记的情况下, 汇聚节点的处理方式. 系统工作一段时间之后, 当 S 在相当长时间段内仅仅只有一个元素时, 则确定此节点为定位节点.

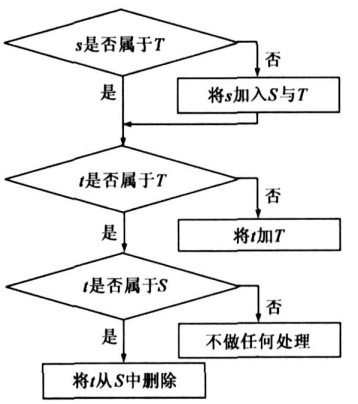

图2　溯源追踪示意图

3　改进标记方法

显然基本标记方法存在一个缺陷, 即上游节点标记可能被下游节点标记覆盖, 因此距离汇聚节点越远的节点, 其标记被收集到的概率就越低, 如果距离汇聚节点很远的节点被俘获并发动攻击, 汇聚节点将需要收集大量的包才能够将其定位. 因此我们将提出两种改进的标记方法.

3.1　等概率包标记方法(EPPM)

如果某节点到达汇聚节点需要 j 跳, 则令 P_j 表示该节点标记数据包的概率, $£_j$ 表示此节点标记数据包且该标记被汇聚节点收集到的概率. p 表示最后一个节点标记数据包的概率, 则 P_j 可以表示为 p 与 j 的函数. EPPM 的目的是对于每一个 j, P_j 的值都约等于 p. 在 EPPM 中, 每个节点标记被收集到的概率都基本相等, 因

此该方法对所有节点更加公平, 同时当攻击节点较远时, 汇聚节点可以更加迅速地将其定位.

但是并不存在一种计算 P_j 的函数可以适用于所有网络, 因此需要根据不同网络拓扑来确定其计算 P_j 的函数. 直观上, 距离汇聚节点越远, 其标记概率应该越大, 但是一个节点标记概率增大, 则覆盖上游节点标记的概率也将随之增大, 因此, 需要大量实验与衡量才能确定节点的标记概率.

我们首先考虑一种典型的网络拓扑结构, 其中普通节点的数据到达汇聚节点最多需要 20 跳, 我们假设 p 为 0.06. 经过大量的实验与衡量, 我们得到如表 1 所示的 P_j, 同时, 该表也列出了相应的 $£_j$. 由表 1 可以看出, P_j 随着 j 的增加而增加, 增加的比例需要经过大量实验才能得出, 而最终可以实现 $£_j$ 与 p 的差距在 1.5% 之内.

表 1 EPPM 与 ENPM 中的节点标记概率

j	EPPM		ENPM
	P_j	$£_j$	P_j
1	0.060	0.0600	0.060
2	0.062	0.0601	0.063
3	0.064	0.0602	0.066
4	0.066	0.0600	0.069
5	0.068	0.0597	0.072
6	0.071	0.0601	0.076
7	0.074	0.0602	0.080
8	0.077	0.0602	0.084
9	0.080	0.0599	0.088
10	0.084	0.0601	0.093
11	0.088	0.0601	0.098
12	0.092	0.0599	0.103
13	0.096	0.0595	0.108
14	0.101	0.0596	0.114
15	0.106	0.0597	0.120
16	0.111	0.0598	0.126
17	0.116	0.0601	0.132
18	0.121	0.0609	0.139
19	0.123	0.0609	0.146
20	0.123	0.0609	0.151

3.2 等数目包标记方法(ENPM)

3.1 节中介绍的 EPPM 使得每个节点标记被汇聚节点收集到的概率基本相等, 解决了基本标记方法中距离汇聚节点较远节点标记被收集到的概率偏低的问题. 但是包括 EPPM 在内, 目前所有基于概率包标记方法的溯源追踪方案都存在这样一个问题: 攻击节点距离汇聚节点越远, 汇聚节点就需要收集越多的包才能够将其定位. 为了解决这一问题, 我们提出了等数目包标记方法(ENPM), 该标记方法的目的是无论攻击者选择哪个节点发动攻击, 汇聚节点都可以通过收集到大致相等数目的包而将其定位(攻击节点为汇聚节点的一步邻居时除外, 因为在这种情况下, 汇聚节点可以通过收集非常少的数据包来将其定位).

同样, 在 ENPM 中也并不存在一种适用于所有网络的函数来为每个节点确定 P_j, 但是基本思想可以确定: 随着 j 的增加, P_j 以比 EPPM 中更快的速度增加. 表 1 也验证了我们这一想法. 假设采用与 3.1 节中相同的场景, 表 1 列出了为实现 ENPM 所需要的 P_j, 其增加速度明显变快. 更多的实验与分析将在第 5 节中给出.

4 安全性分析

图 3 为一个攻击示意图, 节点 0 与节点 5 被俘获, 节点 0 发送大量虚假信息, 节点 5 进行配合, 节点 7 为汇聚节点. 下面将针对各种可能的攻击模型, 对本方案的安全性进行全面的分析.

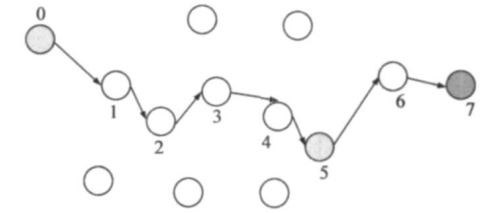

图 3 攻击示意图, 节点0与5被俘获, 节点7为汇聚节点

不标记: 如果节点 0 总是不标记它所产生的包, 则汇聚节点 7 可以追踪到节点 1, 即节点 0 的一步邻居节点.

移除标记: 即节点 5 移除上游节点的标记. 因为我们将每个节点的 id 标记都进行了加密, 所以节点 5 不能选择性地移除某一特定节点的标记, 如果它将所有标记都移除, 并且自己也不标记或者错误标记, 则汇聚节点最终可以定位到节点 6, 即被俘获节点 5 的一步邻居. 而如果它正确标记, 则最终会定位到自身.

标记重排: 即节点 5 试图重新排列上游节点的标记. 假设节点 1 与节点 2 标记了包, 因为节点将自己的 id 进行了加密, 节点 5 不可能知道是哪一个节点标记了此包, 所以也没有办法对标记进行重新排序.

变更标记: 即节点 5 试图更改上游节点的标记. 同样因为这个节点的 id 进行了加密, 所以节点 5 并不知道哪些节点标记了包, 如果它随机地更改包, 则必然有一些包能够到达汇聚节点 7, 最后追踪到节点 0 或者节点 0 的一步邻居内. 如果它更改所有包, 并且自己也不标记包或者错误标记包, 则汇聚节点在回溯追踪的过程中最终会定位到节点 6, 同样定位到被俘获节点 5 的一步邻居; 如果它正确标记包, 则定位到自身.

选择性丢弃: 节点 5 选择性地丢弃上游节点的标记. 分析同变更标记.

由以上分析可以看出, 该方案可以应对各种可能的攻击形式, 因此具有很强的实用性.

5 仿真与分析

我们进行了仿真实验来验证之前的分析以及衡量各个方案的性能. 我们仍然考虑 3.1 节中提到的典型网络设置, 在 EPPM 与 ENPM 中仍然令 $p = 0.06$, 而在 BPPM 中令 $p = 0.1$. 因为可以证明, 在 BPPM 中, 当 $np = 2$ 时(n 表示数据包被转发的次数), 数据包被两个节点标记的概率最大, 因此为了保证当 $n = 20$ 时, BPPM 有

最好的性能, 我们令 p= 0.1.

汇聚节点使用 2.3 节方法定位被俘获节点, 我们测量在 100 次运行中, 有多少次运行不能正确定位节点. 图 4 显示了当汇聚节点收到 400 个虚假数据包时, 采用每种标记方法, 分别有多少次运行不能正确定位节点. 图 5 显示的是汇聚节点收到 600 个虚假数据包的情况. 由两图都可以看出, 当攻击节点距离汇聚节点较近时(小于 10 跳), 采用 BPPM 的方案可以获得比较好的性能, 但是当攻击路径长度再增加时, 采用 BPPM 的方案失效率急剧上升. 因此当网络范围较小时, BPPM 已经可以取得较好效果. 当攻击路径相对较长时(大于 10 跳), 采用 EPPM 与 ENPM 的方案比采用 BPPM 的方案性能高很多.

通过图4 与图5 也可以看出, 采用 EPPM 的方案, 其失效率基本是随着攻击路径长度的增加而增加, 所以攻击者会倾向于选择距离汇聚节点较远的节点发动进攻. 而采用 ENPM 的方案, 其失效率与攻击路径的长度并没有如此明显的正相关, 因此当网络规模较大时, 采用 ENPM 的方案可以得到最好的性能. 由图 5 可以看出, 当有 600 个包被收集到的时候, 无论攻击者选择哪个节点发动攻击, 汇聚节点都可以以非常高的概率定位源节点(约 90%), 这样的性能达到了实用的标准.

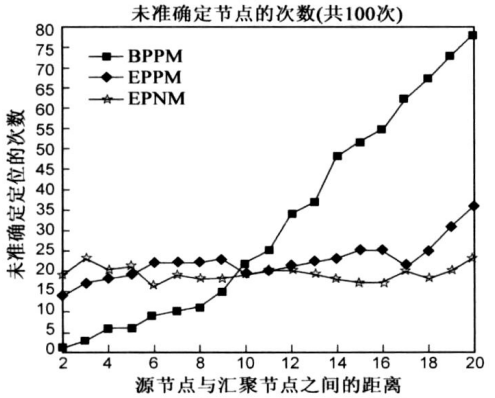

图4　汇聚节点收到400个包的情况

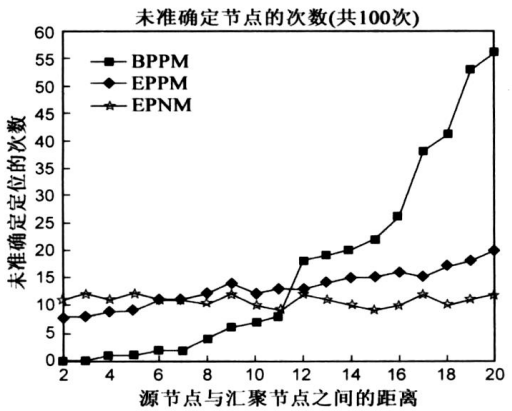

图5　汇聚节点收到600个包的情况

6　总结与展望

如何应对传感器网络中的虚假数据注入问题已经引起人们越来越多的重视[2,3,9~11], 本文提出一种实用的无线传感器网络恶意节点溯源追踪解决方案. 转发节点按照一定概率标记数据包中的固定域, 汇聚节点通过收集一定数目的数据包定位被俘获节点. 同时本文针对基本标记方法中上游节点标记会被下游节点标记覆盖的问题, 给出了两种改进标记方法, 特别是 ENPM 能够应用于较大规模的网络并取得较好的性能. 在实际应用中, 依据不同的网络拓扑可以选取不同的标记方法, 因此本方案有很强的实用性.

溯源追踪方案的目的是定位正在发送大量数据的节点. 但是该方案本身并不能完全解决无线传感器网络中发送虚假数据攻击问题. 例如该方案将攻击节点定位于某一很小的区域后, 仍需要其他方案的配合将攻击节点清除或者隔离; 当攻击节点正在发送大量虚假信息, 而恰好有合法节点也在发送较多合法信息时, 该方案可能同时定位攻击节点与合法节点, 在这种情况下, 也需要其他方案的配合, 区分攻击节点与合法节点. 如何建立一套完整的定位与清除攻击节点的机制将是下一步的研究重点.

参考文献:

[1] A Perrig, J Stankovic, D Wagner. Security in wireless sensor networks[J]. Communications of the ACM, 2004, 47(6): 53-57.

[2] F Ye, H Luo, S Lu, L Zhang. Statistical en-route filtering of injected false data in sensor networks[J]. IEEE Journal on Selected Areas in Communication, 2005, 23(4): 839-850.

[3] S Zhu, S Jajodia, P Ning. An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks[A]. In Proc. IEEE Symposium on Security and Privacy'04[C]. California: Los Alamitos, Calif, 2004. 259-271.

[4] L Garber. Denial-of-service attacks rip the internet[J]. Computer, 2000, 33(4): 12-17.

[5] H Burch, B cheswick. Tracing anonymous packets to their approximate source[A]. In Proc. USENIX Conference on System administration'00[C]. Berkley, California, USA: USENIX Association, 2000. 319-327.

[6] G Sager. Security fun with OCxmon and cflowd[OL]. http://www.caida.org/funding/ngi/content/security/1198/mt0000.htm

[7] S Savage, D Wetherall, A Karlin, T. Anderson. Practical network support for IP traceback[A]. In Proc. ACM SIGCOMM'00[C]. New York, USA: ACM NY, USA, 2000: 295-306.

[8] Advanced and authenticated marking schemes for IP traceback

[A]. In Proc. IEEE INFOCOM '01[C]. New Jersey: Piscataway, N.J., 2001. 878- 886.
[9] Z Yu, Y Guan. A dynamic scheme for err-route filtering false data[A]. In Proc. of ACM International conference on Embedded Networked Sensor Systems' 05[C]. New York, USA: ACM N.Y., USA, 2005. 294- 295.
[10] F Ye, H Yang, Z Liu. Catching "moles" in sensor networks [A]. In Proc. IEEE ICDCS' 07[C]. Washington, DC, USA: IEEE Computer Society, 2007. 69- 77.
[11] R Wang, W Du, P Ning. Containing denial-of-service attacks in broadcast authentication in sensor networks[A]. In Proc. ACM Mobihoc' 07[C]. New York, USA: ACM N.Y., USA, 2007. 71- 79.

作者简介:

杨 峰 男, 1982 年出生于山东烟台, 中国科学技术大学博士研究生, 主要研究方向为: 无线传感器网络.
E-mail: yfus@mail.ustc.edu.cn

周学海 男, 1966 年生于安徽淮南, 中国科学技术大学计算机系教授, 博士生导师, 主要研究方向为: 计算机体系结构, 嵌入式系统设计, 无线传感器网络等.
E-mail: xhzhou@ustc.edu.cn

张起元 男, 1983 年生于河南渑池, 中国科学技术大学博士研究生, 主要研究方向为: 无线传感器网络.

谢 婧 女, 1984 年生于江苏南京, 中国科学技术大学硕士研究生, 主要研究方向为: 嵌入式系统设计与验证, 无线传感器网络安全性等.

章曙光 男, 1970 年生于安徽淮南, 安徽建筑工业学院电子与信息工程学院副教授, 主要研究领域为: 无线传感器网络.