

传感器网络失效节点识别技术研究

赵涛 蔡皖东 李勇军

(西北工业大学计算机学院 西安 710072)

摘要 本文提出一种基于 Bloom Filter 的传感器网络失效节点的识别方法,该方法在不增加传感器网络节点负担的情况下,对传感器网络中无法正常工作的传感器节点进行识别。分析了算法中不同参数选取对于识别性能的影响,并通过仿真对算法准确性进行了验证。

关键词 无线传感器网络, Bloom Filter, 失效节点

Failure Node Identification in Wireless Sensor Network

ZHAO Tao CAI Wan-dong LI Yong-jun

(College of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract Due to the inherent stringent bandwidth and energy constraints, it is usually impractical to carry out periodic query to judge whether the node is failure in wireless sensor network. This paper provides a method of identifying failure nodes based on the Bloom Filter technique in the wireless sensor network. The method can effectively identify the failure sensor node in the wireless sensor network without incurring any additional overhead at sensor nodes. We analyze the performance of the identification algorithm influenced by selection of different parameters, and verify the correctness of the proposed method through simulation.

Keywords Wireless sensor network, Bloom filter, Failure node

传感器节点可以部署在不适于人类工作的环境中,通过采集数据(如温度、声音等)监测周围环境。与其他分布式监测系统最大不同之处在于,传感器网络具有自组织和能量带宽有限的特性。随着工作时间的增加,传感器节点的电池能量逐渐被耗尽,节点会因此而失效,从而退出网络。当网络中某区域失效节点较多时,就会影响传感器网络的正常运行和降低传感数据的健壮性和精确性。对此问题,有效解决方法就是可以快速准确地识别出失效的传感器节点,并对失效节点比率较大的网络区域抛洒新的传感器节点,因此传感器网络中失效节点的识别在传感器网络性能监测和管理方面显得非常重要。

国内外对传感器网络健康状况的性能监测和管理的研究越来越多, Jerry Zhao^[1]提出了传感器网络监测模型^[2]并对传感器节点剩余能量测量^[3]进行了研究, C. Hart^[4]和 Li^[5]对传感器网络内部链路的报文丢失性能进行基于端到端测量推测。由于能源和带宽有限的特性,传感器网络中的性能监测相对于互联网络、电信网络和电力网络中的性能监测更具有挑战性。对网络中的每个节点进行周期查询去判断节点是否失效的方法不符合传感器网络的能源带宽有限的特性。针对传感器网络的特点,本文提出一种基于 Bloom Filter^[6]的失效传感器节点识别方法,它通过在传感器网络传送报文中增加节点标识对失效节点进行有效的识别,并通过仿真对失效识别算法的正确性进行验证。

1 Bloom filter 技术

早在 1970 年, Bloom 提出一种基于多哈希函数映射来压缩参数空间、实现快速参数查找判定的 Filter 方法,并在计算机许多领域得到了广泛的应用。为了证明某一元素是否在集合中,可以用 Bloom filter 来简明地表示这个集合,为了表示一个具有 n 个元素的集合 $S = \{s_1, s_2, \dots, s_n\}$, 用一个长度为 m 的数组来描述 Bloom filter, 数组中的每一个元素初始化为 0。

一个 Bloom filter 使用 k 个相互独立的哈希函数 h_1, h_2, \dots, h_k , 它们的值域为 $\{0, 1, \dots, m-1\}$ 。为了运算方便, 假设关键字经过哈希函数得到一个随机的地址, 以便使一组关键字的哈希地址均匀分布在地址空间 $\{0, 1, \dots, m-1\}$ 。对于每一个元素 $s \in S$, 数组中对应于 $h_1(s), h_2(s), \dots, h_k(s)$ 的元素被置成 1。如图 1 所示, 假设使用 3 个相互独立的哈希函数, 即 $k=3$ 的元素被置成 1。

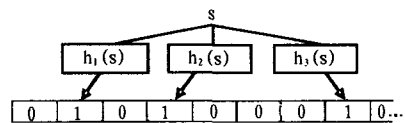


图 1 Bloom filter 原理示意图

这样, 描述一个元素 $s \in S$ 就可以用它的哈希函数值 $h_1(s), h_2(s), \dots, h_k(s)$ 在数组上对应的元素位置全部变成 1 来表示, 如果有某个存储单元为 0, 则表示这个元素不在集合 S 上。

2 传感器网络失效节点识别算法

本文主要针对进行周期性数据收集的传感器网络, 即传感器网络内所有节点周期性采集数据, 并将传感报文通过逐层汇聚的方式传递到传感器网络的汇聚节点。在传感器网络中, 假设每一个传感器节点都有唯一 ID 标识, 且汇聚节点知道传感器节点的物理位置, 每个传感器节点物理位置的定位可以通过文献^[7]中的方法去获取。

在不影响传感器网络正常工作的前提下, 失效传感器节点识别算法利用 Bloom filter 技术在每个传感器节点产生感应报文中加入节点 ID 信息, 并使传感器节点的 ID 信息在数据融合过程中不丢失, 在最终传送到汇聚节点的感应报文中, 仍然知道该报文是由哪些传感器节点采集的感应数据融合而成的。在汇聚节点对每轮收集到的节点 ID 信息进行

统计,对一定轮次连续没有感应报文发送到汇聚节点的传感器节点,根据其距离汇聚节点的距离(跳数),传感器网络内部链路的丢包率,对该传感器节点是否失效进行判定。本算法主要包括两部分,基于 Bloom filter 的节点 ID 信息表示和传送方法和节点失效判别算法。

2.1 基于 Bloom 的节点 ID 信息表示和传送

假设 Bloom Filter 的哈希空间长度为 m ,用位串 V 表示,传感器网络中共有 n 个传感器节点,数据集合相应为 $S = \{0, 1, \dots, n-1\}$,有 k 个相互独立且具有均匀分布特性的哈希函数 $h_1, h_2, \dots, h_k, s \in S, h_i(s) \in \{0, 1, \dots, m-1\}$,实现该传送技术主要分为 3 个步骤。

2.1.1 节点 ID 表示

在传感器数据产生节点利用集合表示算法将节点 ID 信息添加到传感数据的报文中。Bloom filter 集合表示算法如下:

```
BitString V_represent(BitString V, Var s){
  for(i=1; i<=k; i++){
    V[hi(s)]=1;
  }
  return V;
}
```

2.1.2 ID 信息融合

由于传感器网络在传送数据的过程中,会在融合节点进行数据融合以便节约能量,当两个或两个以上的传感报文到达融合节点的时候,在传感器网络的融合节点利用 Bloom Filter 的合并算法将两个或多个传感器节点的 ID 信息保留在经过数据融合过后的数据报文中。这样,虽然多个传感器报文融合成一个报文,但是这个报文是由哪些传感器节点采集的信息并没有丢失。Bloom filter 集合合并算法如下:

```
BitString V_union(BitString V1, BitString V2){
  BitString V[0..m-1]=0;
  for(i=0; i<m; i++){
    V[i]=V1[i]V2[i];
  }
  return V;
}
```

2.1.3 ID 信息统计查找

在传感器网络的汇聚节点,用查找算法对包含节点 ID 的报文进行处理,得到本轮的感应报文是由哪些节点采集的,并将统计的结果记录下来,用于后面的失效节点判断算法中。Bloom filter 查找算法如下:

```
Boolean V_lookup(BitString V, Var s){
  // 如果 s ∉ S, 返回 0; 如果 s ∈ S, 返回 1
  int i=1;
  while((V[hi(s)]=1) && (i<=k)) i++;
  if(i<=k) return 0; else return 1;
}
```

由上述算法可以看出,Bloom filter 技术仅仅在网络传输的感应报文中增加固定大小的数据位,主要处理都集中在能量充足的汇聚节点,不增加节点额外的负担,符合传感器网络带宽和能量有限的特性。

2.2 传感器节点失效判别方法

影响传感器报文失效丢失判断的有三个主要因素,节点的连续未发送数据次数,该节点距离汇聚节点的距离(跳数),传感器网络链路的报文丢失率。节点失效的概率与数据连续未发送次数成正比,与节点距离汇聚节点的距离和传感器网络链路的报文丢失率成反比。随着节点未发送次数的增加,该节点失效的可能性越来越大,但是由于存在该节点产生的报文连续丢失的情况,因此随着传感器网络链路丢失率和距离汇聚节点跳数的增加,节点产生的报文连续丢失的可能性也越来越大。假设 N 是连续节点未收集数据的轮次, p 是传感器网络链路报文丢失的概率(这里为了描述方便,假设传感器网络中的链路报文丢失率都为 p), r 是节点距离汇聚节点的条数。假设节

点 i 距离汇聚节点有 r 跳,则节点 i 一次报文发送失败的概率为 $1-(1-p)^r$,连续 N 次发送失败的概率为:

$$P(\text{false}) = (1 - (1 - p)^r)^N \quad (1)$$

由此设置一个阈值 e ,对于连续报文发送失败的节点来说,如果 $P(\text{false}) < e$,就做出节点失效的判定。这样就可以对传感器网络失效节点进行有效的判别。当传感器网络某个区域失效节点达到一定比例时,可以据此抛撒新传感器节点,保证传感器网络的正常工作和数据健壮性。下面通过仿真试验对算法进行验证。

3 仿真试验

首先分析选取合适的 Bloom filter 参数值,然后对算法的识别性能进行分析。

3.1 Bloom filter 参数选择

由于 Bloom filter 表示集合算法的随机特性,存在某元素不属于数据集合而被指称属于该数据集合的可能性,这时就出现了误判(false positive)。虽然出现误判,但是只要误判的概率足够小,该算法在实际中还是可以应用的。因此首先需要选择合适的 Bloom filter 参数,用 $Peer$ 表示误判的概率,根据文献[6]可得到公式:

$$p_{err} \approx (1 - e^{-\frac{km}{n}})^k \quad (2)$$

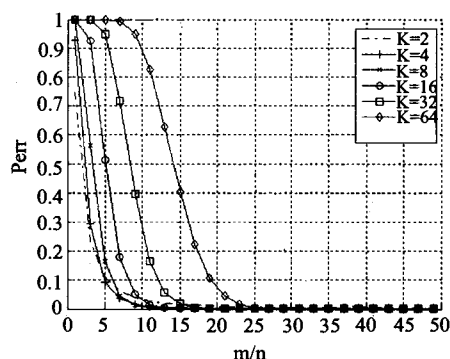


图2 Bloom filter 参数选择误差图

从图2中可以看出,Peer 值并不随着 k 的增加而成单调递减的趋势。如果想减小 Peer 值,需要增加 m/n ,由于 n 表示的是传感器网络内部的节点数,因此只有增加 m 的值,这样会增加网络中传输的数据量,从而缩短网络生存期。另外,考虑到传感器节点上的计算量与 k 值满足单调递增关系, k 值小可以减少节点上的计算量,在 $k=4, 8$ 时,图2中 Peer 值曲线比较接近,在相同 Peer 值的情况下,比 k 取 2 时需要的 m/n 要小。考虑传输数据量和节点上计算量两个方面,取 k 值为 4,从图中曲线上可以看出,当 m 和 n 的比值大于 10 时,再增加 m 和 n 的比值,Peer 值的变化并不显著,因此在算法实现时选择 $m/n=10$ 。

3.2 失效节点仿真试验

基于 3.1 节 Bloom filter 参数分析,利用 NS2^[8] 建立传感器网络仿真系统模型。传感器网络为 6×6 网格并且节点均匀分布,每个传感器节点之间相距 50m,传感器网络 MAC 协议是 CSMA/CA 协议,并使用 LAR 路由由协议进行数据传输。传感器网络拓扑中链路报文丢失率设置为 0.2,内部链路报文丢失率可由文献[4,5]中方法得到。如图3所示,汇聚节点在方格的左下处,随机设定一些传感器节点失效,用黑色节点表示,白色节点表示正常工作的传感器节点。

(下转第 105 页)

(2)路由前缀不聚合现象

虽然普遍情况下路由前缀按照从桩网络向接入点路由器的方向聚合,但有些情况下网管人员对某些路由器并没有设置前缀聚合功能,这样可能出现路由前缀短的路由器节点更接近桩网络,而路由前缀长的节点反而更接近根节点的情况。这种情况和前缀聚合树模型不吻合,但这种情况不具备普遍性。

结束语 非协作方式下 IPv6 接入网络拓扑发现的最大问题在于 IPv6 网络海量地址空间的特性使得探测效率和覆盖率的之间的矛盾不可调和。本文就此问题将 IPv6 接入网络按照其前缀聚合的单向性特征刻画成 PATM 模型,而后提出了基于 PATM 模型的网络拓扑发现方法。通过实际测试及分析,可以得出结论:本文提出的方法能够满足非协作方式下的 IPv6 网络拓扑发现总的要求,且对提高拓扑发现覆盖率有显著的帮助。

参考文献

[1] Internet Mapping Project. Available URL: http://research.lumeta.com/ches/db/
 [2] Astic I, Festor O. A hierarchical topology discovery service for IPv6 networks [C] //IEEE/IFIP Network Operations and Management Symposium (NOMS), Apr. 2002: 497-510
 [3] 李云琪,杨家海. 一个面向 IPv6 的网络拓扑管理系统的实现 [J]. 计算机工程与应用, 2004, 40(29): 73-75,181

[4] 沈曾伟,周刚. IPv6 接入网拓扑结构自动发现方法研究 [J]. 计算机工程, 2006, 19: 136-138
 [5] Waddington D G, Chang Fangzhe, Ramesh V, et al. Topology Discovery for Public IPv6 Networks [J]. ACM SIGCOMM Computer Communications Review, 2003, 33(3): 59-68
 [6] IPv6 Scamper. http://www.wand.net.nz/~mj12/ipv6-scamper/
 [7] Floyd S, Kohler E. Internet Research Needs Better Models. - ACM SIGCOMM Computer Communication Review (CCR), 2003, 33 (1): 29-34
 [8] 张宇,张宏莉,方滨兴. Internet 拓扑建模综述. 软件学报, 2004, 15(8): 1220-1226
 [9] Alderson D, Doyle J, Govindan R, et al. Toward an Optimization-Dreven Framework for Designing and Generating Realistic Internet Topologies. ACM SIGCOMM CCR, 2003, 33(1): 41-46
 [10] Waxman B M. Routing of Multipoint Connections. IEEE Journal on Selected Areas in Communications (JSAC), 1988, 6 (9): 1617-1622
 [11] Doar M B. A Better Model for Generating Test Networks// Proc. 1996 IEEE Global Telecommunications Conf. (GLOBECOM 1996). London, Nov. 1996: 86-93
 [12] Calvert K L, Doar M B, Zegura E W. Modeling Internet topology. IEEE Communications, 1997, 35 (6): 160-163
 [13] Media A, Lakhina A, Matta I. An approach to universal topology generation//Proceedings of MASCOTS. Washington, 2001: 346-353
 [14] Winick J, Jamin S. Inet-3.0: Internet topology generator. Technical Report CSE-TR-456-02. Department of EECS, University of Michigan, 2002
 [15] Zhou S, Mondragon R J. The rich-club phenomenon in the Internet topology. IEEE Communication Letters, 2004, 8 (3): 180-182

(上接第 65 页)

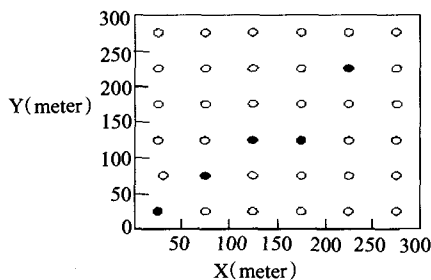


图 3 预设传感器网络节点分布图

参照 2.2 节的失效节点识别算法,设置 $e=0.1$,即可根据公式(1)对失效节点进行识别,识别结果如图 4。

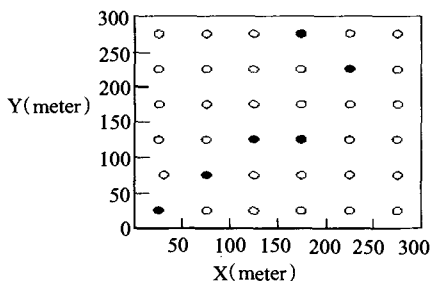


图 4 $e=0.01$ 失效节点识别示意

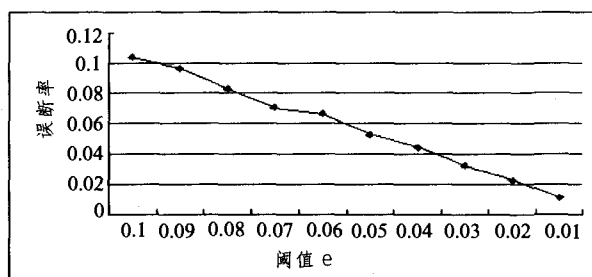


图 5 误断率与阈值 e 关系图

从图中可以看出,随机设置的失效传感器节点可以被准

确地识别出来,但是有一个距离传感器汇聚节点距离较远的节点被误判为失效节点。由于该节点距离传感器汇聚节点较远,数据发送到汇聚节点可能需要经过更多的跳数,在设置的阈值 e 的范围内,产生了报文的连续丢失,因此需要提高阈值的精度去做出更准确的判断,将阈值 e 设置为 0.01,就可以对失效的传感器节点做出准确的判断而不会产生误判。

下面对误判率和阈值 e 的关系进行分析,进行 10000 次数据收集,并对距离汇聚节点跳数为 10 的节点,研究其随着阈值 e 的变化,节点失效的误断率变化情况,从图 4 中可以看出当 e 取值在 0.01 时,误断率可以控制在 1% 左右,对其他节点进行同样的抽样,也得出类似的结果,由此可以看出,节点失效误判的情况可以由阈值 e 很好地控制。

结束语 本文利用 Bloom filter 技术,提出了一种传感器网络失效节点的识别方法,并通过 NS2 仿真模型进行了验证。但是本文提出的算法是针对周期性数据采集的传感器网络,具有一定的局限性,如何在事件驱动和非周期性传感数据收集的传感器网络利用传感器节点采集数据的时空相关性进行失效节点的判断是未来研究的目标。

参考文献

[1] Zhao J, Ramesh G, Deborah E. Sensor network tomography: Monitoring wireless sensor networks. Computer Communication Review, 2002, 32(1): 64
 [2] Zhao J. Measurement and Monitoring In Wireless Sensor Networks. Ph. D. Thesis. Department of Computer Science, University of Southern California, Dec. 2003
 [3] Zhao J, Govindan R, Estrin D. Residual energy scan for monitoring sensor networks// IEEE Wireless Communications and Networking Conference, vol. 1, IEEE USA, 2002: 356-362
 [4] Hartl G, Li Baochun. Loss inference in wireless sensor networks based on data aggregation//Third International Symposium on Information Processing in Sensor Networks, ACM USA, 2004: 396-404
 [5] Li Yongjun, Cai Wandong, Tian Guangli, et al. Loss Tomography in Wireless Sensor Network Using Gibbs Sampling// Proceeding of EWSN 2007. LNCS 4373, 2007: 150-162
 [6] 肖明忠, 代亚非. Bloom Filter 及其应用综述. 计算机科学, 2004, 31(4): 180-183
 [7] He T, Huang C, Blum B M, et al. Range-free localization schemes for large scale sensor networks// Proceedings of the 9th Annual International Conference on Mobile Computing and Networking. 2003: 81-95
 [8] The Network Simulator 2, www.isi.edu/nsnam/ns2, 2005