

基于单源分段编码的无线对等感知网数据收集模型

尹子铭¹, 袁松²

(1. 徐州市第一人民医院, 江苏 徐州 221000; 2. 杭州电子科技大学 电子信息学院, 杭州 310018)

摘要: 针对触发型数据场景下大规模无线对等感知网络的部分数据收集问题, 提出一种基于单源分段编码的数据收集模型; 利用单个源节点的多个分段之间编码而成的数据来记录源数据, 提高数据可靠性的同时使其适应大规模网络下的部分数据收集; 通过使用游走包进行编码操作, 避免源节点过多的能量消耗; 同时针对节点存储空间提出动态划分编码单元策略, 利用邻居间的信息交换动态调整源数据切分的编码单元大小, 实现节点存储空间与收集效率的动态调整; 并针对灾难场景下的数据收集提出了危险感知编码冗余量动态调整策略, 通过对邻居状态的感知动态调整发送随机游走包的个数, 自适应地提升编码冗余量, 提高数据恢复率。

关键词: 大规模无线对等感知网络; 动态自适应编码; 触发型数据; 部分数据收集

Data Collection Model of Wireless Peer-to-Peer Sensing Network based on Single Source Segmented Coding

YIN Ziming¹, YUAN Song²

(1. Xuzhou No. 1 Peoples' s Hospital, Xuzhou 221000, China;

2. Information Engineering College, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: Aiming at the problem of partial data collection in large-scale wireless peer-to-peer awareness networks in trigger data scenario, a data collection model based on single source segmented coding model (SSSCM) is proposed. The data encoded between multiple segments of a single source node is used to record the source data, which improves the reliability of the data and makes it suitable for partial data collection under large-scale networks. The random walk coding packets are used to perform the encoding operations to avoid the excessive energy consumption of the source node. At the same time, the dynamic segmentation of coding units (DSCU) is proposed, using the information exchange to dynamically adjust the size of the coding unit, and achieve the dynamic balance between node storage space and collection efficiency. In addition, for data collection in disaster scenarios, a dynamic adjustment of disaster sensing coding redundancy (DADSCR) is proposed, which dynamically adjusts the number of random walk packets sent through the perception of neighbor status, adaptively increases redundancy, and further improve data recovery rate.

Keywords: large-scale wireless peer-to-peer sensing network; dynamic adaptive coding; trigger data; partial data collection

0 引言

无线对等感知网络在灾难特殊场景下的容错数据收集问题按照有无汇聚节点 (Sink), 通常被分为实时数据收集问题和延时数据收集问题两类^[1-2]。其中, 实时数据收集是指汇聚节点的收集与传感数据的产生同时发生, 即传感数据产生的同时, 汇聚节点则实时进行数据收集并译码。该场景面临的主要问题是确保汇聚节点在任意时刻的译码效率, 同时使网络中尚未被收集与译码的数据保持一定的持久性。一般典型的实时收集感知网络主要包括三大部分组成: 控制中心、汇聚节点和传感器节点, 如图 1 所示。

无线对等感知网络的实时数据收集问题已经被广泛研究^[3-4]。实时数据收集问题主要用于实时监测地震、火灾、

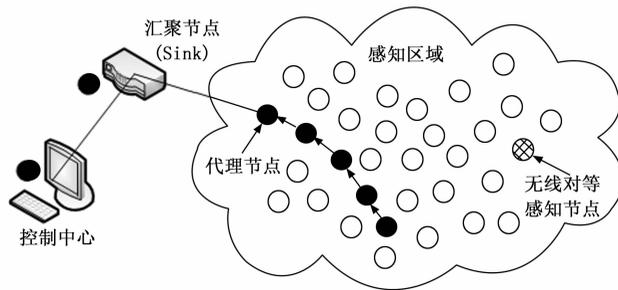


图 1 实时数据收集模型

山体滑坡等自然灾害场景中, 这些场景需要传感器节点将所感知的数据及时发送到汇聚节点。文献 [5] 最早对无配

收稿日期: 2023-02-14; 修回日期: 2023-03-19。

基金项目: 浙江省重点研发计划项目 (2020A01009)。

作者简介: 尹子铭 (1987-), 男, 大学本科, 助理工程师。

引用格式: 尹子铭, 袁松. 基于单源分段编码的无线对等感知网数据收集模型[J]. 计算机测量与控制, 2023, 31(12): 195-202.

置灾难场景收集的数据持久性进行系统分析,提出了一种新颖的 Growth Codes 编码技术来提高网络的数据持久性,但是该方案未有效地解决冗余数据的重复发送问题,使得网络在数据回收后期的整体数据采集速度快速降低。文献 [6] 通过对 Growth Codes 方案进行一系列改进,使得基于 Growth Codes 的数据收集方案能适用于静态场景和动态场景的数据收集,而且表现出了比较好的数据收集性能,但是该方案的改进策略缺乏理论性和创新性。文献 [7] 发现随机网络编码的性能在边缘监测网络中优于 Growth Codes 编码方案,然而这种集中控制的方式在很多场景中并不适用。文献 [8] 针对零配置网络场景中多拍数据的持久性和收集效率问题,提出将 Growth Codes 单拍数据收集问题一般化,以解决多拍数据的持久性和收集问题,但是未能获得理想的效果。文献 [9] 同样针对多拍数据的持久性和收集问题,首先提出了数据的效用问题,然后设计了交叉编码和不交叉编码两类策略,以实现在多拍数据收集中的数据效用的最大化。文献 [10] 针对 Growth Codes 的度分布等问题进行了探讨,设计了特殊的数据分发算法。

延时数据收集模型如图 2 所示,整个网络由大量被散布在感知区域中的无线对等感知节点和一个在收集器组成。收集器在感知节点被散布到感知区域中一段时间后,然后从外部进入感知区域内将感知节点该段时间所存储的编码数据收集起来,接着再在收集器中将收集到的编码包解码恢复出原始数据。

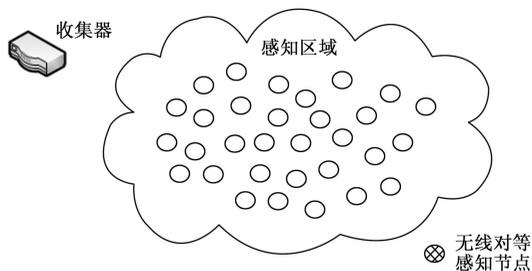


图 2 延时数据收集模型

在延时数据收集的场景中,网络被视为一个分布式存储系统,各感知节点间的感知数据采用编码的方式被存储在整个网络中。LT 码^[11]是第一个无比率擦除的数字喷泉码,被用于感知数据的编码,通过利用无比率的前向纠错来达到提高点对点通信吞吐量的目的。文献 [12] 提出两种基于 LT 码的分布式存储方案,但是由于这两种方案主要是利用节点之间的数据进行编码,所以无法实现部分区域数据的收集。文献 [13] 在喷泉码的基础上提出了两种分布式的编码方法来提高大型传感器网络中的数据持久性问题,但是这两种方案需要已知网络中节点数和源节点个数,而且还需要计算节点转发的概率转移矩阵,计算复杂度较大。文献 [14] 在分布式网络中提出了基于地理路由的分布式喷泉码,该方案的缺点是其基于地理路由以及基于查询的方法不适用于灾难场景的数据编码。文献 [15] 利用

最优局部可恢复码构造了分散擦除码,并利用构造码的生成矩阵性质实现了 Blom 密钥预分配协议。文献 [16] 提出了一种基于局部性码的分散冗余生成方案 D2CP。文献 [17] 引入了一种新的高效能源数据传播方案,称为数据节能生存能力 (DSwEE),将 DSwEE 与数据存储中的两种主要协议进行比较,即数据生存的分散 erasure code (DEC-DS) 和分散 erasure code 编码一分发 (DEC-EaD)。文献 [18] 提出了两层无线对等感知网络体系结构,从网络体系结构、通信协议到传感器单元体系结构,阐述了各个层次的节能策略。文献 [19] 针对故障区域为矩形或圆形的异构网络,利用节点故障模式的空间属性设计了专用的擦除码策略,可以提高数据可持久性并降低通信代价,虽然该方案效率非常高,但是其只适用于空间形状规则的场景。文献 [20] 改进了适用于无线多跳网络的 LT 码,该方案基于置信传播算法,虽然很好地降低了解码算法的复杂度,但是也会带来较大的额外通信和收敛时间开销。文献 [21] 提出利用单个节点周期性采集的多代数据之间进行数据编码的方式,将编码数据的相关性绑定在单个节点的多代数据之间,从而提供了部分区域数据采集的能力,但是该方案要求感知节点要周期性地采集数据,并不适用于触发型的数据采集方式。

普通的集中式 sink 收集存储模型容易产生单点故障,而且容易吸引攻击。集中式存储也可能导致性能瓶颈,因为所有数据收集和访问都必须在汇聚节点完成,因此关注点主要在延时数据收集问题上。尽管以上延时数据收集方式在各自的应用场景中取得了一定的效果,但是它们仍然具有各自的局限性。大部分方案的研究点是基于源节点周期性数据收集的模式,大规模网络下触发式数据采集方式的研究还有待探索,而且大部分传统数据收集方法若要收集部分源节点中的数据则需要解码出所有源节点中的数据,在大规模网络中此种劣势尤为突出,因此需要能以较低成本恢复出网络中部分节点中的数据的方案。

为此,提出了一种可适用于灾难场景下触发型数据在大规模无线对等感知网络环境下实现部分数据收集的模型,即单源分段编码的数据收集模型 (SSSCM, single source segmented coding model)。所给方案首先将单个源节点的数据进行分段编码来进行记录,并通过发送游走包增加源数据编码冗余量来提高数据可靠性,以使其适应大规模网络下的部分数据收集;然后针对节点存储空间提出动态划分编码单元策略,通过利用邻居间的信息交换动态调整源数据切分的编码单元大小,实现节点存储空间与收集效率的动态调整;最后针对灾难场景下的数据收集提出危险感知编码冗余量动态调整策略,通过感知邻居状态动态调整发送随机游走包的个数,以自适应地优化编码冗余量,从而有效提高数据恢复率。仿真结果表明,所给 SSSCM 数据收集方案能可靠地提升数据恢复率,且能有效降低能量消耗,可以很好地适用于灾难场景下的大规模对等感知网络中触发型数据的部分收集。

1 单源分段编码模型 (SSSCM)

所提单源分段编码模型 (SSSCM) 的基本思路如下: 第一步是给出 SSSCM 收集模型的总体设计, 将采集到的单个源节点的数据按编码单元进行分段, 然后采用 LT 度分布对分段的源数据进行异或编码, 这需要引入编码冗余量来恢复源数据, 最后计算处期望发送到网络中的编码包数量, 其中编码单元大小和编码冗余量是影响期望发送到网络中的编码包数量的两个因素; 第二步是针对 SSSCM 模型的编码单元大小优化问题, 通过计算直到采集结束周边节点所需要存储数据量以及当前网络中所存储的编码数据量之间的关系, 给出一种自适应的编码单元计算方法; 第三步是针对 SSSCM 模型的编码冗余量优化问题, 通过判断对节点邻居存活状态感知的方式, 即根据邻居总的减少量、邻居节点在最近一段时间跨度内邻居数量的减少趋势以及当前邻居的个数, 计算出感知周边环境危险度参数, 从而获得最优的编码冗余量。

1.1 所提数据收集模型总体设计

当某单个源节点 N_i 在 t 时刻被触发了采集数据的信号后, 将采集大小为 $x_i^{(t)}$ 的源数据。随后节点根据当前收集到的网络状态信息计算出合适的编码单元 $c_i^{(t)}$, 并依此来计算出切分后的分段个数 $k_i^{(t)}$, 如式 (1) 所示:

$$k_i^{(t)} = \text{ceil}\left(\frac{x_i^{(t)}}{c_i^{(t)}}\right) \quad (1)$$

式中, $\text{ceil}()$ 表示取整函数。

为了提高网络中的数据可靠性, 将 $k_i^{(t)}$ 段数据按照 LT 度分布进行异或编码, 将这些编码包尽量分布到不同的节点中去。用 LT 度分布进行数据编码需要一定的编码冗余度 ϵ 作为保证来恢复出所有的原始数据, 增加一定编码冗余量后总的编码包个数为 $k_i^{(t)}(1 + \epsilon)$ 。

然而, 面对恶劣环境有可能造成的节点失效, 需要提高数据的编码冗余量来增强数据的可靠性, 从而提高数据恢复率。通过节点对周围环境危险态势的感知来动态的调整编码冗余量 $\epsilon_i^{(t)}$ 。首先, 通过设定的编码冗余量的最小值 ϵ_{\min} 来计算出至少需要发送的游走包的个数 P_{\min} , 如式 (2) 所示:

$$P_{\min} = (1 + \epsilon_{\min})k_i^{(t)} \quad (2)$$

随后, 依次发送 P_{\min} 个随机游走包, 在发送完成这 P_{\min} 个游走包的时候, 由于感知区域中环境的变化, 其中部分节点可能因为各种灾难原因而失效, 并且可能继续扩散, 所以此时通过计算出动态冗余量 $\epsilon_i^{(t)}$ 来调整最大所需要发送的随机游走包的个数。此时期望发送到网络中的编码包的数量 P , 如式 (3) 所示:

$$P = (1 + \epsilon_i^{(t)})k_i^{(t)} \quad (3)$$

如果发送的游走包数量没有达到期望的数量, 则继续发送游走包, 直到发送的游走包的数量为 P 。

总之, 首先由源节点发送出一个源数据广播包, 将源数据以及编码单元大小广播到节点周边的一定区域内, 而后单播发送至少需要发送的 P_{\min} 个随机游走包进行编码存

储, 随后通过计算动态冗余量 $\epsilon_i^{(t)}$ 来决定是否需要追加更多的随机游走包来补偿在过去一段时间可能被灾难摧毁的节点以及将来可能会失效的节点中的编码数据。其他节点在收到对应数据包后将会进行一系列判断与处理, 对源数据进行编码存储, 从而实现一次完整的数据收集与编码存储。

1.2 动态划分编码单元 (DSCU)

1.2.1 编码单元大小的影响

编码单元大小 c 是指每个需要存储该批次源数据的编码包的大小, 其将会影响每个存储该编码包的节点所需要占用的存储空间。由于网络中节点剩余编码存储空间的限制, 若编码单元 c 过大, 则节点的剩余空间无法满足其编码数据的存储要求, 从而导致数据无法被妥善存储在网络中。然而, 若编码单元 c 过小, 其分段数 k 将增大, 为了从编码数据中恢复出源数据, 一般需要 $k(1 + \epsilon)$ 个编码包, 其中编码冗余量 $\epsilon > 0$ 。而编码包的个数由随机游走包的发送个数控制, 因此分段增大将导致所需要发送的随机游走包的个数提升。由于节点可能因为各种原因而失效, 所以编码包尽量分布在不同的节点中来提高数据的可靠性, 游走包需要在网络中游走一定的次数来找到合适的节点进行数据的编码存储。进而, 更多的随机游走包需要尽量的分布在不同的节点中, 这势必造成数据收集效率的降低。

1.2.2 编码单元大小的计算

由于上述编码单元大小带来的影响, 希望提出一种自适应的编码单元计算方法, 在保证周边节点有足够存储空间的前提下提高数据收集效率。为了计算出当前期望的编码度, 希望通过计算直到采集结束周边节点所需要存储数据量以及当前网络中所存储的编码数据量之间的关系来实时计算。

首先, 尝试计算直到采集结束单个节点剩余所需的存储空间的期望, 假设期望的数据采集周期在时间范围 $t \in [t_0, T]$ 内, 这表示至少希望网络中的节点在该时间范围内能正常工作, 在此期间收集器可以随时进入特定区域进行数据收集。由于节点需要事先知道收集器将在何时到来, 因此该网络属于非零配置网络。设源节点为 N_i , 其在时刻 t 采集到源数据 S 。则可以根据网络中已使用的编码存储区的空间增速以及直到收集器到来的剩余时间 T_{rest} 来计算出直到收集周期结束还需要存放的编码数据大小的期望 E_c 。假定源节点 N_i 在 t 时刻的邻居个数是 $n_i^{(t)}$, 由于能从 KeepAlive 包中得知各个邻居节点的空间增长速度 $v_1, v_2, \dots, v_{n_i^{(t)}}$, 则可求出邻居存储空间平均增速 $v_i^{(t)}$, 如式 (4) 所示:

$$v_i^{(t)} = \frac{1}{n_i^{(t)}} \sum_{j=1}^{n_i^{(t)}} v_j \quad (4)$$

然后可以计算出感知网络直到收集周期结束剩余所需采集的时间 T_{rest} , 如式 (5) 所示:

$$T_{\text{rest}} = T - t \quad (5)$$

在感知区域中触发采集的数据量的增长较为稳定的情况下, 能通过以上数据计算出直到 T 时刻单个节点剩余所需的存储空间大小的期望 E_c , 如式 (6) 所示:

$$E_c = v_i^{(t)} T_{rest} \quad (6)$$

下面需要知道周边节点当前存储空间的使用情况，同样能利用邻居发来的 KeepAlive 数据包中传递的信息。数据包中包含邻居节点 N_j 编码存储区的剩余存储空间 C_j 以及邻居节点 N_j 的所有邻居的剩余存储空间的均值 M_j 。由于单单知道邻居的剩余存储空间对周围网络空间中节点的剩余存储空间的了解太过片面，希望多一层对邻居的邻居的剩余存储空间信息的掌握，从而拓宽对网络空间的剩余存储空间的了解。

首先，计算源节点邻居的存储空间的均值 $M_i^{(t)}$ ，如式 (7) 所示：

$$M_i^{(t)} = \frac{1}{n_i^{(t)}} \sum_{j=1}^{n_i^{(t)}} C_j \quad (7)$$

随后计算所有邻居的邻居们的剩余存储空间的均值 $\hat{M}_i^{(t)}$ ，如式 (8) 所示：

$$\hat{M}_i^{(t)} = \frac{1}{n_i^{(t)}} \sum_{j=1}^{n_i^{(t)}} M_j \quad (8)$$

至此，得到了源节点邻居的剩余存储空间的均值 $M_i^{(t)}$ 和其邻居的邻居们的剩余存储空间的均值 $\hat{M}_i^{(t)}$ 。下面判断使用哪个值作为源节点周围网络空间中剩余存储空间的值。引入一个权衡系数 $\lambda (0 < \lambda < 1)$ ，以此来控制这两个变量的影响权重。计算出源节点 N_i 周围网络空间中的剩余存储空间的均值 M^* ，如式 (9) 所示：

$$M^* = \lambda M_i + (1 - \lambda) \hat{M}_i \quad (9)$$

可以通过 M^* 和 E_c 的差值来计算出当前合适的编码单元大小 U ，如式 (10) 所示：

$$U = M^* - E_c = \lambda M_i + (1 - \lambda) \hat{M}_i - v_i^{(t)} T_{rest} = \frac{\lambda}{n_i^{(t)}} \sum_{j=1}^{n_i^{(t)}} C_j + (1 - \lambda) \frac{1}{n_i^{(t)}} \sum_{j=1}^{n_i^{(t)}} M_j - \frac{1}{n_i^{(t)}} \sum_{j=1}^{n_i^{(t)}} v_j (T - t) \quad (10)$$

假设给定一个最小编码单元大小的范围 $c_i^{(t)} \in [K_{\min}, K_{\max}]$ ，由于计算出的 U 可能超出所允许的范围，需要将 U 限定在该范围内，所以最终的编码单元 $c_i^{(t)}$ 的计算方式，如式 (11) 所示：

$$c_i^{(t)} = \begin{cases} K_{\min}, & U \leq K_{\min} \\ U & K_{\min} < U < K_{\max} \\ K_{\max} & U \geq K_{\max} \end{cases} \quad (11)$$

1.3 危险感知编码冗余量动态调整 (DADSCR)

在灾难场景下的大规模无线对等感知网络中，虽然能够通过某些特殊的传感器感知到危险信号并发送到网络中，但是由于感知区域中灾难场景的多样性，可能无法有效捕捉危险信号。所以提出一种利用对节点邻居存活状态感知来感知周边环境危险程度的方式，进而对数据收集策略中的数据冗余方式进行调整，从而提高数据的恢复率。

首先，通过对 t 时刻源节点 N_i 周围邻居存活状态的感知来衡量出该节点周围环境的危险系数 $R_i^{(t)}$ 。源节点 N_i 的邻居个数减少量意味着节点周边环境危险程度提升，综合考虑邻居总的减少量 $\Delta_i^{(t)}$ 、邻居节点在最近一段时间跨度 $r (r \in Z^+)$ 内邻居数量的减少趋势 $\tau_i^{(t)}$ 以及当前邻居的个数 $n_i^{(t)}$

来计算其危险程度。记在各个时间节点 N_i 邻居的个数为 $n_i^{(0)}, n_i^{(1)}, \dots, n_i^{(t)}$ ，节点 N_i 直到当前时间 t 总的邻居减少量为 $\Delta_i^{(t)} = n_i^{(0)} - n_i^{(t)}$ ，定义邻居节点在最近一段时间跨度 r 内邻居数量的减少趋势为 $\tau_i^{(t)} = \delta_i^{(t)} e^{p_i^{(t)}}$ ，其中 $\delta_i^{(t)}$ 为 r 时间跨度内邻居的减少量，且有 $\delta_i^{(t)} = n_i^{(t-r)} - n_i^{(t)}$ ， $p_i^{(t)}$ 表示 t 时刻节点 N_i 邻居在时间跨度 r 内的减少速度的变化量，从当前时刻 t 往后分别取 3 个时间跨度为 r 的邻居个数 $n_i^{(t-2r)}, n_i^{(t-r)}, n_i^{(t)}$ ，则有式 (12)：

$$p_i^{(t)} = \delta_i^{(t)} - \delta_i^{(t-r)} = (n_i^{(t-r)} - n_i^{(t)}) - (n_i^{(t-2r)} - n_i^{(t-r)}) = 2n_i^{(t-r)} - n_i^{(t-2r)} - n_i^{(t)} \quad (12)$$

根据邻居减少速率的变化量来对时间跨度 r 内的邻居减少量进行缩放来衡量时间跨度 r 内邻居数量的减少趋势 $\tau_i^{(t)}$ 。当 $p_i^{(t)} < 0$ 时，邻居的减少速度递减，此时 $\tau_i^{(t)} < \delta_i^{(t)}$ ；当 $p_i^{(t)} = 0$ 时，邻居的减少速度不变，此时 $\tau_i^{(t)} = \delta_i^{(t)}$ ；当 $p_i^{(t)} > 0$ 时，邻居的减少速度递增，此时 $\tau_i^{(t)} > \delta_i^{(t)}$ 。 $\tau_i^{(t)}$ 与 $\Delta_i^{(t)}$ 反应出邻居减少量相关数据，与危险程度成正相关；当前的邻居个数 $n_i^{(t)}$ 越少，其邻居减少所带来的负面影响会相较于有更多邻居的情况要大，因此 $n_i^{(t)}$ 与危险程度的关系应为负相关。再考虑到节点邻居个数可能为 0 的极端情况，尝试用以下表达式来描述周围环境的危险程度 $R_i^{(t)}$ ，如式 (13) 所示：

$$R_i^{(t)} = \begin{cases} +\infty & n_i^{(t)} = 0 \\ \frac{\tau_i^{(t)} + \Delta_i^{(t)}}{n_i^{(t)}} & n_i^{(t)} \neq 0 \end{cases} \quad (13)$$

考虑在当前邻居个数 $n_i^{(t)} \neq 0$ 的情况下，若总的邻居减少量 $\Delta_i^{(t)} = 0$ ，则 $\tau_i^{(t)}$ 也必然为 0，从而危险系数 $R_i^{(t)} = 0$ ，此时的危险系数取最小值；如若总的邻居减少量 $\Delta_i^{(t)} > 0$ 且时间 r 跨度范围内邻居的减少量激增将导致 $\tau_i^{(t)} \gg \delta_i^{(t)}$ ，如恰巧此时 $n_i^{(t)} \rightarrow 0$ ，则 $R_i^{(t)} \rightarrow +\infty$ 。从而可以得知危险系数 $R_i^{(t)}$ 的取值范围在区间 $[0, +\infty)$ 之间，然后根据该危险系数计算出所需要的编码冗余量。给定允许的编码冗余量的范围为 $\epsilon_i^{(t)} \in [\epsilon_{\min}, \epsilon_{\max}]$ ，当 $n_i^{(t)} \neq 0$ 时， $0 \leq R_i^{(t)} < +\infty$ ，希望当 $R_i^{(t)}$ 为 0 的时候， $\epsilon_i^{(t)}$ 取 ϵ_{\min} ；当 $R_i^{(t)} \rightarrow +\infty$ 时， $\epsilon_i^{(t)}$ 取 ϵ_{\max} 。仅考虑节点的邻居仍然存在的情况，即 $n_i^{(t)} \neq 0$ 时的情况，当节点的邻居都已不存在时，其并不能找到发送随机游走包的对象。由此，最终的编码冗余量 $\epsilon_i^{(t)}$ 的计算如式 (14) 所示：

$$\begin{aligned} \epsilon_i^{(t)} &= \epsilon_{\max} - \frac{1}{R_i^{(t)} + \frac{1}{\epsilon_{\max} - \epsilon_{\min}}} = \\ &= \epsilon_{\max} - \frac{1}{\frac{\tau_i^{(t)} + \Delta_i^{(t)}}{n_i^{(t)}} + \frac{1}{\epsilon_{\max} - \epsilon_{\min}}} = \\ &= \epsilon_{\max} - \frac{1}{\frac{\delta_i^{(t)} e^{p_i^{(t)}} + n_i^{(t)} - n_i^{(t)}}{n_i^{(t)}} + \frac{1}{\epsilon_{\max} - \epsilon_{\min}}} = \\ &= \epsilon_{\max} - \frac{1}{\frac{(n_i^{(t-r)} - n_i^{(t)}) e^{2n_i^{(t-r)} - n_i^{(t-2r)} - n_i^{(t)}}{n_i^{(t)}} + n_i^{(t)} - n_i^{(t)}}{n_i^{(t)}} + \frac{1}{\epsilon_{\max} - \epsilon_{\min}}} \end{aligned} \quad (14)$$

通过该冗余量便能够计算出最终所需要分布在网络中的编码包的个数, 也就是所需要发送的随机游走包的个数。

2 实验结果与性能评估

2.1 实验方案

考虑到面对的网络场景为大规模的无线对等感知网路, 其真实实验成本相对高昂且实验灵活性较低, 所以采用 Matlab 软件建立仿真平台的方式来进行实验方案的近似模拟, 具体方法是通过在 Matlab 仿真软件的 Simulink 界面上根据网络模型及参数设置建立仿真平台, 模拟仿真后通过示波器输出仿真结果。仿真环境设定如下: 实验覆盖的感知区域大小为 $2\ 800\ \text{m} \times 2\ 800\ \text{m}$, 感知区域内随机散布 8 000 个节点, 所有感知节点均相同且具有全局唯一标识和自定位功能, 这些节点自组织生成采集网络, 感知区域外配置 1 个收集器, 收集数据时与感知节点采用 ZigBee 方式进行通信, 网络将从 0 时刻一直运行到 T 时刻。仿真实验的详细参数如表 1 所示。

所给 SSSCM 方案设计为部分收集策略, 虽然可以在 $0 \sim T$ 时刻的任意一个时刻进行部分的数据收集, 但为了产生更多的数据来减少意外的偏差, 选择在网络运行到 T 时刻时进行数据的部分收集工作。在 T 时刻, 选取网络中任意一块 $1\ 400\ \text{m} \times 1\ 400\ \text{m}$ 的区域来进行数据采集, 随后进行解码以及实验信息的统计工作。

表 1 仿真实验参数表

参数	值
感知区域规模	$2\ 800\ \text{m} \times 2\ 800\ \text{m}$
节点个数	8 000 个
采集周期 T	2 000 s
节点通信半径	70 m
源节点采集数据概率	0.000 1
节点单次采集数据量	20 000 字节
源数据暂存区大小	3 000 000 字节
编码包存放区大小	5 000 字节
源数据最大广播跳数	10 跳
随机游走最大步长	300 步
编码单元最小值 K_{\min}	66 字节
编码单元最大值 K_{\max}	130 字节
编码冗余量最小值 ϵ_{\min}	0.5
编码冗余量最大值 ϵ_{\max}	0.8

2.2 不同编码单元的比较

为了验证 DSCU 方案在实际中的有效性, 对不同的编码单元大小进行了多组实验来验证其性能。对于编码单元大小, 除了提出的 DSCU 方法外对比了其取一些固定值的方案。这些固定值分别为编码单元范围内的最小值 K_{\min} , 中间值 $\frac{(K_{\min} + K_{\max})}{2}$, 以及最大值 K_{\max} , 为了便于描述将其分别记为 MIN, MID, MAX。

2.2.1 数据恢复率

数据恢复率定义为从当前网络中解码出的数据量与当

前网络中产生的原始数据分段个数的比。分别对 100 次不同编码单元类型的实验最终数据的恢复率求均值, 得到了不同编码单元方案下数据恢复率的对比关系, 如图 3 所示。

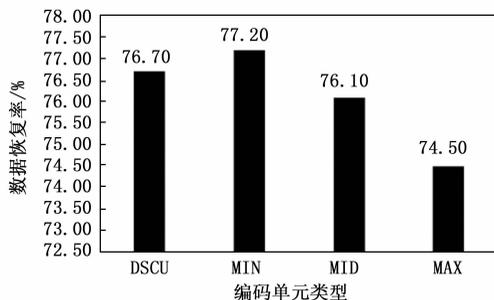


图 3 不同编码单元方案数据恢复率对比图

由于选择在 T 时刻时进行数据收集, 此时网络中仍然可能有许多处于游走状态的数据包, 许多的数据包还未编码完成, 所以最高的数据的恢复率并没有接近 100%。整体来说, 考虑仍然在网络中的未完成编码的包, 数据恢复率仍然处于一个较高的水平。不难看出, 数据的恢复率是按照 MIN、DSCU、MID、MAX 方案的顺序递减的, 所给 DSCU 方案较之 MIN 略低一些, 但是高于 MID 以及 MAX 方案。

不同编码单元类型在 100 次实验中数据恢复率的散点如图 4 所示。从图 4 可以看出, 分布在较高恢复率处的编码单元类型基本为所提出的 DSCU 方案, 但该方案与其它方案相比分布跨度稍大。而 MIN 方案相较于其他方案其分布相对较紧凑, 且 MIN 方案的下界与其它方案相比更高, 恢复率的稳定性较好。分布率的下界从低到高依次为 MAX、MID、DSCU、MIN, 下界最低的方案为 MAX。

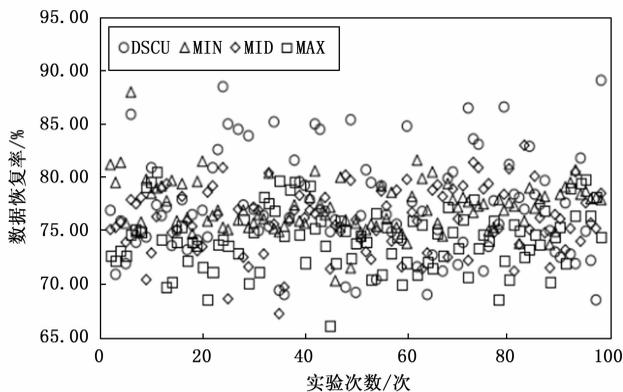


图 4 不同编码单元类型 100 次实验恢复率分布图

下面为验证 DSCU 方案的数据恢复率性能优劣, 将其与传统 SSPLBP^[22] 和 LTCDS^[23] 两种方案进行比较, 如图 5 所示。当数据采集率为 1 时, LTCDS 方案能将全部源数据恢复, 但是数据采集率越低, LTCDS 方案的数据恢复率就越低, 这是由于 LTCDS 方案没有足够的数据冗余量致使无法较好地恢复数据。而 SSPLBP 方案在低触发率情况下具有较高数据恢复率, 但是当触发率超过 0.6 后数据恢复率

开始下降,且随着触发率越高而下降速度越快,这是因为 SSPLBP 方案低触发率下只采集单个源节点的数据,并通过分段进行编码,能很好地适应低触发率数据采集环境,但随触发率升高使得更多节点进行数据采集,而源数据缓存区有限导致没有收到随机游走包的数据因为收到新的源数据被消除了,使得数据恢复率快速下降。所给 DSCU 方案在低触发率情况下同样具有较高数据恢复率,并且当触发率超过 0.8 时数据恢复率才开始下降,而且随着触发率越高而下降速度比较趋于平缓,这是因为所给 DSCU 方案不仅在低触发率下只采集单个源节点的数据并进行分段编码,而且通过动态调整编码大小能有效减少所需存储的源数据段大小,从而能够进一步改善数据恢复率。

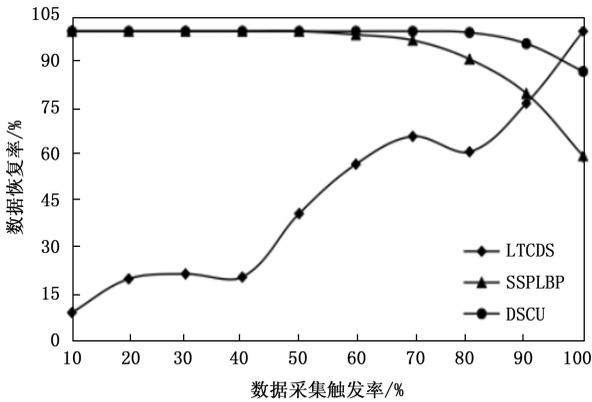


图 5 不同方案下数据恢复率对比图

2.2.2 总分段数与总游走包个数

由于编码单元的划分方式不同,采集节点每产生一批源数据其分段的个数是不尽相同的,因此其所发出的随机游走包的个数也是有所差别的。通过大量实验求均值得出了各个编码单元方案下总的源数据分段数与总游走包个数对比情况,如图 6 所示。

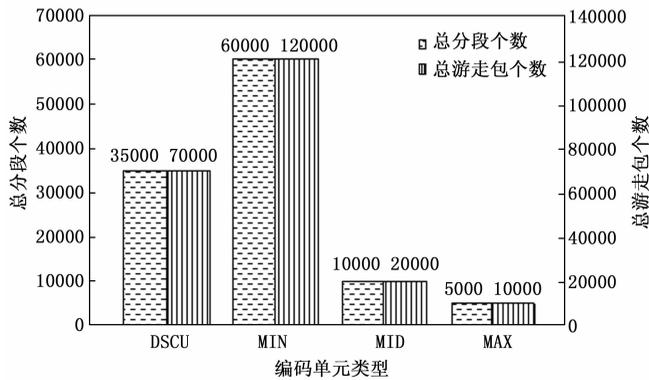


图 6 不同编码单元方案下总分段数与总游走包个数对比图

在 MIN 方案中,网络中的节点无论何时其划分编码单元的方式都是取其最小值,从而其分段个数和随机游走包的个数远大于其他方案。由于大量的随机游走包需要在网络中游动,感知节点的能耗将大大增加。尽管 DSCU 方案的解码率稍低于 MIN 方案,但是 DSCU 方案中节点转发随

机游走包的次数为 MIN 的一半,大大减少了总体的转发能量消耗。

2.2.3 存储空间限制

随机游走包在感知网络中的节点之间游走的过程中,由于编码存储区的剩余空间的限制,其无法在本节点完成存储的操作,在随机游走包剩余游走步长未用完的情况下将继续转发,而当步长用完时将直接丢弃。对不同编码单元方案下因编码存储区空间不足而造成的游走包的转发率以及游走包剩余游走步长走完因存储空间不足而丢弃的包在所有随机游走包中所占的比例进行分析,如图 7 所示。

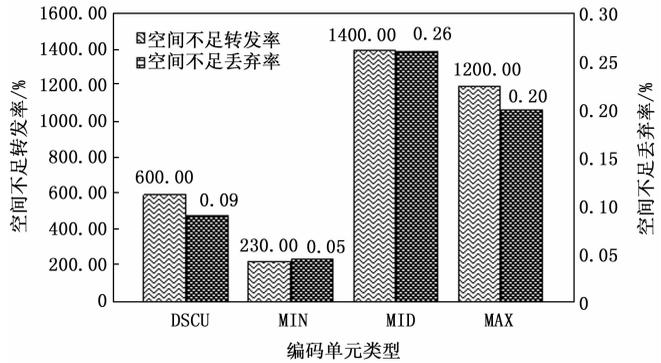


图 7 不同编码单元方案下存储空间不足转发率及丢弃率对比图

数据的恢复率和收集到的编码数据包的个数有关,收集到的编码数据包越多则原始数据有更大的恢复概率。而对于存储空间对于 MAX 和 MID 方案,由于编码单元较大,大量节点没有足够的存储空间进行数据的存储,于是便造成了数据包的丢弃,从而导致了恢复率的下降。而由于提出的 DSCU 方案能够对网络空间的剩余存储空间进行感知,从而动态地调整编码单元的大小来减少存储空间限制所带来的负面影响,所以其因存储空间限制而产生的丢包比例要远小于 MAX 和 MID 方案。

2.3 灾难场景下 DADSCR 的实验效果

由于感知区域往往存在一定的危险,下面通过实验来模拟灾难的发生,并验证 DADSCR 在其中的效果。

2.3.1 增加灾难场景

当感知区域处于灾难多发的场景下时,存储在节点中的数据将面临丢失的风险,为了减少因为节点失效而带来的影响,考虑提高编码数据在网络中的可靠性。为了提高数据的可靠性,考虑根据感知区域的危险程度来提高编码冗余度,从而增加网络中的编码包的个数。将每个时刻点灾难发生的概率 $p_{disaster}$ 设为 0.005,将 DADSCR 方案中用于计算的时间跨度 r 设为 8,将在后续的工作中进一步研究其与灾难扩散速度的关系。灾难的发生中心设置在 $1400\text{ m} \times 1400\text{ m}$ 区域中的随机位置,灾难的形状设置为半径不断扩张的圆形,其初始半径设置为 5,每个时间点半径延长 0.2,总共扩散 300 次。仿真实验都是基于前文的 DSCU 方案,从 2000 个时间点的实验过程抽出其中的 4 个绘制出网络状态便于读者理解,如图 8 所示,从图中能大致了解灾

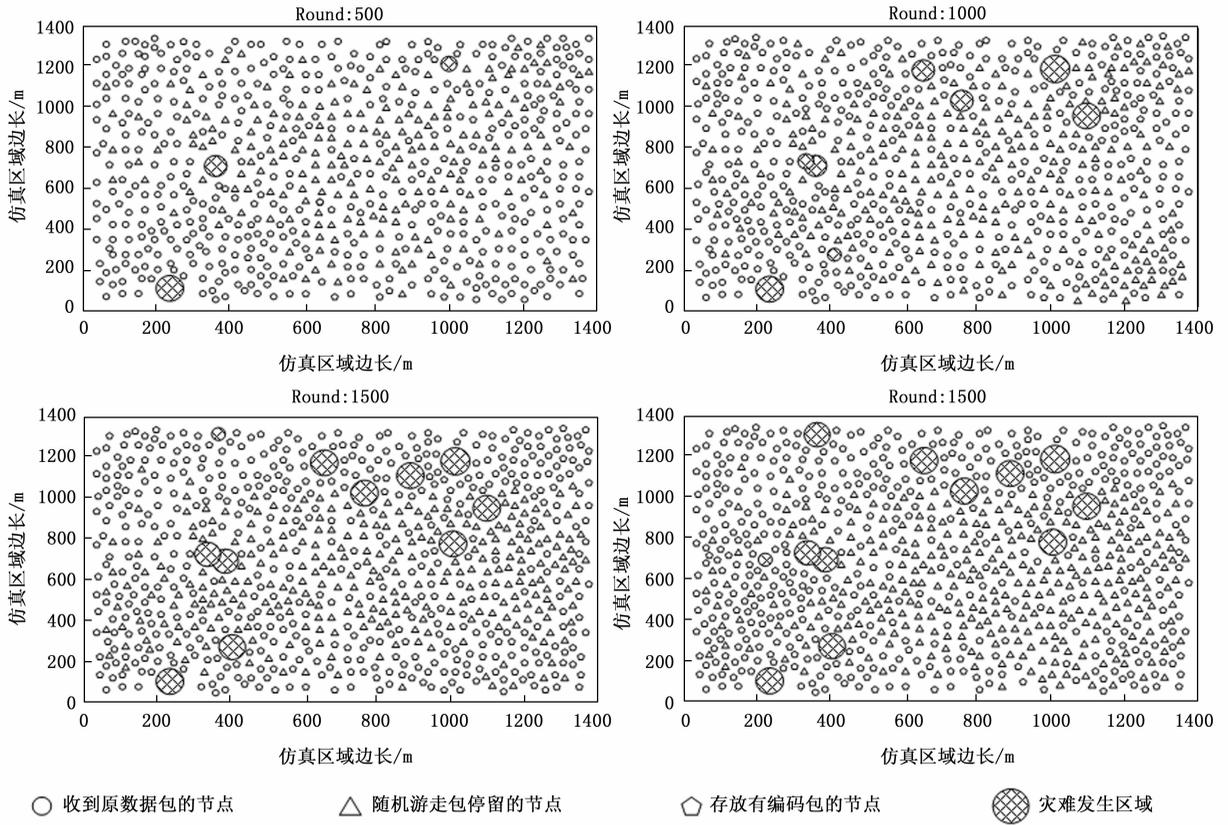


图 8 各个时间点网络状态图

难场景下数据收集的状态。

2.3.2 不同的最大随机游走步长

对采用 DADSCR 方案和仅仅使用原始最低冗余度的源端编码方案 (是指将源节点数据编码后通过多个单播传递) 进行多次实验, 得出的在不同的随机游走最大步长下的数据恢复率如图 9 所示。

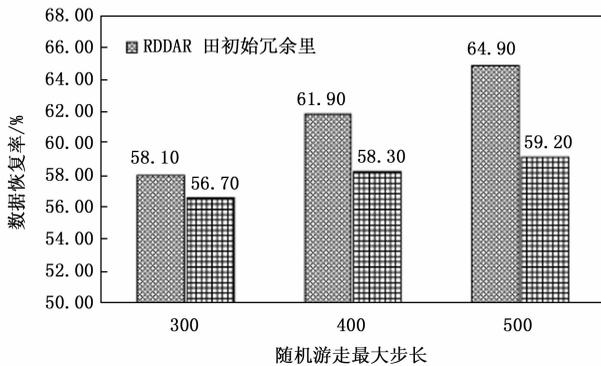


图 9 DADSCR 与初始冗余量在不同步长下的数据恢复率

DADSCR 方案在随机游走最大步长较小的时候优势不明显, 而在随机游走步长较大的时候则能够体现出较大的优势。这是因为当随机游走最大步长较小时, 游走包无法充分的分散在网路中, 在步长用完后要么被无源数据的节点丢弃, 要么存放在已经有该同批次数据的节点中。但是理想情况是希望节点能够尽量分散在不同的节点中, 从

而对节点的失效能够有较大的容忍度。因为当节点中存在多个同批次的编码包时, 若此节点失效则会丢失更大比例的该批次的编码包, 从而导致恢复该批次源数据的时候拥有较小的恢复率。

2.3.3 弥散型灾难和非弥散型灾难场景

同时还对比了 DADSCR 方案在弥散型灾难和非弥散型灾难场景下解码率的区别。非弥散型灾难发生的地点依然是随机的, 但是其灾难区域是非扩散型的, 其发生之初的灾难半径将不再改变。两者的恢复率如图 10 所示。

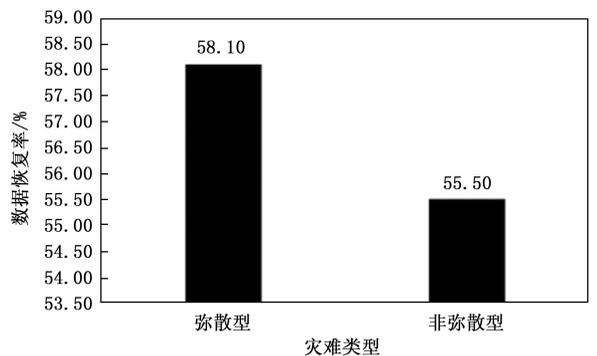


图 10 DADSCR 在弥散型灾难与非弥散型灾难下的数据恢复率

由于 DADSCR 在设计之初考虑的是一种通用的灾难感知策略, 利用节点之间邻居个数的变化来进行灾难的感知,

所以并没有采用特殊传感器来感知灾难的发生。因此, DADSCR 在面对非弥散型灾难的场景下, 能获取到的节点变动信息较少, 从而结果并不理想。

3 结束语

所给单源分段编码模型实现了大规模触发型数据在灾难场景下的部分区域数据延时可靠收集。其中, 为了平衡网络中受限的存储空间以及收集效率之间的矛盾, 针对源数据编码单元的划分进行了动态的调整, 提出了动态划分编码单元。在此基础上, 提出了危险感知编码冗余量动态调整, 利用感知网络中固有的节点之间的邻居个数的变化, 来推测感知区域中灾难的发生状态, 从而动态的调整所需要散布在网络中的编码数据的个数, 进而提高数据的恢复率。

参考文献:

- [1] 魏连锁, 马敬云, 郭 媛. 压缩感知下最短路径的无线网络数据收集算法 [J]. 东北石油大学学报, 2022, 46 (3): 98-106.
- [2] 钟志峰, 吴佳杰, 周 民, 等. 基于 WSN 节点部署中数据采集能量优化研究 [J]. 计算机测量与控制, 2018, 26 (3): 306-309.
- [3] 牟 青, 冯 琳, 魏振春. 基于烟花算法的 WSN 按需充电与数据采集策略 [J]. 合肥工业大学学报 (自然科学版), 2021 (11): 1476-1482.
- [4] 钟 萍, 徐爱昆, 张艺雯, 等. 一种高效低能耗移动数据采集与无线充电策略 [J]. 软件学报, 2021, 32 (9): 2867-2886.
- [5] KAMRA A, FELDMAN J, MISRA V, et al. Growth codes: Maximizing sensor network data persistence [J]. ACM SIGCOMM Computer Communication Review, 2006, 36 (4): 255-266.
- [6] MUNARETTO D, WIDMER J, ROSSI M, et al. Resilient coding algorithms for sensor network data persistence [C] // European Conference on Wireless Sensor Networks, 2008: 156-170.
- [7] WU M Q, KARANDE S S, RADHA H. Network-embedded FEC for optimum throughput of multicast packet video [J]. Signal Processing-Image Communication, 2005, 20 (8): 728-742.
- [8] ZHAO R, LI F Q, TIAN L L, et al. Comprehensive analysis of the whole coding and non-coding RNA transcriptome expression profiles and construction of the circRNA-lncRNA co-regulated ceRNA network in laryngeal squamous cell carcinoma [J]. Functional & Integrative Genomics, 2019, 19 (1): 109-121.
- [9] LIU B D, DONG P, ZHANG Y Y. Design of a network coding concurrent multipath transmission algorithm in complex mobile environment [J]. 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN 2019), 2019: 768-772.
- [10] BRUNISHOLZ P, ERDENE-OCHIR O, ABDALLAH M, et al. Network coding versus replication based resilient techniques to mitigate insider attacks for smart metering [J]. 2015 International Journal of Distributed Sensor Networks, 2015: 127-136.
- [11] LUBY M. LT codes [C] //The 43rd Symposium on Foundations of Computer Science (FOCS 2002), Vancouver, BC, Canada, Proceedings, 2002: 271-280.
- [12] ALY S A, KAMAL A E, AL-KOFAHI O M. Network protection codes: Providing self-healing in autonomic networks using network coding [J]. Computer Networks, 2012, 56 (1): 99-111.
- [13] LIN D D, LIM T J. Subspace-based active user identification for a collision-free slotted ad hoc network [J]. IEEE Transactions on Communications, 2004, 52 (4): 612-621.
- [14] DIMAKIS A G, PRABHAKARAN V, RAMCHANDRAN K. Distributed fountain codes for networked storage [C] //IEEE International Conference on Acoustics, 2006: 1149-1152.
- [15] KUMAR C P, SELVAKUMAR R. Reliable and secure data communication in wireless sensor networks using optimal locally recoverable codes [J]. Peer-to-Peer Networking and Applications, 2020, 13 (3): 742-751.
- [16] PEI X Q, WANG Y J, MA X K, et al. A decentralized redundancy generation scheme for codes with locality in distributed storage systems [J]. Concurrency and Computation-Practice & Experience, 2017, 29 (8): 17.
- [17] ASIRI M, SHELAMI T, AL-AWAMI L, et al. A novel approach for efficient management of data lifespan of IoT devices [J]. IEEE Internet of Things Journal, 2020, 7 (5): 4566-4574.
- [18] LIU S C, KOTTAPALLI V A, KIREMIDJIAN A S, et al. Two-tiered wireless sensor network architecture for structural health monitoring [C] //Smart Structures and Materials 2003; Smart Systems and Nondestructive Evaluation for Civil Infrastructures, 2003: 8-19.
- [19] YI S, AZIMI-SADJADIT B, KALYANARAMAN S, et al. Error control code combining techniques in cluster-based cooperative wireless networks [C] // 2005 IEEE International Conference on Communications (ICC 2005), 2005: 3510-3514.
- [20] ZHANG Y, CHEN J, TANG M, et al. Distributed LT codes on multi-hop networks [C] //Proceedings of 2nd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2017), Atlantis Press, 2017: 548-555.
- [21] CAO N, WANG Q, REN K, et al. Distributed storage coding for flexible and efficient data dissemination and retrieval in wireless sensor networks [C] //Proceedings of IEEE International Conference on Communications (ICC 2010), Cape Town, South Africa, 2010: 1-5.
- [22] LIU X C, ZHANG J P, SUN G L, et al. Multi-timeslots data collection with low rank and modified second-order horizontal total variation for wireless sensor networks [J]. IEEE Access, 2021, 9 (1): 7921-7929.
- [23] ALY S A, KONG Z, SOLJANIN E. Fountain codes based distributed Storage algorithms for large-scale wireless sensor networks [C] //2008 International Conference on Information Processing in Sensor Networks (IPSN 2008), MO: IEEE, 2008: 171-182.