LoRa网络中基于深度强化学习的信息年龄优化

程克非[①] 陈彩蝶[①] 罗 佳*[①②] 陈前斌[②]

[①](重庆邮电大学网络空间安全与信息法学院 重庆 400065)

[②](重庆邮电大学通信与信息工程学院移动通信技术重点实验室 重庆 400065)

摘 要：信息年龄(AoI)是信息新鲜度的衡量指标，针对时间敏感的物联网，最小化AoI显得尤为重要。该文基于LoRa网络的智能交通环境，分析Slot-Aloha协议下的AoI优化策略，建立了Slot-Aloha协议下数据包之间传输碰撞和等待时间的系统模型。通过分析指出，在LoRa上行传输过程中，随着数据包数量增多，AoI主要受到数据包碰撞影响。为克服优化问题中动作空间过大导致难以实现有效求解的问题，该文采用连续动作空间映射离散动作空间的方式，使用柔性动作-评价 (SAC)算法对LoRa网络下的AoI进行优化。仿真结果显示，SAC算法优于传统算法与传统深度强化学习算法，可有效降低网络的平均AoI。

关键词：信息年龄；LoRa；柔性动作-评价算法；深度强化学习；优化策略

中图分类号：TN929.5 **文献标识码**：A **文章编号**：1009-5896(2025)02-0541-10

DOI: 10.11999/JEIT240404

1 引言

在当今数字化时代，智能交通管理技术正在迅速崛起，为实现交通系统的高效运作和安全性提升提供了前所未有的机遇。其中，大规模传感器网络在交通数据采集和决策制定中扮演着至关重要的角色。长距离低功耗射频(Long Range low power radio frequency technology, LoRa)网络以其广泛的覆盖范围和低功耗的特性而备受瞩目，成为连接交通传感器的理想选择。在智能交通场景中，交通监控摄像头需要频繁向LoRa网关进行数据传输，且城市内交通监控摄像头数量有限，其对实时监测的需求和数据包的新鲜度较高，以便及时采取措施来优化交通状况。为了更好地衡量数据的新鲜度，本文采用AoI (Age of Information)来量化数据包的时效性。AoI体现了数据在调度和传输过程中的延迟情况，反映了数据在被采集/生成后多长时间内可用于进行决策、分析或其他应用。AoI被定义为接收端最新接收到数据包的生成时间距离当前时间的时间间隔[1]。

最初，针对AoI的研究主要集中于调度策略[2]，特别是在单服务器排队系统中对不同调度策略的AoI表现进行对比分析。文献[3]研究提出了一种“后生成先服务”(Last Generated First Served, LGFS)调度策略，并在不同服务时间分布的条件下，优化单跳排队系统中的AoI。此外，文献[4]在信道状态未知的情况下，探讨了最优调度策略以实现AoI的优化。近年来，AoI的应用逐渐扩展至无线网络领域。文献[5]提出了一种基于调度策略的AoI优化方法，用以解决链路传输调度问题。文献[6]则考虑了一种依赖资源管理方案的边缘设备辅助智能医疗系统，设计了两种基于强化学习的算法，以最小化系统能耗和平均AoI。文献[7]提出了一种基于改进的决斗深度双Q网络(Improved Dueling Double Deep Q-Network Computation Offloading, ID3CO)算法，用于优化工业物联网中的AoI，仿真结果表明，该算法在性能上优于现有方法。文献[8]基于无人机人群感知的数据采集模式，提出了一种去中心化的多智能体深度强化学习框架DRL-UCS，仿真结果显示该框架在特定应用场景中能够有效优化AoI。目前，关于LoRa网络中AoI的研究相对较少。文献[9]首次在LoRa网络中引入AoI及其优化问题，重点分析了丢包模型，并提出了FreshLoRa-A和FreshLoRa-B两种算法以改善AoI。仿真结果将这些算法与文献[10]提出的两种自适应速率分配算法进行了对比分析。

在LoRa网络中，上行传输采用纯Aloha协议，设备可以在任意时间发送数据，无需等待特定的时隙。每个设备独立决定何时发送数据，然而当两个或多个设备同时发送时，可能会发生信号碰撞。文献[11]探讨了采用纯Aloha机制信道接入LoRa的可扩展性和性能，通过数学模型和模拟分析发现，纯Aloha机制在高密度网络中可能导致性能下降。LoRa网络依赖于啁啾扩频技术，其中扩频因子(Spreading Factor, SF)定义为扩频后啁啾速率与原始信号

收稿日期：2024-04-04；改回日期：2025-01-08；网络出版：2025-01-25

*通信作者： 罗佳 s220802003@stu.cqupt.edu.cn

基金项目：重庆市教委科学技术研究项目(KJQN202400643)

Foundation Item: The Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202400643)

速率的比率。许多文献基于纯Aloha协议提出分配算法为设备分配适当的SF用以提高LoRa网络的可扩展性。文献[12]比较了传统的SF分配方案，提出了一种基于瞬时信道状态的动态SF分配算法，以优化LoRa网络的整体性能。该算法根据设备的信道增益动态调整SF，信道条件较差的设备分配较高的SF，信道条件较好的设备则分配较低的SF。通过误码率(Symbol Error Rate, SER)评估，该动态分配方法在性能上优于传统的分配方案。文献[13]提出了一种干扰感知的SF分配算法，旨在降低信号干扰的显著性，从而提高数据包成功传输率，该算法在仿真中表现出良好的性能。文献[14]则提出了基于自适应速率(Adaptive Data Rate, ADR)的两种SF分配算法：EXPLoRa-SF和EXPLoRa-AT。EXPLoRa-SF通过改进传统的SF自适应算法，根据终端数据包的重传次数动态调整SF，避免过多终端使用高SF，从而减少数据包碰撞的概率。EXPLoRa-AT则根据终端灵敏度1.569在确认和非确认模式下动态分配SF。仿真结果表明，这两种方法在性能上相比原有算法均有显著提升。

然而，在数据包碰撞的情况下，设备可能无法成功传输其数据包。随着网络中设备数量的增加，多个设备在同一时间发送数据包的概率上升，导致较高的碰撞率。为减少LoRa网络中的数据包碰撞，Aloha协议引入了时隙机制，将时间划分为一系列时隙，每个时隙对应一个固定的时间窗口，终端仅在时隙开始时进行数据发送。文献[15]的仿真结果表明，采用时隙Aloha可以有效降低数据包的碰撞率。文献[16]通过将时隙Aloha建模为马尔可夫链，分析了物联网系统中的AoI表达式，并通过仿真探讨了系统中可调参数的影响。文献[17]则对基于时隙Aloha的LoRa物联网进行了性能评估，推导出平均吞吐量和预期延迟的表达式。近年来，许多研究通过引入深度强化学习来优化LoRa网络的资源分配，以增强其可扩展性。文献[18]提出了一种由电网和可再生能源混合供电的绿色LoRa无线网络，并使用深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)和近端策略优化(Proximal Policy Optimization, PPO)算法对能耗和绿色能源的管理进行优化。文献[19]基于SF分配模型，提出了LR-opt-pro, LR-greedy和LR-RL共3种深度强化学习算法，并与现有的SF分配方法(如ADR和EXPLoRa-AT)进行比较。仿真结果表明，LR-RL算法在解决数据包冲突方面具有显著的性能优势。

当前关于AoI优化的研究中，针对LoRa网络的成果相对较为稀缺，且少数相关研究尚未充分考虑在时隙Aloha协议下的系统性能。此外，由于AoI的动态特性，现有研究大多采用基于深度强化学习的优化方案。然而，针对LoRa网络的研究主要聚焦于纯Aloha协议，对于时隙Aloha协议下的AoI优化，尤其是结合深度强化学习的解决方案，仍然缺乏系统性的探讨。基于此，本文的贡献体现在以下两个方面：

(1)基于时隙Aloha协议的LoRa网络AoI优化研究：本文结合智能交通场景中对信息新鲜度的要求，研究了时隙Aloha协议下LoRa网络的数据包AoI优化问题。针对LoRa网络中存在的频繁数据传输，建立了一个基于LoRa数据包碰撞的系统模型，重点分析了时隙Aloha协议下的数据包碰撞与空中传输时间对LoRa网络中AoI的影响，为提高信息传输效率提供理论支持。

(2)采用SAC (Soft Actor-Critic)算法优化AoI：考虑到AoI的时间演化具有马尔可夫性，本文将优化问题建模为马尔可夫决策过程(Markov Decision Process, MDP)，并采用深度强化学习中的SAC算法进行求解。通过仿真对比不同深度强化学习算法及传统贪婪算法的系统性能，结果显示，SAC算法在本文所述的场景下能够有效提升系统性能，实现更优的AoI优化效果。

2 系统模型

如图1所示，考虑基于LoRa物联网上行传输的智能交通管理场景。N个监控摄像头作为终端设备分布在道路和交通枢纽中，监测和控制交通信号状态，以实时优化交通信号。单个LoRa网关作为基站对终端数据进行收集和传输。下文将基于LoRa网络的空中传输时间以及数据包碰撞来推导相应的AoI。

在LoRa网络中，数据包从终端发出后需要经历一段时间才能被网关接收。这个时间称为空中传

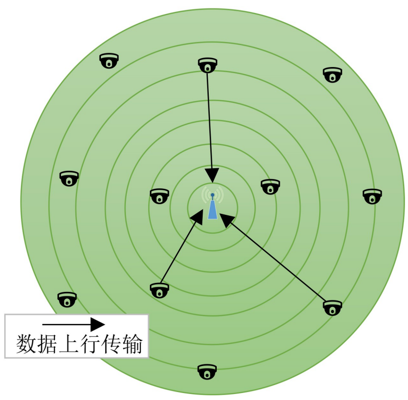


图 1 系统模型

输时间(Time on Air, ToA)，SF越大终端的ToA就越大，信息的新鲜度就越低。由此可见，时隙Aloha协议下，同一时隙不同数据包在相同的SF和信道条件下进行传输会发生碰撞。

2.1 空中传输时间

在LoRa网络中，数据包的传输速率(R_b)与SF、带宽(BW)以及编码率(CR)密切相关。SF是影响符号原始承载比特数的重要参数，其符号时间T_s可由公式$T_\mathrm{s}=2^{\mathrm{SF}}/\mathrm{BW}$表示。更高的SF值意味着更长的符号时间，BW定义为LoRa上行传输的带宽，编码率为数据流中非冗余部分的比例，其可通过$\mathrm{CR}=4/(4+\theta)$表示[20]，其中θ取值为1, 2, 3或4。根据上述关系，数据包速率可通过式(1)计算

$$R_\mathrm{b}=\frac{\mathrm{SF}\times\mathrm{BW}\times\mathrm{CR}}{2^{\mathrm{SF}}} \tag{1}$$

空中传输时间T^a为

$$T^\mathrm{a}=\frac{L_\mathrm{d}}{R_\mathrm{b}} \tag{2}$$

其中，L_d为数据包长度，为了便于计算，假设在LoRa网络中数据包长度为50 Byte，且在带宽125 kHz、编码率4/5的情况下，随着SF的变化，空中传输时间的大小如表1所示。

2.2 信息年龄

用AoI来衡量LoRa网络中数据的及时性和新鲜性，AoI的定义是接收端接收到的最新数据包的当前时间距离其生成时间的时间间隔。为了增加应用的多样性，LoRa通过协调功耗和通信时延之间的关系规定了3种工作模式，分别为双工终端模式(Class A)，划定接收时隙的双向传输模式(Class B)，可最大化接收时隙的双向传输模式(Class C)。其中，A类传输模式的终端在每次上行传输后都会紧跟两个短暂的下行接收数据窗口来实现双向传输，这种A类传输模式有非常高的能效。B类传输模式可在指定时间打开下行接收窗口，提供双向通信。C类传输模式处于完全打开接收窗口状态，仅在发送时短暂关闭，也更耗能。在LoRa网络中A类上行传输遵循纯Aloha协议，该协议在终端频繁向网关发送数据包的情况下极易产生严重数据包碰撞，导致性能低下，因此本文基于LoRa网络的时隙Aloha协议进行分析。为了易于分析，假设所有终端设备围绕一个网关设备呈均匀分布，每个终端设备在单个时隙内均匀随机生成数据包，设数据包在固定带宽和编码率的信道条件下进行传输，时隙长度为T_sl，数据包在发生碰撞后直接被丢弃。$\phi_\mathrm{c}(t)=\{0,1\}$代表在$t$时隙下数据包是否发生碰撞，如图2(a)所示，终端只在每个时隙开始时传输数据，基于定义可知，数据包之间不发生碰撞$\phi_\mathrm{c}(t)=0$，传输成功后网关处AoI更新为新数据包的AoI，如图2(b)所示，数据包之间发生碰撞$\phi_\mathrm{c}(t)=1$，网关一直没有接收到新的数据包，导致AoI持续增加，网关处最新数据则会变得“不新鲜”。

表 1　空中传输时间

SF	7	8	9	10	11	12
T^a (ms)	73.1	128	227.6	409.6	744.7	1365.3

每个终端在单位时隙中均匀随机地生成数据包，且终端队列容量仅为1，在队列里有数据包的情况下生成新的数据包，应该丢弃队列中的数据包，将新的数据包放入队列，确保上一个数据包传输完成后，下一个要传输的数据包最“新鲜”。

数据包的AoI计算公式为

$$A^\mathrm{o}(t)=\begin{cases}T^\mathrm{a}(t)+W^\mathrm{t}(t),\varphi_\mathrm{c}(t)=0\\A^\mathrm{o}(t-1)+T_\mathrm{sl},\varphi_\mathrm{c}(t)=1\end{cases} \tag{3}$$

其中，W^t为数据包的生成到发送所经历的时间，

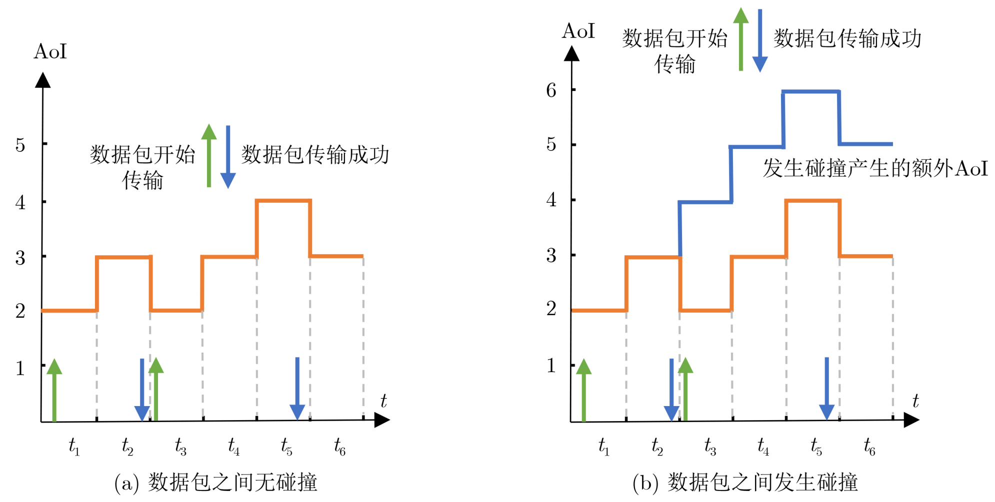


图 2 基于时隙Aloha的数据包AoI变化情况

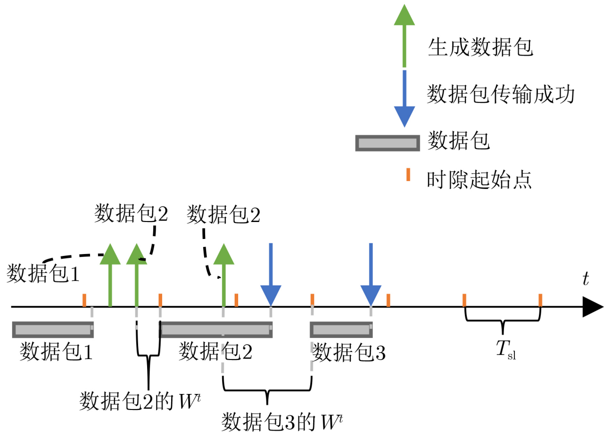


图 3 数据包传输情况

由于在上一个数据包传输过程中，都可能随机生成其余数据包，因此数据包的生成到发送时间范围为$[0, T_{sl})$，其中T_{sl}为数据包经历的时隙长度，则W^t的概率密度函数为

$$f(x)_{W^t} = \begin{cases} \dfrac{1}{T_{sl}}, 0 \le x < T_{sl} \\ 0, 其他 \end{cases} \tag{4}$$

由于单个LoRa网关范围内的城市交通监控摄像头数量较小，且数据传输频繁，因此系统模拟每个时隙均会产生数据包。图3描述了数据包传输的情况，当终端在下一个时隙还未到达期间生成了数据包1和数据包2，在终端的队列中保留数据包2丢弃数据包1。在下一个时隙到达时，数据包2开始传输，数据包2的传输时间横跨两个时隙，并在第二个时隙结束之前提前传输完成，若终端在数据包传输期间生成了数据包，则需等待数据包传输完成且等待下一个时隙到达再进行传输，数据包3在数据包2还未传输完的期间生成，数据包3的等待时间分为等待数据包2传输的时间以及等待下一个时隙到达的时间。同时，数据包的传输时间可能跨越几个时隙，也可能小于1个时隙。

如图4所示，展示了各终端的数据包传输情况。在终端1的时间线上，时隙1内产生了数据包1，此时终端队列为空，因此将数据包1存入队列并在下一个时隙开始时进行传输。时隙2内产生了数据包2，但由于数据包1尚未完成传输，终端继续传输数据包1，并将数据包2存入队列。时隙3内，数据包1传输完成，同时生成了数据包3。为保证数据的“新鲜度”，终端丢弃了队列中的数据包2，将数据包3加入队列，并在时隙4开始时传输数据包3。因此，在时隙3末尾，更新到系统的终端1的AoI，A_1^{o}为$2T_{sl}$。

当两个终端在同一时隙选择相同的SF和信道，且它们正在传输的数据包在时间上重叠时，数据包将发生碰撞。比如在图4中，若终端2和终端3在时隙2选择了相同的SF和信道，则数据包8和数据包15会发生碰撞并被丢弃。碰撞发生后，时隙3开始时将传输数据包9和数据包16。碰撞会导致AoI在原有的基础上增加1个时隙的时间，即叠加T_{sl}。

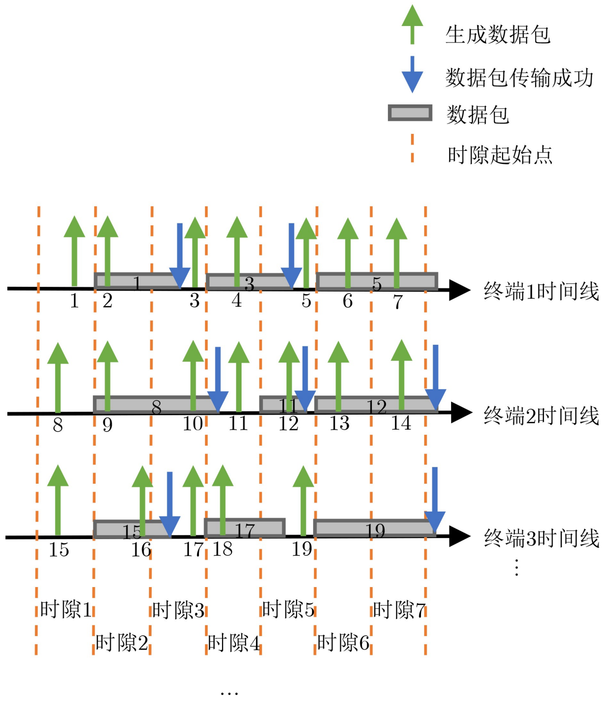


图 4 各终端数据包传输情况

3 问题建模与求解算法

3.1 传统算法模型

在系统模型中，碰撞和空中传输时间会影响LoRa网络中整体AoI，贪婪算法(算法1)将选择适当信道和SF使得当前时隙下的AoI得到最优，通过式(3)计算AoI，使在当前时隙下整体AoI最低，贪婪算法步骤如下：

(1) 随机初始化终端的SF和信道。

(2) 遍历所有信道和SF的组合找到当前终端最适合的信道和SF，以达到当前熟悉下AoI的最小化。

(3) 重复上述步骤直到所有终端分配完成。

上述贪婪算法中，每次更新result都会遍历所有SF和信道的组合，其时间复杂度为$O(c)$，而要最终得到result需要遍历所有终端，因此时间复杂度为$O(cn)$。从时间平均的角度看，贪婪算法仅考虑了当前时隙下的最优AoI，没有考虑到多个时隙之间的关联性，而MDP考虑了当前时隙状态和之前多个时隙状态的关系，能够更好地适应时隙之间的关联性从而达到更好的AoI优化效果。

3.2 问题建模

在深度强化学习中，MDP考虑了智能体在不

算法 1 贪婪算法

```
输入：终端集合sensors = {sen_1, sen_2, ⋯, sen_n}
输出：每个终端分配的SF和信道集合
result = {(SF_1, C_1), (SF_2, C_2), ⋯, (SF_n, C_n)}
result ← ∅, sf ∈ {7, 8, ⋯, 12}, 信道C ∈ {0, 1, ⋯, c}
for i = 1, 2, ⋯, N do
    SF_i ← 随机从sf集合中选择
    C_i ← 随机从C集合中选择
    result ← result ∪ (SF_i, C_i)
end for
for j = 1, 2, ⋯, N do
    记录第j时隙下最优的SF bestSF_j = 0
    记录第j时隙下最优的信道C bestC_j = 0
    minavgAoI = M_INT
    for sf ← 7 to 12 do
        for C ← 0 to c do
        avgAoI = average (AoI)
        if avgAoI < minavgAoI
            minavgAoI=avgAoI
            bestSF_j = sf
            bestC_j = C
        end for
        end for
    result ← result ∪ (bestSF_i, bestC_i)
    end for
return result
```

同状态下采取的动作对环境的影响，以及由此产生的奖励反馈。在LoRa网络中，数据包之间的碰撞将导致丢包率的上升，进而增加系统的AoI。通过合理分配SF和信道，可以有效降低数据包的碰撞率，从而优化AoI。

在第3节的模型中，为了在某个状态下实现整体AoI的最优，需要为每个终端分配合适的SF和信道。此时，终端作为智能体(agent)，其行为通过深度强化学习进行决策。由于各终端的动作是相互独立的，即每个终端的信道和SF分配独立于其他终端，因此该问题中的动作空间是离散的。通过对智能体在不同状态下的奖励信号进行优化，系统能够最小化数据包碰撞，进而实现AoI的优化目标。

(1) 环境：如图5所示，与智能体交互的LoRa网络为环境，根据策略函数π智能体在步骤t生成动作A_t，根据环境情况，分配奖励R_t，同时进入下一个状态S_{t+1}。在本文的深度强化学习框架中，将每个步骤(step)定义为对所有终端分配信道和SF，每个时隙都是1个step，$T_{\rm st}$个step构成1个episode，

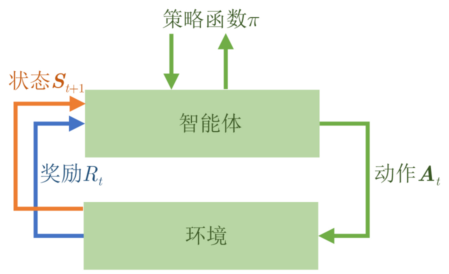


图 5 状态转移图

由于每个episode之间相互独立，因此每个episode开始时，累计奖励初始化为0。

(2) 动作和状态：动作集合为$\boldsymbol{A} = \{{\rm SF}(i), \boldsymbol{C}(i)\}$，其中$\mathbf{SF}(i)$为每个终端在第$i$个时隙分配的SF的集合，$\mathbf{SF}(i) = \{{\rm SF}_1(i), {\rm SF}_2(i), \cdots, {\rm SF}_n(i)\}$；$\boldsymbol{C}(i)$为每个终端第$i$个时隙分配的信道的集合，$\boldsymbol{C}(i) = \{C_1(i), C_2(i), \cdots, C_n(i)\}$，设状态集合为$\boldsymbol{S} = \{\boldsymbol{A}^{\rm o}(i), \boldsymbol{T}^{\rm a}(i), \boldsymbol{A}(i), EX(i), \boldsymbol{W}^{\rm t}(i)\}$，其中$A^{\rm o}(i)$为第$i$个时隙每个终端发送到网关处的最新数据包的AoI集合，$\boldsymbol{A}^{\rm o}(i) = \{A_1^{\rm o}(i), A_2^{\rm o}(i), \cdots, A_n^{\rm o}(i)\}$；$T^{\rm a}(i)$为对应每个终端上一个时隙中数据包剩余的空中传输时间的集合，如果在第i个时隙时数据包未传输完，$T_j^{\rm a}(i)$为第j个终端之前剩余的空中传输时间，如果第j个终端之前的数据包已经传输完成，那么$T_j^{\rm a}(i)$为0；$A(i)$为数据包之前时隙下分配的动作集合，若之前时隙的数据包在第i个时隙仍然在传输，则$A_j(i)$为第j个终端之前分配的动作；$\mathbf{EX}(i)$为第i个时隙还没传输完的数据包在不发生碰撞的情况下，会经历的整个时隙长度；$\boldsymbol{W}^{\rm t}(i)$表示第$i+1$个时隙的数据包的等待时间。为了便于计算，将状态扁平化为一维向量，则状态$\boldsymbol{S}$是长度为$4N$的一维数组。

(3) 奖励函数：本文的优化目标是让整体LoRa网络的AoI最优化，则奖励函数应该为每个step结束后整体LoRa网络平均AoI的负值，因此，时隙i奖励函数可计算为

$$r_i = -(A_1^{\rm o}(i) + A_2^{\rm o}(i) + \cdots + A_n^{\rm o}(i))/N \tag{5}$$

3.3 基于深度强化学习的优化方案

SAC算法是一种基于最大熵的随机策略深度强化模型框架，SAC的主要特性之一为熵正则化(entropy regularization)，其中熵的计算是其核心部分之一，熵主要用于鼓励策略(policy)探索更多不同的行为，同时用来量化策略的随机性或不确定性，熵的增加会增强策略的随机性，从而增加更多的探索，加快后续的学习速度。从文献[21]可知，熵的计算公式如

$$H(\pi(\cdot|\boldsymbol{s})) = -E_{a\sim\pi(\cdot|s)}[\ln P(\pi(\boldsymbol{a}|\boldsymbol{s}))] \tag{6}$$

其中，$\pi(\cdot|\boldsymbol{s})$为在状态$\boldsymbol{s}$下的策略，它给出了在该状态下采取不同行动的概率分布。$E_{\boldsymbol{a}\sim\pi(\cdot|\boldsymbol{s})}$表示对于策略$P$在状态$\boldsymbol{s}$下选择行动$a$的期望值。$\ln P(\pi(\boldsymbol{a}|\boldsymbol{s}))$为行动$\boldsymbol{a}$在状态$\boldsymbol{s}$下的策略概率的对数。

标准的深度强化学习算法中，学习目标是找到累计奖励最多的策略，表达式为

$$\pi^* = \underset{\pi}{\operatorname{argmax}} \sum\nolimits_t E_{(\boldsymbol{s}_t,\boldsymbol{a}_t)\sim\rho_\pi}[R(\boldsymbol{s}_t,\boldsymbol{a}_t)] \tag{7}$$

而引入了熵的目标策略，即

$$\pi^* = \sum\nolimits_t E_{(\boldsymbol{s}_t,\boldsymbol{a}_t)\sim\rho_\pi}[R(\boldsymbol{s}_t,\boldsymbol{a}_t) + \alpha H(\pi(\cdot|\boldsymbol{s}_t))] \tag{8}$$

其中，α为熵正则化系数，$\underset{(\boldsymbol{s}_t,\boldsymbol{a}_t)\sim\rho_\pi}{E}$表示在根据策略$\pi$生成的轨迹中，对状态-动作对(state-action pair)的期望值，其中ρ_π是策略π生成的状态-动作对的分布。

SAC算法中的状态价值函数为

$$\begin{aligned} V_{\text{soft}}^{\pi}(s) = E_{(\boldsymbol{s}_t,\boldsymbol{a}_t)\sim\rho_\pi} \sum\nolimits_{t=0}^{\infty} [\gamma^t (R(\boldsymbol{s}_t,\boldsymbol{a}_t) \\ + \alpha H(\pi(\cdot|\boldsymbol{s}_t)))| \boldsymbol{s}_0 = \boldsymbol{s}] \end{aligned} \tag{9}$$

动作价值函数为

$$\begin{aligned} Q_{\text{soft}}^{\pi}(\boldsymbol{s},\boldsymbol{a}) = E_{(\boldsymbol{s}_t,\boldsymbol{a}_t)\sim\rho_\pi} \Big[\sum\nolimits_{t=0}^{\infty} \gamma^t (R(\boldsymbol{s}_t,\boldsymbol{a}_t) \\ + \alpha \sum\nolimits_{t=1}^{\infty} \gamma^t H(\pi(\cdot|\boldsymbol{s}_t)))|\boldsymbol{s}_0 = \boldsymbol{s}, \boldsymbol{a}_0 = \boldsymbol{a}\Big] \end{aligned} \tag{10}$$

由此可见状态价值函数和动作价值函数的关系为

$$Q_{\text{soft}}^{\pi}(\boldsymbol{s},\boldsymbol{a}) = E_{(\boldsymbol{s}_t,\boldsymbol{a}_t)\sim\rho_\pi}[R(\boldsymbol{s}_t,\boldsymbol{a}_t) + \gamma V_{\text{soft}}^{\pi}(\boldsymbol{s})] \tag{11}$$

与标准深度强化学习算法不一样的是SAC算法中，如果行动者网络输出的动作使综合指标(包含动作价值和熵)越大越好，动作价值网络Q输出的动作价值越准确越好，状态价值网络V输出的状态价值越准确越好。

使用tanh函数将动作控制在(–1,1)之间

$$\boldsymbol{a}_t = \tanh(\boldsymbol{a}_t) \tag{12}$$

本文的动作为给所有终端分配相应的信道和SF，对于此离散动作来说，若信道数量为c，SF数量为6时，整个离散动作空间为$(6c)^N$，可以看出这是一个非常巨大的值，普通的适用于离散动作空间的算法无法进行计算，因此本文将动作在(–1,1)的值映射到相应离散的信道和SF来达到动作为离散的。基于SAC的AoI优化见算法2。

4 仿真结果

为了使算法之间差异比较更为明显，表2为实验中各算法参数设置值。

由于增加终端数量过大会导致空间维度过大GPU过载的情况，因此降低信道数量让信道数量和SF数量的组合数小于等于终端数量，使得每个终端不能独立地分到对应的SF和信道，从而发生碰撞行为。取信道数量为2，终端数量为12，SF数量为6。LoRa网络中数据包的大小一般为10～100 Byte，因此本文取数据包大小为50 Byte。

图6为不同算法的收敛曲线，图6(a)给出了4种算法分配方式下平均AoI变化，可以看到SAC算法在3 000轮左右便几近收敛，而TD3在4 000多轮才达到收敛且收敛后性能低于SAC的相应性能。一方面，SAC算法采用最大熵策略，平衡了探索和利用

算法 2 基于SAC的AoI优化算法

输入：终端数量N
输出：动作a_t
1. 初始化：评价者网络$Q_{\omega_1}(\boldsymbol{s},\boldsymbol{a})$参数$\omega_1$，$Q_{\omega_2}(\boldsymbol{s},\boldsymbol{a})$参数$\omega_2$，行动者网络$\pi_\theta(\boldsymbol{s})$参数 θ,目标网络$Q_{\omega_1}(\boldsymbol{s},\boldsymbol{a})$，$Q_{\omega_2}(\boldsymbol{s},\boldsymbol{a})$，$\pi_\theta(\boldsymbol{s})$参数$\omega_1^-,\omega_2^-,\theta^-$，初始化经验回放池$R$
2. **for**序列episode $=1,2,\cdots,E$ **do**
3. 　初始化状态：$S=\{3T_{\rm sl},3T_{\rm sl},\cdots,3T_{\rm sl},\boldsymbol{0},\boldsymbol{0},\boldsymbol{0},\text{random}(0,T_{\rm sl}),\text{random}(0,T_{\rm sl}),\cdots,\text{random}(0,T_{\rm sl})\}$
4. 　初始累计奖励：$R_{tt}=0$
5. 　**for**时间步$t=1,2,\cdots,T$**do**
6. 　　根据当前策略选择动作$a_t=\pi_\theta(\boldsymbol{s}_t)$
7. 　执行动作a_t
8. 　获得奖励$r_t=A_{\rm av}^{\rm o}$
9. 　将该轮时间步的奖励累加到累计奖励中$R_{tt}=R_{tt}+r_t$
10. 　环境状态变为s_{t+1}
11. 　　将$(\boldsymbol{s}_t,\boldsymbol{s}_{t+1},\boldsymbol{a}_t,\boldsymbol{r}_t)$保存到经验记忆池$R$中
12. 　　**for**训练轮数$k=1\to K$**do**
13. 　　从R中采样N个元组$\{(\boldsymbol{s}_i,\boldsymbol{s}_{i+1},a_i,\boldsymbol{r}_i)\}_{i=1,2,\cdots,N}$
14. 　　计算最小化损失函数，更新两个评价者网络
15. 　　更新当前行动者网络
16. 　　更新熵正则项的系数α
17. 　　更新目标网络
18. 　　end**for**
19. 　**end for**
20. **end for**

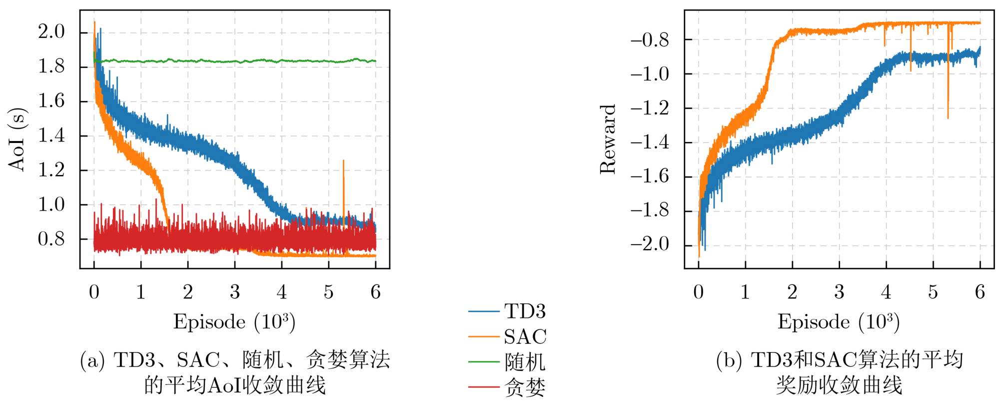
表 2 实验参数值

参数名	值
信道数量(c)	2
终端数量(N)	12
SF数量	6
编码率(CR)	4/5
带宽(BW)	125 kHz
数据包大小($L_{\rm d}$)	50 Byte
step总数($T_{\rm st}$)	500
时隙长度($T_{\rm sl}$)	500 ms

的权衡，智能体会更积极地探索环境，从而更好地学习到环境的特性，并且能够更有效地利用数据，而TD3使用确定性策略，倾向于选择在当前策略下估计最好的动作，这将导致其在探索方面不够积极；另一方面，SAC算法采用通过最小化 KL 散度来优化策略，有助于提高算法的稳定性和收敛性。贪婪算法优于TD3，劣于SAC算法，随机算法性能最差。两种强化学习算法的奖励函数为每个step终端的AoI年龄取负，如图6(b)。

图7为不同终端下的AoI，其中图7(a)为SAC算法在不同终端设备数下平均AoI的收敛曲线，图7(b)为算法在平均AoI趋于平衡时，不同终端数量下的平均AoI值。由于智能交通中监控摄像头的数量大概为8～20，因此训练终端时将终端数量设置为10～14，可以看到图7(b)中LoRa网络平均AoI随着终端数量N的增加而增加，一方面，是因为网络中数据包碰撞的概率增加，另一方面，终端分配到更大SF的概率增加，可能获得更大的空中传输时间，进而影响整体AoI。同时可以看到，在终端数量较少的情况下贪婪算法甚至优于SAC算法，这是因为在系统资源较为充裕的情况下，终端之间的数据包碰撞概率较低。在简单场景下，SAC算法可能会进行很多探索性的动作，以期望找到最优的策略，这些额外的探索动作可能会浪费资源，而贪婪算法直接利用现有信息进行决策，避免了不必要的复杂度和学习过程，反而能获得更好的即时表现。随着终端数量增加，网络变得复杂，贪婪算法由于缺乏全局视角和长期优化能力，表现会变得不稳定，而SAC算法在复杂场景下能够逐渐展现其稳定性和优越性。

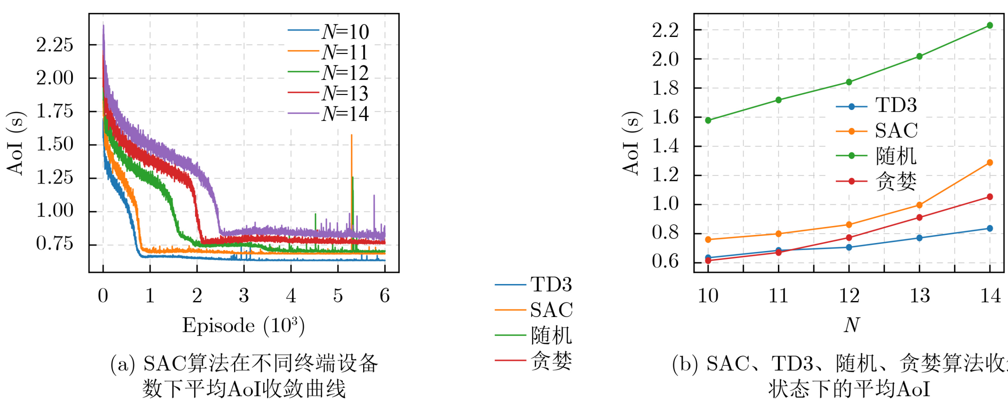
图8给出了SAC算法在不同时隙长度T_{sl}下平均AoI收敛曲线结果、TD3算法在不同时隙长度T_{sl}下平均AoI收敛曲线结果以及两者达到收敛时平均AoI的变化折线图。经过实验，100 ms以上随着时隙的减少趋于收敛时的AoI呈明显下降趋势，因此为了使平均AoI的收敛图形清晰、不重叠，100 ms以上间隔100 ms, 100 ms以下间隔20 ms。在图8(c)中，平均AoI随时隙长度T_{sl}的减少呈先减小后增加的趋势，这是因为时隙长度的减少会减少等待下一个时隙到达的时间，这样整体AoI也会减小，但若时隙长度过于小时，数据包停留的时隙数增加，在数据包密集的情况下会导致碰撞大大增加，从而使网络中的AoI变大，由此可见时隙长度的设置在优化平均AoI方面极为重要。



(a) TD3、SAC、随机、贪婪算法的平均AoI收敛曲线

(b) TD3和SAC算法的平均奖励收敛曲线

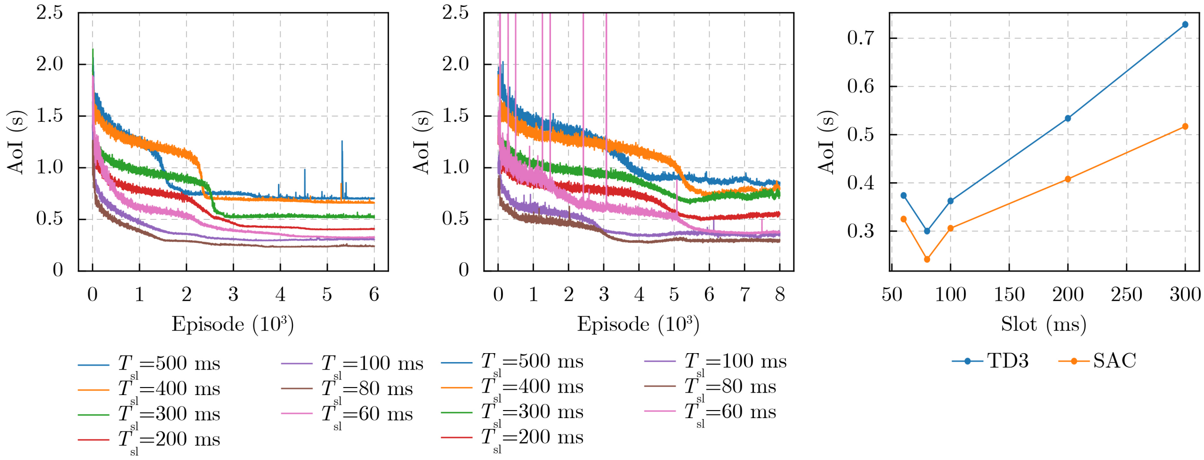
图 6 不同算法的收敛曲线



(a) SAC算法在不同终端设备数下平均AoI收敛曲线

(b) SAC、TD3、随机、贪婪算法收敛状态下的平均AoI

图 7 不同终端数量下的AoI



(a) SAC算法在不同时隙长度T_{sl}下平均AoI收敛曲线

(b) TD3算法在不同时隙长度T_{sl}下平均AoI收敛曲线

(c) TD3算法和SAC算法在收敛状态下的平均AoI

图 8 TD3算法和SAC算法在不同时隙长度T_{sl}下平均AoI变化

5 结束语

针对LoRa网络中AoI相关研究较少的情况，本文基于智能交通管理环境研究了LoRa上行数据包传输的AoI优化问题，提出了在时隙Aloha协议下的数据包碰撞模型，分别通过贪婪算法和SAC算法来优化AoI，仿真结果显示，贪婪算法优于传统深度强化学习算法劣于SAC算法，SAC算法可以有效地提高LoRa网络中的AoI优化问题。此外，本文仅考虑了网络中AoI优化问题并没有联合考虑能耗、丢包率等问题。针对这方面的不足，未来的研究可以进一步考虑能耗、丢包率、AoI优化的平衡问题，降低能耗和丢包率。另外，本文尚未涉及异构场景的研究。在LoRa网络与其他通信技术(如Wi-Fi, Bluetooth, NB-IoT等)共存的传输环境中，不同通信协议和设备类型之间的互操作性、数据一致性和网络管理将带来新的挑战。通过在异构传输环境中进行AoI优化研究，可以进一步提升LoRa网络在智能交通管理等复杂应用场景中的性能和可靠性。

参 考 文 献

[1] KAUL S, YATES R, and GRUTESER M. Real-time status: How often should one update?[C]. IEEE INFOCOM, Orlando, USA, 2012: 2731–2735. doi: 10.1109/INFCOM.2012.6195689.

[2] INOUE Y, MASUYAMA H, TAKINE T, *et al*. A general formula for the stationary distribution of the age of information and its application to single-server queues[J]. *IEEE Transactions on Information Theory*, 2019, 65(12): 8305–8324. doi: 10.1109/TIT.2019.2938171.

[3] BEDEWY A M, SUN Yin, and SHROFF N B. Minimizing the age of information through queues[J]. *IEEE Transactions on Information Theory*, 2019, 65(8): 5215–5232. doi: 10.1109/TIT.2019.2912159.

[4] HE Qing, YUAN Di, and EPHREMIDES A. Optimal link scheduling for age minimization in wireless systems[J]. *IEEE Transactions on Information Theory*, 2018, 64(7): 5381–5394. doi: 10.1109/TIT.2017.2746751.

[5] KADOTA I, SINHA A, UYSAL-BIYIKOGLU E, *et al*. Scheduling policies for minimizing age of information in broadcast wireless networks[J]. *IEEE/ACM Transactions on Networking*, 2018, 26(6): 2637–2650. doi: 10.1109/TNET.2018.2873606.

[6] WU Beining, CAI Zhengkun, WU Wei, *et al*. AoI-aware resource management for smart health via deep reinforcement learning[J]. *IEEE Access*, 2023, 11: 81180–81195. doi: 10.1109/ACCESS.2023.3299340.

[7] PENG Kai, XIAO Peiyun, WANG Shangguang, *et al*. AoI-aware partial computation offloading in IIoT with edge computing: A deep reinforcement learning based approach[J]. *IEEE Transactions on Cloud Computing*, 2023, 11(4): 3766–3777. doi: 10.1109/TCC.2023.3328614.

[8] WANG Hao, LIU Chi, YANG Haoming, *et al*. Ensuring threshold AoI for UAV-assisted mobile crowdsensing by multi-agent deep reinforcement learning with transformer[J]. *IEEE/ACM Transactions on Networking*, 2024, 32(1): 566–581. doi: 10.1109/TNET.2023.3289172.

[9] WANG Zhuoyao, XU Xiaokang, and ZHAO Jin. Spreading factor allocation and rate adaption for minimizing age of information in LoRaWAN[C]. 2022 IEEE 24th International Conference on High Performance Computing & Communications; 8th International Conference on Data Science & Systems; 20th International Conference on Smart City; 8th International Conference on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys), Hainan, China, 2022: 482–489. doi: 10.1109/HPCC-DSS-SmartCity-DependSys57074.2022.00092.

[10] CUOMO F, CAMPO M, CAPONI A, *et al*. EXPLoRa: Extending the performance of LoRa by suitable spreading factor allocations[C]. Proceedings of the 13th International Conference on Wireless and Mobile Computing, Networking and Communications, Rome, Italy, 2017: 1–8. doi: 10.1109/WiMOB.2017.8115779.

[11] GEORGIOU O and RAZA U. Low power wide area network analysis: Can LoRa scale?[J]. *IEEE Wireless Communications Letters*, 2017, 6(2): 162–165. doi: 10.1109/LWC.2016.2647247.

[12] HAMDI R, QARAQE M, and ALTHUNIBAT S. Dynamic spreading factor assignment in LoRa wireless networks[C]. 2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 2020: 1–5. doi: 10.1109/ICC40277.2020.9149243.

[13] FARHAD A, KIM D H, STHAPIT P, *et al*. Interference-aware spreading factor assignment scheme for the massive LoRaWAN network[C]. International Conference on Electronics, Information, and Communications, Auckland, New Zealand, 2019: 1–2. doi: 10.23919/ELINFOCOM.2019.8706416.

[14] FARHAD A, KIM D H, and PYUN J Y. Resource allocation to massive internet of things in LoRaWANs[J]. *Sensors*, 2020, 20(9): 2645. doi: 10.3390/s20092645.

[15] BELTRAMELLI L, MAHMOOD A, ÖSTERBERG P, *et al*. LoRa beyond aloha: An investigation of alternative random access protocols[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(5): 3544–3554. doi: 10.1109/TII.2020.2977046.

[16] WANG Jiwen, YU Jihong, CHEN Xiaoming, *et al*. Age of information for frame slotted aloha[J]. *IEEE Transactions on Communications*, 2023, 71(4): 2121–2135. doi: 10.1109/TCOMM.2023.3244214.

[17] CHEIKH I, SABIR E, AOUAMI R, *et al*. Throughput-delay tradeoffs for slotted-aloha-based LoRaWAN networks[C]. International Conference on Wireless Communications and Mobile Computing, Harbin, China, 2021: 2020–2025. doi: 10.1109/IWCMC51323.2021.9498969.

[18] HAMDI R, BACCOUR E, ERBAD A, *et al*. LoRa-RL: Deep reinforcement learning for resource management in hybrid energy LoRa wireless networks[J]. *IEEE Internet of Things Journal*, 2022, 9(9): 6458–6476. doi: 10.1109/JIOT.2021.3110996.

[19] HONG Shengguang, YAO Fang, ZHANG Fengyun, *et al*. Reinforcement learning approach for SF allocation in LoRa network[J]. *IEEE Internet of Things Journal*, 2023, 10(20): 18259–18272. doi: 10.1109/JIOT.2023.3279429.

[20] WARET A, KANEKO M, GUITTON A, *et al*. LoRa throughput analysis with imperfect spreading factor orthogonality[J]. *IEEE Wireless Communications Letters*, 2019, 8(2): 408–411. doi: 10.1109/LWC.2018.2873705.

[21] HAARNOJA T, ZHOU A, ABBEEL P, *et al*. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]. The 35th International Conference on Machine Learning, Stockholm, Sweden, 2018: 1856–1865.

程克非：男，博士生导师，研究方向为无线通信网络、云计算与大数据、嵌入式系统及应用、网络空间安全等.

陈彩蝶：女，硕士生，研究方向为LoRa物联网.

罗　佳：男，讲师，博士，研究方向为下一代无线通信网络、人工智能、区块链等.

陈前斌：男，教授，博士生导师，研究方向为个人通信、多媒体信息处理与传输、异构蜂窝网络等.

责任编辑：　马秀强

Optimizing Age of Information in LoRa Networks via Deep Reinforcement Learning

CHENG Kefei[①]　CHEN Caidie[①]　LUO Jia[①②]　CHEN Qianbin[②]

[①](*School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

[②](*Key Laboratory of Mobile Communication Technology, School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: Age of Information (AoI) quantifies information freshness, which is critical for time-sensitive Internet of Things (IoT) applications. This paper investigates AoI optimization in an LoRa network under the Slot-Aloha protocol in an intelligent transportation environment. A system model is established to characterize transmission collisions and packet waiting times. Analytical results indicate that in LoRa uplink transmission, as the number of packets increases, AoI is primarily influenced by packet collisions. To address the challenge of a large action space hindering effective solutions, this study maps the continuous action space to a discrete action space and employs the Soft Actor-Critic (SAC) algorithm for AoI optimization. Simulation results demonstrate that the SAC algorithm outperforms conventional algorithms and traditional deep reinforcement learning approaches, effectively reducing the network's average AoI.

Objective With the rapid development of intelligent transportation systems, ensuring the real-time availability and accuracy of traffic data has become essential, particularly in transmission systems for traffic monitoring cameras and related equipment. Long-range, low-power radio frequency (LoRa) networks have emerged as a key technology for sensor connectivity in intelligent transportation due to their advantages of low power consumption, wide coverage, and long-distance communication. However, in urban environments, LoRa networks are prone to frequent data collisions when multiple devices transmit simultaneously, which affects information timeliness and, consequently, the effectiveness of traffic management decisions. This study focuses on optimizing data packet timeliness in LoRa networks to enhance communication efficiency. Specifically, it aims to improve AoI under the Slotted Aloha protocol by analyzing the effects of packet collisions and over-the-air transmission time. Based on this analysis, an optimization method using deep reinforcement learning is proposed, employing the SAC algorithm to minimize AoI. The goal is to achieve lower latency and a higher data transmission success rate in an intelligent transportation environment with frequent data transmissions, thereby improving overall system performance and ensuring real-time information availability to meet the freshness requirements of intelligent transportation systems.

Method To address the requirements for information freshness in intelligent transportation scenarios, this study investigates the optimization of packet AoI in LoRa networks under the Slotted Aloha protocol. A system model is established to analyze packet collisions and over-the-air transmission time, providing theoretical support for enhancing information transmission efficiency. Given the Markovian nature of AoI evolution, the optimization problem is formulated as a Markov Decision Process (MDP) and solved using the SAC algorithm in deep reinforcement learning.

Results and Discussions The study examines AoI variations during collisions (Fig. 2) and develops a collision model for data packet transmission (Fig. 4). Simulation results indicate that the SAC algorithm outperforms the Temporal Difference (TD) algorithm and conventional methods (Fig. 6). As the number of terminals increases, the system's average AoI also increases (Fig. 7). Additionally, the variations in average AoI under different time slots for the SAC and TD3 algorithms are analyzed (Fig. 8).

Conclusions Given the limited research on AoI in LoRa networks, this study examines the AoI optimization problem in LoRa uplink packet transmission within an intelligent traffic management environment and proposes a packet collision model under the Slotted Aloha protocol. The greedy algorithm and SAC algorithm are employed for AoI optimization. Simulation results demonstrate that the greedy algorithm outperforms conventional deep reinforcement learning algorithms but remains less effective than the SAC algorithm. The SAC algorithm significantly improves AoI optimization in LoRa networks. However, this study focuses solely on AoI optimization without considering energy consumption and packet loss rate. Future research should explore the trade-offs between energy efficiency, packet loss, and AoI optimization to minimize energy consumption and data loss. Additionally, this study does not address heterogeneous network scenarios. In environments where LoRa networks coexist with other communication technologies (e.g., Wi-Fi, Bluetooth, NB-IoT), challenges related to interoperability, data consistency, and network management arise. Investigating AoI optimization in heterogeneous transmission environments could further enhance the performance and reliability of LoRa networks in complex applications such as intelligent traffic management.

Key words: Age of Information (AoI); LoRa; Soft Actor-Critic algorithm (SAC); Deep reinforcement learning; Optimization strategy