XXX Congresso de Iniciação Científica UNICAMP

Aplicação de *Reinforcement Learning* em Redes de Comunicação Sem Fio Aloha

Palavras-chave: Redes sem fio, *Reinforcement learning*, Aloha

Thiago Nascimento Silva, FEEC - UNICAMP
Prof. Paulo Cardieri, FEEC - UNICAMP

1 Introdução

Para atender as demandas atuais de serviços baseados em comunicação sem fio, o uso dos recursos rádio (os canais) precisa ser cada vez mais eficiente, e as técnicas de controle de acesso tem um papel fundamental nesse contexto. Uma técnica de controle de acesso largamente utilizada hoje em dia é o Aloha, em que os terminais compartilhando um canal iniciam suas transmissões sem qualquer verificação do estado do canal (ocupado ou livre). Essa simplicidade de operação, adequada em muitos cenários, vem às custas de um baixo desempenho e uso pouco eficiente do canal. Neste artigo, apresentamos resultados de um estudo do uso de uma técnica aprendizado de máquina, denominada *Q-Learning* combinado com Aloha, para ensinar os terminais a escolherem de forma eficiente os recursos que serão usados nas suas transmissões, visando maximizar o desempenho da rede, mas mantendo a simplicidade de operação.

2 Redes sem fio usando ALOHA

O foco deste estudo foram as redes de comunicação sem fio com terminais gerando pacotes aleatoriamente e compartilhando um certo número de canais de comunicação. Nesta situação, as técnicas de *controle de acesso ao meio* (o canal) mais adequadas são aquelas ditas *aleatórias*, em que os canais são alocados aos terminais quando solicitado, isto é, sem a reserva de canal. A técnica de controle de acesso aleatório mais simples é o Aloha [1], em que os terminais iniciam a transmissão de seus pacotes, sem qualquer verificação prévia do estado do canal (ocupado ou desocupado). Se mais de um terminal transmitir ao mesmo tempo, as transmissões podem ser mal sucedidas, e o terminais precisarão retransmitir o pacote novamente, ou o pacote será perdido.

Nesse trabalho, consideramos uma rede com acesso *slotted* Aloha, em que o tempo é dividido em *slots*, de forma que as transmissões só podem ser iniciados no início de um *slot*. A divisão do tempo em *slots* diminui a chance de colisões, quando comparado com o caso sem *slots*.

3 Aplicação de *Reinforcement Learning* no Protocolo Aloha

A simplicidade do protocolo Aloha vem às custas de um baixo desempenho, quando comparado com outras técnicas de acesso aleatório. Esse baixo desempenho, medido, por exemplo, pela probabilidade de transmissão com sucesso, deve-se ao fato de o terminal não checar o estado do canal antes de iniciar a sua transmissão. Diversos técnicas auxiliares tem sido propostas para melhorar esse desempenho, mas mantendo a simplicidade do Aloha. Uma dessas técnicas auxiliares é aquela baseada em *Reinforcement Learning* (RL) [2]. Nesse trabalho, usamos como base a estratégia de uso de RL para o *slotted* Aloha proposta por Chu et al. [3]. Essa estratégia consiste em agrupar os *slots* em quadros e usar a técnica *Q-Learning* para "ensinar" os terminais a escolherem o *slot* dentro do quadro que resultará na maior chance de transmissão bem sucedida.

3.1 *Slotted* Aloha com *Q-Learning*

Seguindo o modelo proposto por Chu et al. [3], o tempo é dividido em *slots*, que são agrupados em N *slots* para formar um quadro. No nosso estudo, cada terminal poderá transmitir em apenas um *slot* do quadro. Consideraremos uma rede composta por K terminais, todos sincronizados temporalmente e transmitindo para uma única estação central. Para cada terminal, cada *slot* é associado a uma quantidade Q que representa a preferência que o terminal terá na escolha do *slot* para transmissão. Portanto, a cada quadro t, cada terminal $k \in \{1, \cdots K\}$ atualizará o valor da quantidade Q para cada *slot* $k \in \{1, \cdots K\}$, usando

$$Q_{t+1}(k,n) = Q_t(i,n) + \alpha \left[r - Q_t(k,n) \right], \tag{1}$$

em que α é a taxa de aprendizagem e r é a recompensa. Se a transmissão no slot k foi bem sucedida, então $r = +1$; caso contrário, $r = -1$. Para os *slots* que não foram usados pelo terminal k, a recompensa será $r = 0$. Portanto, os valores de $Q(k,n)$ representam a preferência do *slot* n na escolha do *slot* para a transmissão da estação n.

3.2 Modelo de transmissão e de geração de pacotes

O sucesso da transmissão de um pacote depende da qualidade do sinal recebido, o que pode ser modelada por meio da métrica relação sinal - interferência + ruído (SINR): a transmissão será bem sucedida se a SINR do sinal recebido for acima de um dado valor. A qualidade do sinal recebido por sua vez depende de fatores como potência de transmissão, distância entre transmissor e receptor, ruído e interferência causada por transmissões concorrentes. Nesse nosso estudo, supusemos que todos os sinais transmitidos em um mesmo *slot* chegam ao receptor com a mesma qualidade, e que apenas uma das transmissões será bem sucedida, escolhida de forma aleatória. Dessa forma, poderemos avaliar mais claramente o comportamento da técnica *Q-Learning*.

É suposto também que a cada quadro um pacote chega a um terminal (para ser transmitido à estação central) com probabilidade a. Caso a transmissão de um pacote seja mal sucedida, aquele pacote é descartado. Portanto, não consideramos nesse estudo retransmissões de pacotes.

3.3 Sobre o processo de recompensa

A operação da técnica *Q-Learning* pode ser interpretada da seguinte forma: a cada quadro e para cada *slot*, o *agente* (associado a cada terminal) escolhe uma das ações: *transmitir* ou *não transmitir*. Se a transmissão for bem sucedida, o agente ganha uma recompensa positiva ($r = +1$). Por outro lado, se for mal sucedida, a recompensa é negativa ($r = -1$). Caso a ação tenha sido *não transmitir* naquele *slot*, também haverá uma recompensa negativa, porém, com uma punição menos

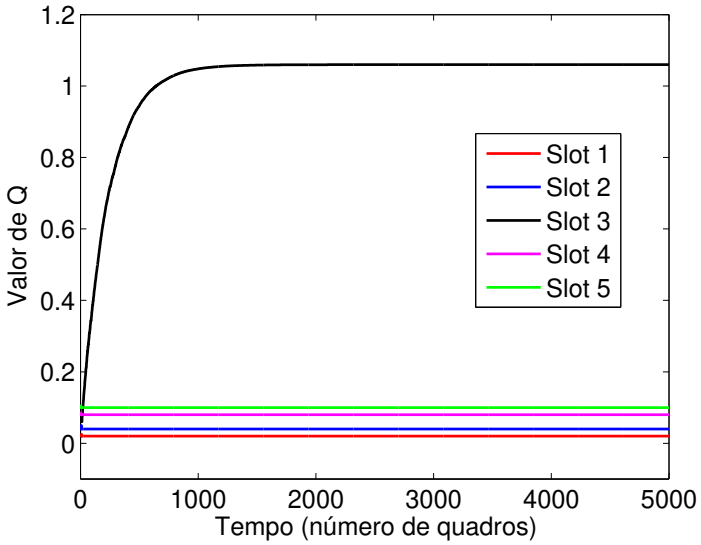

Figura 1: Comportamento dos valores de Q dos *slots* de um dos terminais, para probabilidade de chegada de pacotes $a = 0{,}9$, com $N = 5$ *slots* por quadro e $K = 5$ terminais na rede. Nota-se que o valor de Q de um dos *slots* converge para ser o máximo entre os outros valores de Q. Nota-se também que os valores de Q dos *slots* preteridos não se alteram.

severa ($r = 0$). Após exploração suficiente, os agentes terão conhecimento das ações que levam ao bom desempenho do terminal (qual *slot* usar).

4 Análise do desempenho

Nessa seção são apresentados resultados para um cenário com $N = 5$ *slots* por quadro, K terminais, e $\alpha = 0{,}005$. Os resultados foram gerados por meio de um simulador de rede sem fio, desenvolvido em linguagem Matlab.

Primeiramente, consideraremos o caso com $K = 5$ terminais, ou seja, há tantos *slots* no quadro quanto terminais. Os experimentos mostram que praticamente todos os pacotes de todos os terminais são transmitidos com sucesso, ou seja, a probabilidade de sucesso nesse caso aproximadamente unitária. De fato, o algoritmo *Q-Learning* ajusta os valores de Q de cada *slot* para cada terminal de forma que cada terminal escolhe exclusivamente um *slot* para transmitir. Esse comportamento pode ser visto na Figura 1, que mostra a evolução dos valores de Q com o passar dos quadros de um dos terminais (o mesmo comportamento é observado para os outros terminais). Portanto, os agentes associados a cada terminal são capazes de aprender coletivamente a escolherem um *slot* para cada um terminal, de forma a não causar transmissões concorrentes em um *slot*.

A situação, no entanto, muda quando temos mais terminais do que *slots* no quadro, forçando transmissões de dois ou mais terminais em um mesmo *slot*. Para avaliar esse cenário, consideramos o caso com ainda $N = 5$ *slots* por quadro, mas agora com $k = 10$ terminais na rede. A Figura 2 mostra o comportamento dos valores de Q de um dos terminais. Nota-se que agora não há uma convergência suave dos valores de Q, e o melhor *slot* para transmitir, isto é, aquele com maior valor de Q, muda com o tempo.

No entanto, mesmo sem uma convergência suave dos valores de Q, o uso da técnica *Q-Learning* melhora o desempenho da rede, como indica a Figura 3, que apresenta a probabilidade de sucesso média entre todos os terminais, para (i) uma rede operando com *slotted* Aloha e *Q-Learning* e (ii)

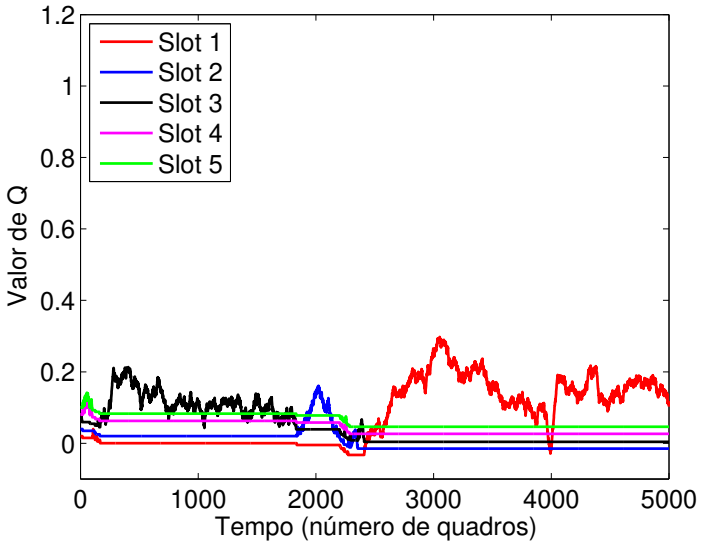

Figura 2: Comportamento dos valores de Q dos *time-slots* do usuário 1, para probabilidade de chegada de pacotes $a = 0{,}9$. Note-se que o melhor *slot* para transmitir (aquele com o maior valor de Q) muda com o tempo, e que os valores de Q dos *slots* preteridos não se alteram.

uma rede operando apenas com o *slotted* Aloha. Nesta segunda rede, o tempo também dividido em quadros de N *slots*, mas o terminais escolhem o *slot* para transmissão de forma aleatória. Nota-se na Figura 3 que o uso do algoritmo *Q-Learning* aumenta a probabilidade de sucesso, exceto para tráfego é muito baixo, quando os desempenhos das duas redes são praticamente iguis. Nesse caso, porém, o desempenho já é muito bom, pois a chance de colisões é baixa.

Os resultados de simulação mostraram também que, quando o *Q-Learning é usado* no caso com $K = 5$ e $N = 10$, a qualquer instante apenas dois terminais transmitem simultaneamente em cada *slot*. De fato, sabemos intuitivamente que essa é a melhor distribuição de terminais por *slot*, quando não há prioridade de transmissão entre os terminais.

5 Conclusões

Os resultados apresentados nesse artigo mostram a efetividade do uso da técnica *Q-Learning* para melhorar o desempenho de uma rede sem fio empregando protocolo de acesso Aloha. O técnica *Q-Learning* foi usada para ensinar os terminais a escolherem um dos *slots* disponíveis para transmissão dentro de um quadro, de forma a evitar transmissões simultâneas num mesmo *slot* de outros terminais. Como evolução desse trabalho, pretende-se investigar o cenário em que os terminais tem prioridades de transmissão diferentes. O desafio nesse cenário é evitar que os terminais de menor prioridade nunca consigam transmitir.

6 Agradecimentos

Agradecemos à Unicamp, à SAE e ao CNPq, que proporcionaram uma bolsa de iniciação para este projeto.

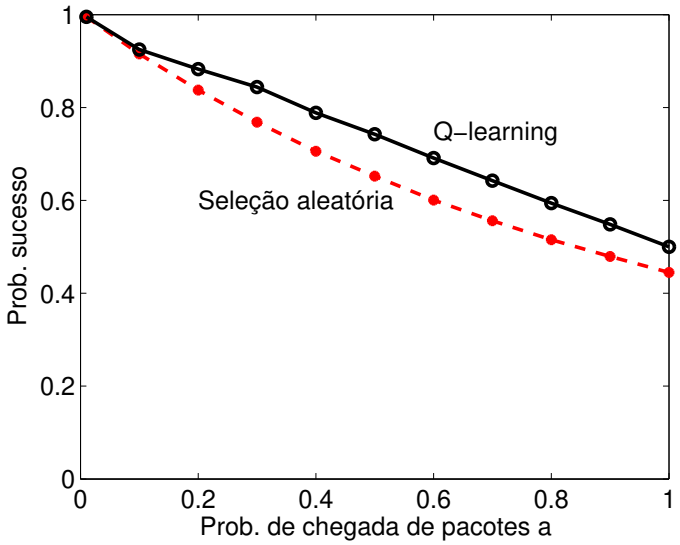

Figura 3: Probabilidade de sucesso de transmissão para os casos com seleção de *slot* aleatória e usando *Q-Learning*.

Referências

[1] R. Rom and M. Sidi, *Multiple Access Protocols: Performance and Analysis*. Berlin, Heidelberg: Springer-Verlag, 1990.

[2] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018.

[3] X. Huang, J. Jiang, S.-H. Yang, and Y. Ding, "A reinforcement learning based medium access control method for lora networks," in *2020 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, pp. 1–6, 2020.