Reinforcement Learning for Dynamic Channel Allocation in Cellular Telephone Systems

Paper by Satinder Singh, Dimitri Bertsekas
Neural Information Processing Systems 1996

Background Information about Cellular Telephone Systems

- Cellular communication is provided via local towers.
- Signals sent from your cell phone first reaches a tower and transferred to a central station afterwards.
- All connections are provided through these towers. Each tower provides connection to a certain spatial division called *cell*.
- Each cell is assigned a certain range of '*channels*' defined as a division of the total system bandwidth.
- Channels are centered around a certain frequency.
- In order to avoid multiple communications sharing the same channel, channels are distributed among the cells according to *channel reuse constraint*, which is the minimum separation distance between simultaneous reuse of the same channel.

Q: Can one channel be time shared among a number of calls?
A: Paper does not answer this question.
C: One of the telephone companies are either using time sharing or making an attempt to.

The Problem

- When a call requests a service in a given cell either it is assigned a channel or the call is 'blocked' from the system. The latter happens if no available channel can be found.
- If callers cross from one cell to another, in this case the call is 'handed off' to the cell of entry; that is a free channel is assigned to the call at the new cell. If no free channel is available the call is 'dropped' from the system.
- Objective is to minimize both dropping and blocking calls however dropping existing calls are generally more undesirable.
- Results are presented in a system where 49^{49} states exist. Therefore brute force is ruled out.

Q: In the case of 3-way calling are the callers assigned to the same channel?
A: 3-way calling is a controlled communication however if multiple calls share a channel the control component is lost.
C: Also for any call participants don't need to use the same channel. Imagine a 3-way call where one of the participants is in US, the other is in Canada and the other is in Mexico, they will be in different cells therefore not forced to use the same channel.

Fixed Assignment Algorithm

- Most popular system
- Set of Channels is partitioned into subsets.
- Each cell is *permanently* assigned a fixed subset of channels.
- Disadvantage:
  Even there are available channels in the neighboring cells a busy cell must block calls if it has no free channels.

Dynamic Channel Allocation

- More efficient than fixed assignment
- Channels are still partitioned as in Fixed Assignment Algorithm
- Channels are not assigned *permanently* therefore can be shared among cells on a need basis unless there is a channel reuse constraint violation.
- In this approach algorithms require to keep track of channel status:{occupied, unoccupied} and be able to identify the event type:{arrival, departure, handoff}

Borrowing with Directional Channel Locking

- Best existing dynamic channel allocation policy developed by Zhang & Yum in 1989.
- <Pseudo-code for the BDCL – not provided in the paper>:
  - on call *arrival* or *handoff*:
    If local channel available then use the smallest numbered one else borrow the largest numbered free channel from neighbors. If no channel is found, block the call (or drop, if handoff).
  - on call departure:
    If channel for this call was borrowed reassign the channel. If channel was not borrowed and there are other existing calls on borrowed channels then assign one of them to the released channel and return the borrowed one.
- Disadvantage:
  Cascading reassignments may occur if cells need their channels which they previously lend to their neighbors. In this case, they must borrow from other cells and in case of a departure every cell must reassign their calls.

RL Algorithm

- The system is continual and rewards are immediate
- The objective is to maximize $J = E\{\int_0^\infty e^{-\beta t} c(t)dt\}$ which assigns decisions to each state. Therefore maximizing J (higher value) is equivalent to minimizing the expected number of blocked calls over an infinite horizon.
- RL approach is used to approximate optimal value function J using Temporal difference (TD(0)) algorithm. Updated value function:
  $J_{new}(x') = (1-\alpha)J_{old}(x') + \alpha(c(x,a,\Delta t) + \gamma(\Delta t)J_{old}(y'))$
- <Pseudo-code for the RL – not provided in the paper>:
  - on call *arrival* or *handoff*:
    If local channel available then use the smallest numbered one else borrow the largest numbered free channel from neighbors with *maximum* expected reward. If no channel is found, block the call (or drop, if handoff).
  - on call departure:
    If channel for this call was borrowed reassign and return the channel with *minimum* expected penalty.

Fixed Rate Traffic Results

- <3 graphs are shown: empirical blocking probability vs. time>
- Results show that RL performs better than both BDCL and FA as the number of calls/hr is low. As the mean call arrival rate increases, the relative difference between the 3 algorithms decreases.
- Learning curve is quite rapid.
- As the number of calls goes to infinity, FA becomes optimal due to the fact that no short term fluctuations exist for RL to exploit.

Q: Although RL seems to perform better in these graphs, how do we know if it won't perform worse as the #calls increase?
A: (explained above, 4^{th} bullet)

Q: Why the empirical blocking probability is initially 0?
A: Since all channels are free initially, no blocking occurs.

Variable Traffic Results

- FA and BDCL do not consider traffic patterns therefore suffer significantly if there is a huge fluctuation in call demands.
- Word 'Non-uniform' appears on graph is used to represent the non-uniformity of call traffic across cells.
- RL adapts its policy to the particulars of the call traffic it is trained on therefore it is less sensitive to different patterns.

- Pattern examples:
  - (mmmm…m)
  - (lhllhlhlhlh)
  - (lhllhllhllhllhllh)
    l:low, m:medium, h:high

Future Work – Decentralized Training

- Policies and values are on a per-cell basis in this problem.
- Small networks pose a problem, decisions impact other cells.
- For large networks, training itself could be decentralized because the decision in a particular cell has a minor effect on far away cells.

Related Work

- Admission Control
- Predicting Temporal Differences (H0)

End Discussions

C: This approach is not used in current telephone systems although the paper is dated 1996.