Matching Queues with Abandonments in Quantum Switches: Throughput Analysis

Martin Zubeldia, Prakirt R. Jhunjhunwala, Siva Theja Maguluri

GEORGIA TECH

MOTIVATION: QUANTUM SWITCHES

- Operate in discrete time
- Maximally entangled qubits (Bell pairs) between the switch and each node are generated at a steady rate, and stored in a local quantum memory
- Stored qubits lose coherence after a random amount of time, and are lost
- Upon request, two nodes are connected by using a two-qubit Bell-state measurement between pairs of locally stored qubits

Connections are generated as matchings!

STYLIZED MODEL: THE ‘W’ TOPOLOGY

\[ \begin{array}{c}
\lambda_1 & R_1(\cdot) & Q_1(\cdot) & \gamma & \mu_1 \\
\lambda_2 & R_2(\cdot) & Q_2(\cdot) & \gamma & \mu_2 \\
\end{array} \]

- Connection requests \( R(\cdot) \): Arrive as i.i.d. process \( A(\cdot) \) of rate \( \lambda_1 \)
- Entangled qubits \( Q(\cdot) \): Generated as i.i.d. processes \( S_j(\cdot) \) of rates \( \mu_j \)
- Decoherence: Stored qubits are lost after \( \sim \text{Geo}(\gamma) \) units of time
- Matchings: Pairs of stored qubits are matched with a request

Qubits of 2nd type have a choice of which matching to make!

MARKOV CHAIN MODEL

- Discrete-time Markov chain \( (R(\cdot), Q(\cdot)) \in \mathbb{N}^2 \times \mathbb{N}^1 \)
- Recursive definition:
  \[ \begin{align*}
  R(t+1) &= R(t) + A(t) - M(t) \\
  Q(t+1) &= Q(t) - D(t) + S(t) - M(t)
  \end{align*} \]
  where
  - \( D(t) \sim \text{Ber}(Q(t), \gamma) \) are the number of abandonments
  - \( M(t) \) are the number of matchings
- Matching policies: Use a randomized control signal \( X(t) \in \{0,1\} \)
  - If \( X(t) = 0 \) \( \Rightarrow \) Connections of 1st type have preference
  - If \( X(t) = 1 \) \( \Rightarrow \) Connections of 2nd type have preference
  - All possible matchings are created ("non-idling")

MATCHINGS:

- Discrete-time Markov chain \( (R(t), Q(t)) \)
- Recursive definition:
  \[ \begin{align*}
  R(t+1) &= R(t) + A(t) - M(t) \\
  Q(t+1) &= Q(t) - D(t) + S(t) - M(t)
  \end{align*} \]
  where
  - \( D(t) \sim \text{Ber}(Q(t), \gamma) \) are the number of abandonments
  - \( M(t) \) are the number of matchings
- Matching policies: Use a randomized control signal \( X(t) \in \{0,1\} \)
  - If \( X(t) = 0 \) \( \Rightarrow \) Connections of 1st type have preference
  - If \( X(t) = 1 \) \( \Rightarrow \) Connections of 2nd type have preference
  - All possible matchings are created ("non-idling")

MAX-WEIGHT POLICY

\[ X(t) = \begin{cases} 
0, & \text{if } R(t) > R_2(t) \\
\text{Ber}(1/2), & \text{if } R(t) = R_2(t) \\
1, & \text{if } R(t) < R_2(t)
\end{cases} \]

COMpletely backlogged (\( R(\cdot) = R_2(\cdot) = \infty \))

- \( Q_2(\cdot) \) and \( Q_3(\cdot) \) behave as 2-sided queue with abandonments
- Total throughput \( C_{1,2} \)
- Throughput of type \( i \) requests \( \in \lceil C_i, \infty \rceil \)

PARTIALLY backlogged (\( R_2(\cdot) = \infty \))

- \( \lambda_2 < C_2 \Rightarrow (R(\cdot), Q(\cdot)) \) is positive recurrent
- Throughput of type \( 2 \) (backlogged) requests \( \in \lceil C_2, \infty \rceil \)

THEOREM: STABILITY REGION

\[ \lambda_1 + \lambda_2 = C_{1,2} \]

THEOREM: FLUID LIMIT

Fix \( T > 0 \), and \( r^0 \in \mathbb{R}_+^2 \) with \( \|r^0\|_1 > 0 \). If \( \|Q(0)\|_1 = 0 \) and
\[ \lim_{n \to \infty} \frac{1}{n} \left\| \sum_{t=0}^{n-1} R(t) - r^0 \right\|_1 = 0, \quad \text{a.s.,} \]
then
\[ \lim_{n \to \infty} \sup_{t \in [0,T]} \left| \frac{1}{n} \left( R(nt) - r(t) \right) \right|_1 = 0, \quad \text{a.s.,} \]
where \( r(\cdot) \) is the unique fluid solution with initial condition \( r^0 \).

TRANSIENT OF STABLE FLUID LIMITS

CONCLUSIONS

- Completely backlogged systems not representative of throughput
- Abandonments generate counter-intuitive transient behaviors