On the Quickest Flow Problem in Dynamic Networks - a Parametric Min-Cost Flow Approach

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dl.acm.org/citation.cfm?id=2722218">http://dl.acm.org/citation.cfm?id=2722218</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Society for Industrial and Applied Mathematics</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Wed Mar 16 06:58:54 EDT 2016</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/100441">http://hdl.handle.net/1721.1/100441</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
On the Quickest Flow Problem in Dynamic Networks – A Parametric Min-Cost Flow Approach∗

Maokai Lin † Patrick Jaillet ‡

October 2014

Abstract
We consider the quickest flow problem in dynamic networks with a single source s and a single sink t: given an amount of flow F, find the minimum time needed to send it from s to t, and the corresponding optimal flow over time.

We introduce new mathematical formulations and derive optimality conditions for the quickest flow problem. Based on the optimality conditions, we develop a new cost-scaling algorithm that leverages the parametric nature of the problem. The algorithm solves the quickest flow problem with integer arc costs in \(O(nm \log(n^2/m) \log(nC))\) time, where \(n, m, \) and \(C\) are the number of nodes, arcs, and the maximum arc cost, respectively.

Our algorithm runs in the same time bound as the cost-scaling algorithm by Goldberg and Tarjan [10, 11] for solving the min-cost flow problem. This result shows for the first time that the quickest flow problem can be solved within the same time bound as one of the fastest algorithms for the min-cost flow problem. As a consequence, our algorithm will remain one of the fastest unless the quickest flow problem can be shown to be simpler than the min-cost flow problem.

1 Introduction
Static network flow models have been extensively studied and widely used in the past decades to formulate many real-world problems. Many applications, however, require a time dimension to be considered: instead of a picture of the network at one instant, we need to consider a video of the network flow over a period of time. These applications arise from various fields of studies, including transportation [14], evacuation planning [23, 12], job scheduling [2], electronic communication [5, 16], network routing [18, 3, 19], and parallel computing [15]. Such a requirement leads to the dynamic network flow model [8, 4, 13]. It is also called network flow over time [6].

A generic approach for solving optimization problems in dynamic networks is due to Ford and Fulkerson [8]. They showed that one can use the time-expansion technique to convert dynamic networks with discrete time horizon into static networks, then solve the problem using algorithms developed for static networks. This approach, however, leads to pseudo-polynomial algorithms because the number of nodes and arcs in time-expanded networks are normally proportional to the time horizon \(T\), which is exponential in the input size \(\log T\).

Although the general approach fails to provide polynomial-time algorithms for every problem in dynamic networks, there are efficient algorithms for some special problems. Ford and Fulkerson [7] showed that the dynamic max-flow problem, which is to send as much dynamic flow as possible through a single-source-single-sink network in a given time \(T\), can be formulated as a min-cost flow problem in a slightly modified static network. As a result, the dynamic max-flow problem can be solved by various min-cost flow algorithms in polynomial time.

A closely related optimization problem in dynamic networks is the quickest flow problem, which is to find the minimum time \(T^*\) needed to send a given amount of dynamic flow \(F\) from a source node \(s\) to a sink node \(t\). An immediate approach is to use binary search to find the minimum time \(T\) such that the maximum amount of dynamic flow that can be sent from \(s\) to \(t\) within time \(T\) exceeds \(F\). This naive approach, however, does not automatically yield a fully polynomial-time algorithm. Burkard et al. [4] improved the naive approach and developed polynomial-time algorithms by using Newton’s method and Megiddo’s parametric search [17]. Their algorithms need to repeatedly call subroutines that solve the min-cost flow problem.

An open question remains: since the quickest flow problem is so similar to the dynamic max-flow problem, is it possible to develop an algorithm that solves the quickest flow problem within the same time bound as the min-cost flow problem? This would be impossible if we need to repeatedly solve the min-cost flow problem as a subroutine.

In this paper, we provide an affirmative answer to this question. We observe that the quickest flow problem is essentially a parametric min-cost flow problem.
By extending the cost-scaling algorithm introduced by Goldberg and Tarjan \cite{10, 11}, we design a new cost-
scaling algorithm for the quickest flow problem that runs in the same time bound as Goldberg and Tar-
jan’s algorithm. Because their algorithm remains one
of the fastest for solving the min-cost flow problem in
terms of the worst-case time bound, our result shows that
the quickest flow problem can be solved within the
same time bound as one of the best algorithms for the
min-cost flow problem. Moreover, unless the quickest
flow problem can be shown to be simpler than the min-
cost flow problem, our algorithm will remain one of the
fastest for solving the quickest flow problem.

In addition, our result shows that the preflow push
algorithm framework is well suited for parametric ex-
tensions. Gallo et al. \cite{9} showed that by extending the
preflow push algorithm for the maximum flow problem,
one could solve the parametric max-flow problem within the
same time bound as of the maximum flow prob-
lem. Our result shows that a similar extension can be
obtained for the quickest flow problem, which can be
viewed as a parametric min-cost flow problem with a
single source and a single sink.

1.1 Related Work The dynamic max-flow problem and the quickest flow problem are two fundamental
optimization problems in dynamic networks. Consider
a network \(G = (V, E)\) with a single source \(s\) and a
single sink \(t\). In this network, each arc \(e\) is associated
with an arc cost (also called transition time) \(\tau_e\) and an
arc capacity (also called maximum flow rate) \(u_e\). The
dynamic max-flow problem is to send as much flow as
possible from \(s\) to \(t\) within a given time \(T\), and the
quickest flow problem is to find the minimum time \(T^*\)
needed to send a given amount of flow \(F\) from \(s\) to \(t\).

To solve the dynamic max-flow problem, Ford and
Fulkerson \cite{7} introduced a solution form called
temporally-repeated flow. With any feasible static flow
\(x\) in a network, we can obtain a temporally-repeated
solution in the following way: Perform a flow decom-
position on flow \(x\) to obtain a set of paths \(P\) and the
corresponding flow \(x(p)\) for each path \(p \in P\). Send flow
on each path \(p \in P\) in a constant rate \(x(p)\) from time 0
to \(t(p) := \max\{T - \tau(p), 0\}\), where \(\tau(p) := \sum_{e \in p} \tau_e\) is the
cost of path \(p\).

Ford and Fulkerson \cite{7} showed that any temporally-
repeated solution corresponding to a feasible static flow
\(x\) is a feasible dynamic flow, and there exists an optimal
solution to the dynamic max-flow problem that is in
a temporally-repeated form. Using this result, they
further showed that a dynamic max-flow problem can be
formulated as a min-cost flow problem by adding an
arc \((t, s)\) with cost \(\tau_{ts} = -T\) and infinite capacity to
the original network. After solving this min-cost flow
problem, one can transform the optimal static flow into
a temporally-repeated flow, which is an optimal solution
to the original dynamic max-flow problem.

The quickest flow problem is closely related to the
dynamic max-flow problem. Burkard et al. \cite{4} showed
that the maximum amount of flow that can be sent
through a network increases with time \(T\). Thus, one
can do a binary search over time \(T\) and solve a dynamic
max-flow problem in each iteration, until the minimum
time needed to send the given amount of flow \(F\) is
found. Burkard et al. \cite{4} further used Newton’s method
to improve the naive binary search and obtained a time
bound of \(O(\log(nU)MCF(n, m))\), where \(n, m, \) and \(U\)
are the number of nodes, arcs, and the maximum arc
capacity, respectively, and \(MCF(n, m)\) is the time bound
for solving one min-cost flow problem. They also showed
that the quickest flow problem can be solved in strongly-
polynomial time using Megiddo’s parametric search \cite{17}.

1.2 Our Results In this paper, we consider the
quickest flow problem in dynamic networks with a single
source \(s\) and single sink \(t\). Given a static flow \(x\), we call a
temporally-repeated flow converted from \(x\) a temporally-
repeated flow of \(x\). By transforming the formulation by
Ford and Fulkerson \cite{7}, we show that the quickest flow
problem can be formulated with the following fractional
programming problem:

\[
T^* = \min \frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v}
\]

\[
\text{s.t. } \sum_{e \in \delta^+_u} x_e - \sum_{e \in \delta^-_u} x_e = \begin{cases} v & w = s \\ -v & w = t \\ 0 & \text{otherwise} \end{cases} \\
0 \leq x_e \leq u_e \quad \forall e \in E
\]

For any optimal solution \(x^*\) of (1.1), a temporally-
repeated flow of \(x^*\) is an optimal dynamic flow for
the quickest flow problem, and \(T^*\) is the shortest time
needed to send \(F\) amount of flow from \(s\) to \(t\).

By setting \(\theta := 1/v\) and substituting \(x_e \rightarrow x_e' = x_e/v = x_e \cdot \theta\), we can linearize the objective function of
(1.1) and obtain a linear programming problem (1.2):

\[
T^* = \min F \cdot \theta + \sum_{e \in E} \tau_e \cdot x_e
\]

\[
\text{s.t. } \sum_{e \in \delta^+_u} x_e - \sum_{e \in \delta^-_u} x_e = \begin{cases} 1 & w = s \\ -1 & w = t \\ 0 & \text{otherwise} \end{cases} \\
0 \leq x_e \leq u_e \cdot \theta \quad \forall e \in E
\]

Let \((\theta^*, x^*)\) be an optimal solution to the linear program
(1.2), then a temporally-repeated flow of the static
flow $x^*/\theta^*$ is an optimal solution to the quickest flow problem.

Observe that if we fix $v$ in (1.1), it becomes a min-cost flow problem, with supply at $s$ and demand at $t$ both equal to $v$. Therefore, the quickest flow problem is essentially a parametric min-cost flow problem with respect to $v$. Using this observation, we derive the following optimality conditions for the quickest flow problem:

**Theorem 1.1.** A temporally-repeated flow of a feasible static flow $x$ is optimal if and only if $x$ is a min-cost flow with respect to its flow value $v$, and satisfies

\[
-d_{st} \leq \frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v} \leq d_{st}
\]

where the flow value $v$ is the amount of static flow sent from node $s$ to $t$, and $d_{st}$ and $d_{ts}$ are the costs of the shortest $s-t$ path and $t-s$ path in the residual network of flow $x$, respectively.

Based on these optimality conditions, we introduce a new algorithm for solving the quickest flow problem with integer arc costs. Our algorithm adds an extra step to each scaling phase of the cost-scaling algorithm by Goldberg and Tarjan [11] for solving the min-cost flow problem, but still runs in $O(n^3 \log(nC))$ time. It can be further improved to run in $O(nm \log(n^2/m) \log(nC))$ time by applying the dynamic tree data structure developed by Sleator and Tarjan [21, 22]. Here $n$, $m$, and $C$ are the number of nodes, arcs, and the maximum arc cost, respectively.

**Organization:** In Section 2, we show that the quickest flow problem can be formulated in terms of the fractional programming problem in (1.1) and derive optimality conditions based on this formulation. In Section 3, we introduce the new cost-scaling algorithm with analysis of its correctness and time complexity. We conclude the paper in Section 4 with some final remarks.

## 2 Mathematical Models and Optimality Conditions

### 2.1 Preliminaries

**Network:** $G = (V, E)$ is a network with a set of nodes $V$ and arcs $E$. Let $n := |V|$ be the number of nodes, and $m := |E|$ be the number of arcs. We only consider networks with a single source $s \in V$ and a single sink $t \in V$. Each arc $e = (w, w') \in E$ is associated with a cost, or free-flow transition time $\tau_e \geq 0$, and a capacity, or maximum flow rate $u_e \geq 0$. In this paper, we assume $\tau_e$ and $u_e$ do not change over time. We define $\delta^+_w := \{(w, w') \in E, \forall w' \in V\}$ as the set of outgoing arcs from node $w$, and $\delta^-_w := \{(w', w) \in E, \forall w' \in V\}$ as the set of incoming arcs into node $w$.

**Feasible Flow:** If a static flow $x$ satisfies the flow conservation conditions

\[
\sum_{e \in \delta^+_w} x_e - \sum_{e \in \delta^-_w} x_e = \begin{cases} v & w = s \\ -v & w = t \\ 0 & \text{otherwise} \end{cases}
\]

where $v \geq 0$, and the capacity constraints

\[
0 \leq x_e \leq u_e \quad e \in E
\]

then we call $x$ a feasible flow. We call $v$ the flow value of $x$, and $x$ a feasible flow with flow value $v$.

**Preflow and Node Excess:** A preflow $\hat{x}$ is a static flow that satisfies the capacity constraints (2.5) but not necessarily the flow conservation conditions (2.4). For a preflow $\hat{x}$, we define flow excess $e(w)$ on a node $w \in V$ as the sum of incoming flow minus the sum of outgoing flow. Formally, we have

\[
e(w) := \sum_{e \in \delta^+_w} \hat{x}_e - \sum_{e \in \delta^-_w} \hat{x}_e \quad \forall w \in V
\]

In a preflow, $e(w) \geq 0$ for every $w \in V \setminus \{s, t\}$. In this paper, we will use $x$ to denote a preflow in order to simplify notations. If $x$ also satisfies the conservation constrains (2.4), we will emphasize that $x$ is a feasible flow.

**Flow Decomposition:** It is well known (see Ahuja et al. [1], Sect. 3.5 for example) that for any feasible flow $x$, there exists a flow decomposition $(P, x(p))$, where $P$ is a set of $s-t$-paths and cycles in $G$, and $x(p) \geq 0, \forall p \in P$ is the flow on the path or cycle $p$.

**Residual Network:** Residual networks are widely used for solving network flow problems. A residual network $G(x) = (V, E(x))$ with respect to a given flow $x$ is defined as follows: Replace each arc $e = (w, w')$ in the original network $G$ by a forward arc $e' = (w, w')$ and a backward arc $e'' = (w', w)$. The forward arc $e'$ has a cost $\tau_{e'} := \tau_e$ and residual capacity $r_{e'} := u_e - x_e$. The backward arc $e'' = (w', w)$ has a cost $\tau_{e''} := -\tau_e$ and residual capacity $r_{e''} := x_e$. In this paper, we use $\hat{\tau}$ to indicate the cost of an arc $e$ in a residual network. We have the following theorem in a residual network with respect to flow $x$:

**Theorem 2.1.** *(Theorem 9.1, Ahuja et al. [1])* If there does not exist any cycle with negative cost in the residual network $G(x)$, flow $x$ is a min-cost flow.

For further discussions of residual networks, we refer the reader to Ahuja et al. [1], Sect. 2.4.

**Node Potentials and Reduced Cost:** Node potentials arise from the dual formulation of the min-cost flow problem. They are the dual variables corresponding
to the flow conservation constraints (2.4). With a set of node potentials \( \pi \), the reduced cost of arc \( e = (w, w') \) in the residual network is defined as \( c_e' := \pi_{w'} + \tau_e - \pi_w \), where \( \tau_e \) is the cost of arc \( e \) in the residual network \( G(x) \).

Shortest Simple Path: A simple \( w \)-\( w' \) path is defined as a path from node \( w \) to \( w' \) without loop. We use \( d_{st}(x) \) to denote the cost of a shortest simple \( s \)-\( t \) path in the residual network \( G(x) \), and \( d_{se}(x) \) the cost of a shortest simple \( t \)-\( s \) path in \( G(x) \). Such costs could be either positive or negative. Note that although the cost of a shortest path between two nodes could be minus infinity in a residual network with negative cycles, the cost of a shortest \textit{simple} path is always bounded. In this paper, we use shorthands \( d_{st} \) and \( d_{se} \) if the flow they correspond to is clear in the context.

Dynamic Flow (Flow over Time) A dynamic flow consists of a set of Lebesgue-integrable functions \( f_e(t) \) that represents the rate of flow on each arc \( e \in E \) over time \( t \). The dynamic flow excess over time on node \( w \in V \) is defined as:

\[
(2.7) \quad g_w(t) := \sum_{\delta_w^+} \int_0^{t - \tau_e} f_e(s) ds - \sum_{\delta_w^-} \int_0^t f_e(s) ds
\]

A feasible flow \( f \) must satisfy the capacity constraints \( f_e(t) \leq u_e, \forall e \in E, \forall t \geq 0 \), and the node excess \( g_w(t) \) must be non-negative at all time for all nodes. See Sect. 1 of Skutella [20] for a complete discussion of the definition of dynamic flows.

Temperally-Repeated Flow: First introduced by Ford and Fulkerson [7], temporally-repeated flow is a special type of dynamic flow. It turns any feasible static flow \( x \) into a feasible dynamic flow. For any feasible static flow \( x \) and a time horizon \( T \), one can perform a flow decomposition on \( x \) and obtain a tuple \( \langle P, x(p) \rangle \), where \( P \) is a set of paths and \( x(p) \) is the flow on each path \( p \in P \). A temporally-repeated flow sends flow over each path \( p \in P \) at a constant rate \( x(p) \) from time 0 to \( t(p) := \max\{T - \tau(p), 0\} \), where \( \tau(p) := \sum_{e \in p} \tau_e \) is the cost of path \( p \). Mathematically, a temporally-repeated flow \( f \) corresponding to a flow decomposition \( \langle P, x(p) \rangle \) of a feasible static flow \( x \) and time horizon \( T \) is defined as:

\[
(2.8) \quad f_e(t) := \sum_{p \in P_e(t)} x(p), \forall e = (w, w') \in E, t \in [0, T)
\]

where \( P_e(t) := \{ p \in P : e \in p \text{ and } \tau(p_{s,w}) \leq t \text{ and } \tau(p_{w',t}) < T - t \} \). In the definition, \( \tau(p_{s,w}) \) is the sum of costs of the arcs from node \( s \) to node \( w \) on path \( p \), and \( \tau(p_{w',t}) \) is the sum of costs of the arcs from node \( w' \) to node \( t \) on path \( p \). Again, see Sect. 1 of Skutella [20] for a complete discussion.

2.2 Formulations and Optimality Conditions

Ford and Fulkerson [7] showed that for the dynamic max-flow problem, there exists an optimal solution in the temporally-repeated form. Therefore, a dynamic max-flow problem with time horizon \( T \) can be formulated as:

\[
F^*(T) = \max_{x} T \cdot v - \sum_{e \in E} \tau_e \cdot x_e
\]

\[
(2.9) \quad \text{s.t.} \quad x_e - \sum_{e \in \delta^-_w} x_e = \begin{cases} v & w = s \\ -v & w = t \\ 0 & \text{otherwise} \end{cases}
\]

For any feasible static flow \( x \) and time horizon \( T \)

\[
\sum_{e \in \delta^+_w} x_e - \sum_{e \in \delta^-_w} x_e = \begin{cases} v & w = s \\ -v & w = t \\ 0 & \text{otherwise} \end{cases}
\]

\[
0 \leq x_e \leq u_e \quad \forall e \in E
\]

Lemma 2.1. (Burkard et al. [4]) \( F^*(T) \) is non-decreasing as time \( T \) increases.

Lemma 2.1 implies that finding an optimal solution to the quickest flow problem with a given amount of flow \( F \) is equivalent to finding the minimum time \( T^* \) such that \( F^*(T^*) \geq F \). This observation is used to prove Theorem 2.2.

Theorem 2.2. The quickest flow problem can be formulated as the fractional programming problem (1.1) shown in Section 1.2.

Proof. By Lemma 2.1, the quickest flow problem can be formulated as:

\[
T^* = \min T
\]

\[
\text{s.t.} \quad T \cdot v - \sum_{e \in E} \tau_e \cdot x_e \geq F
\]

\[
(2.10) \quad \sum_{e \in \delta^+_w} x_e - \sum_{e \in \delta^-_w} x_e = \begin{cases} v & w = s \\ -v & w = t \\ 0 & \text{otherwise} \end{cases}
\]

\[
0 \leq x_e \leq u_e \quad \forall e \in E
\]

In this formulation, \( T, x, \) and \( v \) are variables. The first constraint in (2.10) can be rearranged as:

\[
(2.11) \quad T \geq \frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v}
\]

Note that \( v \neq 0 \). Because if \( v = 0 \), no flow will be sent, and the first constraint in (2.10) cannot hold. Since (2.11) is the only constraint for the variable \( T \), we can move it to the objective function and obtain the fractional programming formulation (1.1).

Note that by setting \( \theta := 1/v \) and substituting \( x_e \) with \( x'_e := x_e/v = x_e \cdot \theta \), we can linearize (1.1) and obtain the linear form (1.2).
In (1.1), if we consider \( v \) as a parameter, the objective function becomes a summation of the term \( F/v \) and \( 1/v \) times the optimal value of a min-cost flow problem. In order to use this property to prove the optimality conditions in Theorem 1.1, we first define the following function for the min-cost flow problem with flow value \( v \):

\[
g(v) = \min \sum_{e \in E} \tau_e \cdot x_e
\]

(2.12)

\[
\text{s.t. } \sum_{e \in \delta^+_w} x_e - \sum_{e \in \delta^-_w} x_e = \begin{cases} v & w = s \\ -v & w = t \\ 0 & \text{otherwise} \end{cases}
\]

\[0 \leq x_e \leq u_e \quad \forall e \in E\]

We further define function \( T^*(v) \) as:

(2.13)

\[
T^*(v) := \frac{F + g(v)}{v}
\]

Using this definition, we can express the optimal time \( T^* \) defined in (1.1) as:

(2.14)

\[
T^* = \min_{v > 0} T^*(v)
\]

We are now in a position to prove Theorem 1.1. We first show that the optimality condition (1.3) is satisfied if and only if \( v \) is a local minimum of function \( T^*(v) \). We then show that \( T^*(v) \) is unimodal, which implies that a local minimum of \( T^*(v) \) is also a global minimum. Using this result, we finally establish the optimality conditions stated in Theorem 1.1.

**Lemma 2.2.** Flow value \( v \) is a local minimum for function \( T^*(v) \) if and only if

(2.15) \[ -d_{ts} \leq T^*(v) \leq d_{st} \]

**Proof.** See Appendix A. \( \square \)

**Lemma 2.3.** Function \( T^*(v) \) is unimodal.

**Proof.** See Appendix A. \( \square \)

With Lemma 2.2 and Lemma 2.3, we can now prove Theorem 1.1 given in Section 1.2.

**Proof.** [Proof of Theorem 1.1] From Lemma 2.2 and Lemma 2.3, we know that \( v^* \) is an optimal flow value for \( T^*(v) \) if and only if \( -d_{ts} \leq T^*(v^*) \leq d_{st} \).

Note that if \( x^* \) is a feasible flow with flow value \( v^* \), then \( g(v^*) = \sum_{e \in E} \tau_e \cdot x_e^* \) if and only if \( x^* \) is a min-cost flow. Hence, \( T^*(v^*) = (F + \sum_{e \in E} \tau_e \cdot x_e^*) / v^* \) if and only if flow \( x^* \) is a min-cost flow with flow value \( v^* \).

Joining the two equivalence relationships above, we conclude that a feasible flow \( x^* \) with flow value \( v^* \) is an optimal solution to the fractional programming formulation (1.1) if and only if 1) \( x^* \) is a min-cost flow, and 2) \( -d_{ts} \leq (F + \sum_{e \in E} \tau_e \cdot x_e^*) / v^* \leq d_{st} \). Therefore a temporally-repeated flow of \( x^* \) is an optimal solution to the quickest flow problem if and only if the two optimality conditions above hold. \( \square \)

### 3 Cost-Scaling Algorithm

In this section, we introduce a new cost-scaling algorithm for solving the quickest flow problem with integer arc costs. We prove the correctness of the algorithm by showing that when it terminates, the optimality conditions in Theorem 1.1 are met. For the time complexity, we show that the algorithm runs in \( O(n^3 \log(nC)) \) time, and can be further improved to run in \( O(nm \log(n^2/m) \log(nC)) \) time with the dynamic tree data structure introduced by Sleator and Tarjan [21, 22]. Here \( n, m, \) and \( C \) are the number of nodes, arcs, and the maximum arc cost, respectively.

Recall that there are two requirements in the optimality conditions: 1) Flow \( x \) must be a min-cost flow, and 2) the inequality \( -d_{ts} \leq (F + \sum_{e \in E} \tau_e x_e) / v \leq d_{st} \) must be satisfied.

In order to find a min-cost flow, our algorithm follows the cost-scaling framework developed by Goldberg and Tarjan [10, 11]. For some \( \epsilon > 0 \), we allow the reduced cost \( c_e^\pi \) in the residual network \( G(x) \) to be negative, but require that \( c_e^\pi \geq -\epsilon \). A pair of flow \( x \) and node potentials \( \pi \) meeting this requirement is called \( \epsilon \)-optimal. We start with a big \( \epsilon \) and iteratively reduce \( \epsilon \) by half and modify \( x \) and \( \pi \) to satisfy the \( \epsilon \)-optimality. Such iterations continue until \( \epsilon \) is so small that the sum of reduced costs through any cycle in \( G(x) \) is greater than -1. Since we assume that all arc costs are integers, there must be no negative cycles in \( G(x) \). Therefore \( x \) is a min-cost flow. We refer the reader to Goldberg and Tarjan [11] and Ahuja et al. [1], Sect. 10.3 for a more comprehensive discussion.

In order to have the ratio \( r := (F + \sum_{e \in E} \tau_e x_e) / v \) between \( -d_{ts} \) and \( d_{st} \) when the algorithm ends, we add an extra step to each scaling phase. It turns out that \( \pi_s - \pi_t \) serves as a good approximation of both \( -d_{ts} \) and \( d_{st} \). Therefore, we push extra flow from \( s \) to \( t \) to guarantee that the upper bound of the gap between the ratio \( r \) and \( \pi_s - \pi_t \) is reduced by half after each iteration. At the end of the scaling phases, such gap will be less than 1. Using the assumption that all arc costs are integers, we can obtain an optimal solution by solving at most one extra max-flow problem.

The complete algorithm is described from Algorithm 1 to 5. The main procedure is Algorithm 1. It has
a main loop with a sequence of $\epsilon$-scaling phases and a final saturation subroutine. Each scaling phase consists of a Refine subroutine that modifies a $2\epsilon$-optimal flow into an $\epsilon$-optimal flow, and a Reduce-Gap subroutine that closes the gap between the ratio $r$ and $\pi_s - \pi_t$. In the subroutines, we call a node $w$ active if its flow excess $e(w) > 0$. We call an arc in $G(x)$ admissible if its reduced cost $c_{w}^{\pi} < 0$. We define the admissible network of $G(x)$ as the sub-network consisting solely of admissible arcs.

Algorithm 1 : Cost Scaling

Set $\epsilon := C$, flow $x := 0$, node potentials $\pi := 0$

while $\epsilon \geq 1/(8n)$ do

  Refine($\epsilon, x, \pi$)
  Reduce-Gap($\epsilon, x, \pi$)
  $\epsilon := \epsilon/2$
end while

Return optimal flow $x$

Algorithm 2 : Refine($\epsilon, x, \pi$)

Set $x_{e} := 0$ for all $\{e \in E : c_{e}^{\pi} > 0\}$ and $x_{e} := u_{e}$ for all $\{e \in E : c_{e}^{\pi} < 0\}$

Compute flow excess $e(w)$ for all $w \in V$; Put all nodes in a list $L$

while there exists an active node do

  Choose the first active node $w$ in $L$
  Push/Relabel($w$)
end while

Throughout the cost-scaling algorithm, a key metric that we focus on is the gap between the ratio $r$ and $\pi_s - \pi_t$. Define

$$\gamma(x, \pi) := \frac{F + \sum_{e \in E} \tau_e x_e}{v} - (\pi_s - \pi_t)$$

The main goal of the subroutine Reduce-Gap is to reduce the upper bound of the gap $\gamma(x, \pi)$ in each scaling phase, while maintaining a lower bound of $\gamma(x, \pi)$ in order to leave some buffer for future operations.

Subroutine Reduce-Gap consists of two levels of loops. The outer loop runs two main tasks before starting the inner loop: 1) Relabel the source node $s$ if all admissible arcs outgoing from $s$ are saturated, and 2) generate extra flow from node $s$. The inner loop then iteratively pushes the newly generated flow down to the sink. In order to avoid sending too much flow in one iteration, which might breach the lower bound of $\gamma(x, \pi)$, we do not relabel the source node $s$ in the inner loop, and set the excess $e(s)$ to 0 when all admissible arcs outgoing from $s$ are saturated.

Algorithm 3 : Reduce-Gap($\epsilon, x, \pi$)

Put all nodes except $t$ in a list $L$

while $\left(F + \sum_{e \in E} \tau_e x_e\right)/v - (\pi_s - \pi_t) > 7n\epsilon$ do

  Relabel $\pi_s := \pi_s + \epsilon$ if there is no outgoing admissible arc from $s$ in $G(x)$
  Set the excess of $s$ to $\epsilon(s) := \left(F + \sum_{e \in E} \tau_e x_e\right) - (\pi_s - \pi_t + 5n\epsilon) \cdot v / 6n\epsilon$

while there exists an active node in $L$ do

  Choose the first active node $w$ in $L$
  if $w$ is the source node $s$ then
    while $\epsilon(s) > 0$ and $G(x)$ contains an admissible arc $e = (s, w')$ do
      Push $\delta := \min\{\epsilon(s), r_e\}$ units of flow from $s$ to $w'$
    end while
    Set $\epsilon(s) := 0$
  else
    Push/Relabel($w$)
  end if
end while

Algorithm 4 : Push/Relabel($w$)

while $\epsilon(w) > 0$ and $G(x)$ contains a node $w'$ such that arc $e = (w, w')$ is admissible do

  Push $\delta := \min\{\epsilon(w), r_e\}$ units of flow from $w$ to $w'$
end while

if $\epsilon(w) > 0$ then

  Relabel $\pi_w := \pi_w + \epsilon$, and move $w$ to the beginning of list $L$
end if

Algorithm 5 : Saturate($x$)

Compute $d_{st}$, the cost of a shortest $s$-$t$ path in $G(x)$

if $d_{st} < \left(F + \sum_{e \in E} \tau_e x_e\right)/v$ then

  Form a sub-network $G'(x)$ that only includes arcs that lie on some shortest $s$-$t$-path in $G(x)$
  Send maximum amount of flow from $s$ to $t$ in $G'(x)$
end if
As a summary, we prove the following lemmas before showing the correctness of the algorithm and analyzing its time complexity:

1. Lemma 3.1 shows that $\gamma(x, \pi)$ does not change too much after executing the subroutine Refine.
2. Lemma 3.2 shows that $\pi_s - \pi_t$ serves as a good approximation of $-d_{ts}$ and $d_{st}$ during the execution of the subroutine Reduce-Gap.
3. Lemma 3.3 establishes the lower bound of $\gamma(x, \pi)$ in the subroutine Reduce-Gap.
4. Lemma 3.4 shows that the ratio $r$ is non-increasing after each iteration of the outer loop in the subroutine Reduce-Gap.
5. Lemma 3.5 implies that node $s$ is relabeled at least once in every two iterations of the outer loop in the subroutine Reduce-Gap.
6. Lemma 3.6 and Lemma 3.7 show that all nodes are relabeled at most $O(n)$ times in the subroutine Reduce-Gap.

In Lemma 3.1, we use $x^2$ and $\pi^2$ to denote the flow and node potentials when the 2-scaling phase ends, and $x^\epsilon$ and $\pi^\epsilon$ to denote the flow and node potentials when the subroutine Refine ends in the scaling phase.

**Lemma 3.1.** If we have

\[ 5n \cdot 2 \epsilon \leq \gamma(x^2, \pi^2) \leq 7n \cdot 2 \epsilon \]

when the 2-scaling phase ends, we have

\[ 5n \epsilon \leq \gamma(x^\epsilon, \pi^\epsilon) \leq 18n \epsilon \]

when the subroutine Refine ends in the $\epsilon$-scaling phase.

**Proof.** See Appendix B. □

The following lemmas show how the range of $\gamma(x^\epsilon, \pi^\epsilon)$ changes during the execution of subroutine Reduce-Gap. Because they all involve only $x$ and $\pi$ in the $\epsilon$-scaling phase, we drop the superscript $\epsilon$ to simplify notations.

**Lemma 3.2.** Throughout the $\epsilon$-scaling phase, we have:

\[ -d_{ts} - n \epsilon \leq \pi_s - \pi_t \leq d_{st} + n \epsilon \]

**Proof.** See Appendix B. □

In the following lemmas, we assume that $x$, $v$, and $\pi$ are the flow, flow value, and node potentials when the inner loop of the subroutine Reduce-Gap starts, respectively, and $x'$, $v'$, and $\pi'$ the flow, flow value, and node potentials when the the inner loop ends, respectively.

**Lemma 3.3.** When the inner loop of the subroutine Reduce-Gap ends, the gap $\gamma(x', \pi') \geq 5n \epsilon$.

**Proof.** See Appendix B. □

**Remark:** A direct corollary of Lemma 3.3 is that when the subroutine Reduce-Gap ends, we have $5n \epsilon \leq \gamma(x', \pi') \leq 7n \epsilon$. Therefore, if the assumption that $5n \cdot 2 \epsilon \leq \gamma(x^2, \pi^2) \leq 7n \cdot 2 \epsilon$ in Lemma 3.1 holds when the scaling phase ends, the inequality $5n \cdot \epsilon \leq \gamma(x^\epsilon, \pi^\epsilon) \leq 7n \cdot \epsilon$ will also hold when the $\epsilon$-scaling phase ends.

**Lemma 3.4.** When the inner loop of the subroutine Reduce-Gap ends, we have

\[ F + \sum_{e \in E} \tau_e \cdot x_e' \leq F + \sum_{e \in E} \tau_e \cdot x_e \]

**Proof.** See Appendix B. □

**Lemma 3.5.** If the source node $s$ is not relabeled in one iteration of the outer loop of the subroutine Reduce-Gap, either it will be relabeled in the next iteration, or we will have $\gamma(x^\epsilon, \pi^\epsilon) \leq 7n \epsilon$ and the subroutine Reduce-Gap will terminate.

**Proof.** See Appendix B. □

We are now in a position to prove the correctness of the algorithm and show that it runs in $O(n^3 \log (nC))$ time.

**Correctness:** Theorem 3.1 shows that the two optimality conditions in Theorem 1.1 are both satisfied when Algorithm 1 terminates. Therefore a temporally-repeated flow of the final static flow $x$ is an optimal solution to the quickest flow problem.

**Theorem 3.1.** When Algorithm 1 terminates, flow $x$ and node potentials $\pi$ satisfy the optimality conditions:

1) Flow $x$ is a minimum-cost flow with respect to its flow value $v$, and 2) $-d_{ts} \leq (F + \sum_{e \in E} \tau_e x_e)/v \leq d_{st}$.

**Proof.** See Appendix B. □

**Time complexity:** There are $O(\log (nC))$ cost-scaling phases. Within each cost-scaling phase, we run the subroutine Refine once and Reduce-Gap once. The Refine subroutine is the same as the wave implementation of the cost-scaling algorithm for the min-cost flow problem. Therefore it runs in $O(n^3)$ time. If we can bound the time complexity of the Reduce-Gap subroutine by $O(n^3)$ as well, we can claim that the overall time complexity of the cost-scaling loop is bounded by $O(n^3 \log (nC))$. 
To show that the subroutine Reduce-Gap indeed runs in $O(n^3)$ time, we first show that the source node $s$ is relabeled at most $O(n)$ times. Based on this result, we then prove that other nodes are also relabeled at most $O(n)$ times. These two bounds lead to three key conclusions: 1) The Reduce-Gap subroutine performs at most $O(n^2)$ relabeling operations, 2) the number of saturating pushes is bounded by $O(nm)$, and 3) the number of non-saturating pushes is bounded by $O(n^3)$. Therefore the execution time of the subroutine Reduce-Gap is bounded by $O(n^3)$.

**Lemma 3.6.** The source node $s$ is relabeled at most $11n$ times in the subroutine Reduce-Gap.

**Proof.** First, consider the first cost-scaling phase. Because $\epsilon = C$, $\pi_s - \pi_t$ can increase from 0 to at most $ne$. Thus, node $s$ can be relabeled at most $n$ times in the first scaling phase.

For all the subsequent iterations, Lemma 3.1 shows that when the subroutine Reduce-Gap starts, we have $5ne \leq \gamma(x, \pi) \leq 18ne$. Because the subroutine Reduce-Gap ends when $\gamma(x, \pi) \leq 7ne$, $\gamma(x, \pi)$ decreases by at most $18ne - 7ne = 11ne$.

Lemma 3.4 shows that the ratio $r$ does not increase, while each time $s$ is relabeled, $\pi_s - \pi_t$ increases by $\epsilon$. As a result, the gap $\gamma(x, \pi) = r - (\pi_s - \pi_t)$ decreases by at least $\epsilon$ each time $s$ is relabeled. Hence, node $s$ can be relabeled at most $11n$ times.

**Lemma 3.7.** Any node $w \in V$ is relabeled at most $13n$ times in the subroutine Reduce-Gap.

**Proof.** See Appendix B.

**Theorem 3.2.** The time complexity of the cost-scaling Algorithm 1 is $O(n^3 \log (nC))$.

**Proof.** From Lemma 3.6 and Lemma 3.7, we conclude that the total number of relabeling operations is bounded by $O(n^2)$. Further, because all nodes are relabeled at most $O(n)$ times, we can use the same analysis as in Goldberg and Tarjan [11] to obtain a $O(nm)$ bound for the number of saturating pushes.

For the number of non-saturating pushes, we note that if the subroutine Reduce-Gap does not relabel any node within $n$ consecutive node examinations, there must be no active node. We will either start a new iteration of the outer loop of Reduce-Gap, or terminate the subroutine Reduce-Gap altogether. Lemma 3.5 implies that we must relabel node $s$ at least once in two consecutive iterations of the outer loop. Thus, there is at least one node relabeling in $2n$ consecutive node examinations. Since the number of total relabeling operations is bounded by $O(n^2)$, the number of node examinations is bounded by $O(n^3)$. Because there is at most one non-saturating push per node examination, we obtain the bound $O(n^3)$ for the number of non-saturating pushes.

In conclusion, the time complexity of the cost-scaling loop is bounded by $O(n^3 \log (nC))$. The final subroutine Saturate might invoke a max-flow algorithm to saturate all shortest paths in the residual network. Since the time complexity of many well-known max-flow algorithms is bounded by $O(n^3 \log n)$, it is not a bottleneck of the algorithm. Thus, the overall time complexity of the cost-scaling algorithm for the quickest flow problem is $O(n^3 \log (nC))$.

If we apply the dynamic tree data structure introduced by Sleator and Tarjan [21, 22], the time complexity of the cost-scaling Algorithm 1 can be further improved to $O(nm \log (n^2/m) \log (nC))$.

**Theorem 3.3.** The quickest flow problem can be solved in $O(nm \log (n^2/m) \log (nC))$ time.

**Proof.** See Appendix B.

### 4 Conclusions and Future Research

In this paper, we show that the quickest flow problem in dynamic networks can be formulated as a fractional programming problem or a linear programming problem. We then derive optimality conditions that lead to a new cost-scaling algorithm for the quickest flow problem. The new algorithm follows the cost-scaling framework designed to solve the min-cost flow problem. We add an extra step in each scaling phase to guarantee that the optimality conditions are met when the algorithm ends. We show that the algorithm runs in $O(n^3 \log (nC))$ time. We can further improve the time complexity to $O(nm \log (n^2/m) \log (nC))$ by applying the dynamic tree data structure.

The result is significant because it shows for the first time that the quickest flow problem can be solved in the same time bound as one of the fastest algorithms for the min-cost flow problem. It implies that the quickest flow problem might not be strictly harder than the min-cost flow problem if measured by time complexity. As a consequence, unless the quickest flow problem could be shown to be simpler than the min-cost flow problem, the dynamic tree implementation of our algorithm will remain one of the fastest algorithms for the quickest flow problem in terms of worst-case time complexity.

We observe that the quickest flow problem is essentially a parametric flow problem. This is evident from the fractional programming formulation (1.1). The reason that the cost-scaling algorithm can solve the quickest flow problem within the same time bound as the
min-cost flow problem might be attributed to this parametric nature of the problem. Gallo et al. [9] showed that the parametric max flow problem can be solved within the same time bound as the maximum flow problem using the preflow-push algorithm. In this paper, we show that the quickest flow problem, as a parametric min-cost flow problem, can also be solved within the same time bound as the min-cost flow problem. Similar to what Gallo et al. [9] observed, the preflow-push algorithm framework exhibits great elasticity for parametric extensions.

Our new formulations and algorithms for the quickest flow problem lead to several new research opportunities. An open question is whether one can modify the cost-scaling algorithm to obtain a strongly-polynomial algorithm. If this is possible, we expect the time bound of such an algorithm to be better than the strongly-polynomial algorithm introduced by Burkard et al. [4]. Secondly, our algorithm might be extended to solve more general parametric min-cost flow problems, such as determining the optimal flow value of any \( s-t \) min-cost flow that satisfies extra side constraints. Lastly, the cost-scaling approach introduced in this paper might inspire the development of new algorithms for a more general optimization problem in dynamic networks – the quickest transshipment problem. The quickest transshipment problem is similar to the quickest flow problem except it allows multiple source nodes and multiple sink nodes with their respective supplies and demands. Hoppe and Tardos [13] introduced the only polynomial-time algorithms for this problem. Similar to the approach used by Burkard et al. [4], their algorithms also search over time \( T \) to find an optimal solution. Our cost-scaling approach provides a different angle for looking at this problem, and new algorithms might be developed to reach better time bounds.

Acknowledgment

We thank Prof. James B. Orlin for the help to simplify the proofs for the optimality conditions.

References

Therefore the change of cost is to keep the flow optimal, the extra flow needs to be cost incurred by sending the extra

\[ x \]

\[ g(A.3) \]

\[ a \] shortest path in the residual network achieve this goal, we need to send the extra flow through

\[ g \]

Increasing a flow value \( \epsilon \) means to send an extra \( \epsilon \) amount of flow back from \( s \) to \( t \) while maintaining a min-cost flow. In order to achieve this goal, we need to send the extra flow through a shortest path in the residual network \( G(x^*(v)) \), where \( x^*(v) \) is a min-cost flow with flow value \( v \). The extra cost incurred by sending the extra \( \epsilon \) amount of flow is \( \epsilon \cdot d_{st} \). Therefore we have:

\[ g(v + \epsilon) = g(v) + \epsilon \cdot d_{st} \]

(A.3)

Using (A.3), we can reduce (A.1) to:

\[ \frac{F + g(v)}{v} \leq \frac{F + g(v + \epsilon)}{v + \epsilon} = T^*(v + \epsilon) \]

(A.1)

\[ T^*(v) = \frac{F + g(v)}{v} \]

\[ F + g(v) - \epsilon \cdot d_{ts} \]

\[ \frac{F + g(v)}{v} \]

\[ d_{st} \geq \frac{F + g(v)}{v} = T^*(v) \]

(A.2)

\[ \Rightarrow \]

\[ \epsilon > 0, \quad \text{we have:} \]

\[ \text{From (A.4) and (A.6), we conclude that } v \text{ is a local minimum for } T^*(v) \text{ if and only if } -d_{ts} \leq T^*(v) \leq d_{st}. \]

\[ \square \]

Proof. [Proof of Lemma 2.3] Function \( g(v) \) defines a linear programming problem with parameter \( v \), and it is a minimization problem, therefore \( g(v) \) is convex and piecewise linear with respect to \( v \). As a result, we can express \( g(v) \) as:

\[ g(v) = \begin{cases} 
\lambda_0 \cdot v + \beta_0 & 0 = \alpha_0 \leq v < \alpha_1 \\
\lambda_1 \cdot v + \beta_1 & \alpha_1 \leq v < \alpha_2 \\
\cdots \\
\lambda_K \cdot v + \beta_K & \alpha_K \leq v < \infty 
\end{cases} \]

(A.7)

where \( K \) is the total number of break points, and \( \alpha_i \) is the \( i \)th break point. In (A.7), we have \( \lambda_0 < \lambda_1 < \cdots < \lambda_K \) and \( \beta_0 > \beta_1 > \cdots > \beta_K \).

Based on this expression, we have:

\[ T^*(v) = (F + g(v))/v = \lambda_i + (F + \beta_i)/v \]

for \( v \) between \( \alpha_i \) and \( \alpha_{i+1} \).

Let \( k \) be the smallest index of the pieces such that \( F + \beta_i < 0 \). Mathematically, we define \( k := \min_i \{ i : F + \beta_i < 0 \} \). Let \( k = K + 1 \) if \( F + \beta_i \geq 0 \) for all \( i \).

Then for all \( i < k \), we have \( F + \beta_i \geq 0 \). Consequently, \( T^*(v) = \lambda_i + (F + \beta_i)/v \) is non-increasing on each piece \( i \) if \( i < k \). On the other hand, for all \( i \geq k \), we have \( F + \beta_i < 0 \). Therefore \( T^*(v) = \lambda_i + (F + \beta_i)/v \) is monotone increasing on each piece \( i \) if \( i \geq k \).

Because function \( T^*(v) \) is continuous, we conclude that \( T^*(v) \) is non-increasing when \( v < \alpha_k \) and increasing when \( v \geq \alpha_k \). Hence, \( T^*(v) \) is unimodal. \( \square \)

\[ \sum_{e \in E} \tau_e x_e^* \leq \sum_{e \in E} \tau_e x_e \leq \sum_{e \in E} \tau_e x_e^* + \epsilon v \]

Proof. The left inequality is by definition. For the right inequality, consider an \( \epsilon \)-optimal flow \( x \) and node potentials \( \pi \). Let \( A \) be the set of arcs in the original
network whose forward arc or backward arc in the residual network \( G(x) \) is admissible. We alter the cost of each arc \( e = (w, w') \in A \) in the following way: set \( \tau'_e := \tau_w - \tau_{w'} \), \( \forall e \in A \). For the arcs not in set \( A \), we keep \( \tau'_e = \tau_e \). Note that after this change, \( c'_e \geq 0 \) for every arc \( e' \) in \( G(x) \). Therefore flow \( x \) is optimal in the network with the altered arc costs \( \tau'_e \). Mathematically, we have:

\[
\sum_{e \in E} \tau'_e x_e \leq \sum_{e \in E} \tau'_e x_e^*
\]

Due to the \( \epsilon \)-optimality of \( x \) and \( \pi \), we have \( \pi_{w'} + \tau_e - \pi_w \geq -\epsilon \) and \( \pi_{w'} - \pi_e - \pi_{w'} \geq -\epsilon \) for all \( e = (w, w') \in A \). Therefore we have

\[
-\epsilon \leq \tau_e - \tau'_e \leq \epsilon
\]

Note that \( x \) and \( x^* \) are both feasible flows with flow value \( v \), so the flow on any arc cannot exceed \( v \). That is,

\[
0 \leq x_e \leq v, \quad 0 \leq x'_e \leq v, \quad \forall e \in E
\]

Combining all the inequalities above, we have

\[
\sum_{e \in E} \tau_e x_e = \sum_{e \in A} (\tau'_e + (\tau_e - \tau'_e))x_e + \sum_{e \in A} \tau'_e x_e
\]

\[
\leq \sum_{e \in E} \tau'_e x_e^* + \sum_{e \in A} \tau_e x_e^* + \tau_e x_e^*
\]

By (B.9),

\[
= \sum_{e \in A} \tau_e (\tau_e + (\tau'_e - \tau_e))x_e
\]

\[
+ \sum_{e \in A} \tau_e x_e^* + \sum_{e \in A} (\tau_e - \tau'_e) x_e
\]

\[
= \sum_{e \in E} \tau_e x_e^* + \sum_{e \in A} (\tau_e - \tau'_e)(x_e - x_e^*)
\]

\[
\leq \sum_{e \in E} \tau_e x_e^* + \sum_{e \in A} \epsilon \cdot x_e^* . v \quad \text{by (B.10) and (B.11)}
\]

To complete the proof, we use the assumption that the admissible network of the residual network \( G(x) \) is acyclic. It implies that the number of arcs in set \( A \) cannot exceed \( n - 1 \). Combining this result with inequality (B.12), we obtain \( \sum_{e \in E} \tau e x_e \leq \sum_{e \in E} \tau e x'_e + n \epsilon \cdot v \), which is the right inequality of (B.8). □

\textbf{Proof.} [Proof of Lemma 3.1] We use \( v^{2\epsilon} \) and \( v^\epsilon \) to denote the flow value of \( x^{2\epsilon} \) and \( x^\epsilon \), respectively. Because Subroutine \( \text{Refine} \) does not send extra flow from \( s \) to \( t \), we have \( v^\epsilon = v^{2\epsilon} \).

Let \( x^\epsilon \) be a min-cost flow with flow value \( v^\epsilon \). Because the Subroutines \( \text{Reduce-Gap} \) and \( \text{Refine} \) both use the push/relabel algorithm that yields admissible networks that are acyclic (see Corollary 5.6 in [11]), by Lemma B.1, we have:

\[
\sum_{e \in E} \tau_e x_e^{2\epsilon} \leq \sum_{e \in E} \tau_e x_e^\epsilon \leq \sum_{e \in E} \tau_e x_e^\epsilon + n \cdot 2\epsilon \cdot v^{2\epsilon}
\]

and

\[
\sum_{e \in E} \tau_e x_e^\epsilon \leq \sum_{e \in E} \tau_e x_e^\epsilon \leq \sum_{e \in E} \tau_e x_e^\epsilon + n \cdot \epsilon \cdot v^\epsilon
\]

Putting (B.13) and (B.14) together, we have:

\[
\sum_{e \in E} \tau_e x_e^{2\epsilon} - n \epsilon \cdot v^\epsilon \leq \sum_{e \in E} \tau_e x_e^\epsilon \leq \sum_{e \in E} \tau_e x_e^\epsilon + 2n \epsilon \cdot v^\epsilon
\]

Moreover, Goldberg and Tarjan [11] showed that Subroutine \( \text{Refine} \) could only increase each node's potential by at most \( 3n \epsilon \) (Lemma 5.8 in [11]). That is, \( 0 \leq \pi^\epsilon - \pi^{2\epsilon} \leq 3n \epsilon \) and \( 0 \leq \pi_i^\epsilon - \pi_i^{2\epsilon} \leq 3n \epsilon \). Therefore,

\[
\pi_i^{2\epsilon} - \pi_i^\epsilon - 3n \epsilon \leq \pi_i^\epsilon - \pi_i^{2\epsilon} \leq \pi_i^{2\epsilon} - \pi_i^\epsilon + 3n \epsilon
\]

Using (3.17), (B.15) and (B.16), we have

\[
\pi_i^\epsilon - \pi_i^{2\epsilon} \geq \pi_s^{2\epsilon} - \pi_i^{2\epsilon} - 3n \epsilon \quad \text{by (B.16)}
\]

\[
\geq \frac{F + \sum_{e \in E} \tau_e x_e^{2\epsilon}}{v^{2\epsilon}} - 17n \epsilon \quad \text{by (3.17)}
\]

\[
\geq \frac{F + \sum_{e \in E} \tau_e x_e^\epsilon}{v^\epsilon} - 17n \epsilon \quad \text{by (B.15)}
\]

\[
= \frac{F + \sum_{e \in E} \tau_e x_e^\epsilon}{v^\epsilon} - 18n \epsilon
\]

and

\[
\pi_i^\epsilon - \pi_i^{2\epsilon} \leq \pi_s^{2\epsilon} - \pi_i^{2\epsilon} + 3n \epsilon \quad \text{by (B.16)}
\]

\[
\leq \frac{F + \sum_{e \in E} \tau_e x_e^{2\epsilon}}{v^{2\epsilon}} - 7n \epsilon \quad \text{by (3.17)}
\]

\[
\leq \frac{F + \sum_{e \in E} \tau_e x_e^\epsilon + 2n \epsilon \cdot v^\epsilon}{v^\epsilon} - 7n \epsilon \quad \text{by (B.15)}
\]

\[
= \frac{F + \sum_{e \in E} \tau_e x_e^\epsilon}{v^\epsilon} - 5n \epsilon
\]

Combining (B.17), (B.18), and the definition of \( \gamma(x, \pi) \) (3.16), we obtain (3.18). □

\textbf{Proof.} [Proof of Lemma 3.2] The pair of flow \( x \) and node potentials \( \pi \) is \( \epsilon \)-optimal throughout the \( \epsilon \)-scaling phase. Consider any shortest simple path \( p_{st} \) from \( s \) to \( t \) in the
residual network \( G(x) \). We have
\[
\sum_{e \in P_{st}} c_e^n = \sum_{e = (w, w')} \left( \pi_{w'} + \hat{t}_e - \pi_w \right) = \pi_t - \pi_s + \sum_{e \in P_{st}} \hat{t}_e
\]
\[
= \pi_t - \pi_s + d_{st}
\]
Since \( c_e^n \geq -\epsilon \), we have \( \pi_i - \pi_s + d_{st} = \sum_{e \in P_{st}} c_e^n \geq -n\epsilon \).
This is equivalent to the right inequality of (3.19).

Applying a similar argument on a shortest simple path \( p_{ts} \) from \( t \) to \( s \) in \( G(x) \), we can obtain the left inequality of (3.19).

We next consider the change of node potentials and the ratio \( (F + \sum_{e \in E} \tau_e \cdot x_e) / v \) during the execution of the subroutine \( \text{Reduce-Gap} \).
Let \( x, v, \pi \) be the flow, flow value, and node potentials when the inner loop of the subroutine \( \text{Reduce-Gap} \) starts, and let \( x', v', \) and \( \pi' \) be the flow, flow value, and node potentials when the the inner loop ends. Let \( \Delta v = v' - v \) be the change of the flow value. Note that because there is excess on node \( s \) when the inner loop starts, but no excess when it ends, we have \( v' \geq v \).

In order to prove Lemma 3.3, Lemma 3.4, and Lemma 3.5, we first prove the following lemma:

**Lemma B.2.** When the inner loop of Subroutine \( \text{Reduce-Gap} \) ends, we have
\[
\frac{F + \sum_{e \in E} \tau_e \cdot x_e'}{v'} \geq \frac{F + \sum_{e \in E} \tau_e \cdot x_e + (\pi_s - \pi_t - n\epsilon) \cdot \Delta v}{v + \Delta v}
\]
and
\[
\frac{F + \sum_{e \in E} \tau_e \cdot x_e'}{v'} \leq \frac{F + \sum_{e \in E} \tau_e \cdot x_e + (\pi_s - \pi_t + n\epsilon) \cdot \Delta v}{v + \Delta v}
\]

**Proof.** If we decompose the flow \( x' - x \) in the residual network \( G(x) \), we have a set \( P \) of feasible \( s-t \) paths in \( G(x) \). Let \( \tau(p) \) and \( x(p) \) be the cost and flow on each path \( p \in P \). Recall that \( d_{st}(x) \) is the cost of a shortest simple \( s-t \) path in the residual network \( G(x) \). By the right inequality of (3.19), we have
\[
\tau(p) \geq d_{st}(x) \geq \pi_s - \pi_t - n\epsilon
\]
Similarly, \( d_{ts}(x') \) is the cost of a shortest simple \( t-s \) path in \( G(x') \). Note that the reverse path of \( p \) is a feasible \( t-s \) path in the residual network \( G(x') \), therefore we have \( d_{ts}(x') \leq -\tau(p) \). Hence, by the left inequality of (3.19), we have
\[
(B.22) \quad \tau(p) \leq -d_{ts}(x') \leq \pi'_s - \pi'_t + n\epsilon
\]

Note that we do not relabel the source node \( s \) in the inner loop of Subroutine \( \text{Reduce-Gap} \), and the sink node \( t \) is never relabeled. Therefore \( \pi_s = \pi'_s \) and \( \pi_t = \pi'_t \).

Based on (B.21) and (B.22), we have
\[
\frac{F + \sum_{e \in E} \tau_e \cdot x_e'}{v'} = \frac{F + \sum_{e \in E} \tau_e \cdot x_e + \sum_{p \in P} \tau(p) \cdot x(p)}{v + \Delta v} \geq \frac{F + \sum_{e \in E} \tau_e \cdot x_e + (\pi_s - \pi_t - n\epsilon) \cdot \Delta v}{v + \Delta v}
\]
and
\[
\frac{F + \sum_{e \in E} \tau_e \cdot x_e'}{v'} = \frac{F + \sum_{e \in E} \tau_e \cdot x_e + \sum_{p \in P} \tau(p) \cdot x(p)}{v + \Delta v} \geq \frac{F + \sum_{e \in E} \tau_e \cdot x_e + (\pi_s - \pi_t + n\epsilon) \cdot \Delta v}{v + \Delta v}
\]

**Proof.** [Proof of Lemma 3.3] At the beginning of an iteration of the outer loop, we generate an extra amount of excess \( \left[ (F + \sum_{e \in E} \tau_e x_e) - (\pi_s - \pi_t + 5n\epsilon) \cdot v \right] / 6n\epsilon \) at the source node \( s \). No other excess is generated before the next iteration of the outer loop. Therefore,
\[
\Delta v \leq \frac{(F + \sum_{e \in E} \tau_e x_e) - (\pi_s - \pi_t + 5n\epsilon) \cdot v}{6n\epsilon}
\]
\[
\Rightarrow (\pi_s - \pi_t + 5n\epsilon) \cdot v + 6n\epsilon \cdot \Delta v \leq F + \sum_{e \in E} \tau_e x_e
\]
\[
\Rightarrow (\pi_s - \pi_t + 5n\epsilon) \cdot (v + \Delta v)
\]
\[
\leq F + \sum_{e \in E} \tau_e x_e + (\pi_s - \pi_t - n\epsilon) \cdot \Delta v
\]
\[
\Rightarrow \pi_s - \pi_t + 5n\epsilon \leq \frac{F + \sum_{e \in E} \tau_e \cdot x_e + (\pi_s - \pi_t - n\epsilon) \cdot \Delta v}{v + \Delta v}
\]
\[
\Rightarrow \pi_s - \pi_t \leq \frac{F + \sum_{e \in E} \tau_e \cdot x_e'}{v'} - 5n\epsilon \quad \text{by (B.19)}
\]

Because we do not relabel node \( s \) in the inner loop of Subroutine \( \text{Reduce-Gap} \), and the sink node \( t \) is never relabeled in Subroutine \( \text{Reduce-Gap} \), we have \( \pi_s = \pi'_s \), and \( \pi_t = \pi'_t \). Therefore, we have \( \gamma(x', \pi') \geq 5n\epsilon \). \( \square \)
Proof. [Proof of Lemma 3.4] By Lemma 3.3, we have $\gamma(x, \pi) \geq 5n\epsilon$, or equivalently,
\begin{equation}
\pi_s - \pi_t \leq \frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v} - 5n\epsilon
\end{equation}
Using (B.20) and (B.23), we have
\begin{align*}
F + \sum_{e \in E} \tau_e \cdot x_e &\leq \frac{F + \sum_{e \in E} \tau_e \cdot x_e + (\pi_s - \pi_t + n\epsilon) \cdot \Delta v}{v + \Delta v} \\
&\leq \frac{(F + \sum_{e \in E} \tau_e \cdot x_e) + (F + \sum_{e \in E} \tau_e \cdot x_e) \cdot \Delta v}{v + \Delta v} \\
&= \frac{(F + \sum_{e \in E} \tau_e \cdot x_e) \cdot (1 + \Delta v/v)}{v + \Delta v} \\
&= \frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v}
\end{align*}
by (B.23)
\[\Box\]

Proof. [Proof of Lemma 3.5] According to Subroutine Reduce-Gap, if there is no outgoing admissible arc from $s$ at the end of an iteration of the outer loop, $s$ will be relabeled in the next iteration. Therefore, we only need to show that if there exists an outgoing admissible arc from $s$ at the end of an iteration of the outer loop, we must have $\gamma(x', \pi') \leq 7n\epsilon$.

We first show that if there exists an outgoing admissible arc from $s$ when an iteration ends, all the extra flow generated at node $s$ at the beginning of the iteration has been sent to the sink. Mathematically, we will show that $\Delta v = [(F + \sum_{e \in E} \tau_e \cdot x_e) - (\pi_s - \pi_t + 5n\epsilon) \cdot v]/6n\epsilon$.

When $s$ is the first active node in the list $L$, we keep pushing excess $e(s)$ to other nodes until either $e(s) = 0$ or the residual network $G(x)$ does not contain any admissible arc outgoing from $s$. Note that any arc $e = (s, w)$ in $G(x)$ cannot become admissible again after it is saturated unless $s$ is relabeled. Since there is still an admissible arc outgoing from $s$ when the iteration ends, we must have $e(s) = 0$ every time we are done pushing excess from node $s$ to other nodes. Therefore, setting $e(s) := 0$ after the pushes does not reduce the overall flow excess. Moreover, we do not lose any flow excess when performing push/relabel operations on other nodes. Thus, we can claim that all extra flow generated at the beginning of a iteration has been sent to the sink when the iteration ends.

With $\Delta v = [(F + \sum_{e \in E} \tau_e \cdot x_e) - (\pi_s - \pi_t + 5n\epsilon) \cdot v]/6n\epsilon$, we have:
\[\begin{align*}
\pi_s - \pi_t + 5n\epsilon \cdot v &+ 6n\epsilon \cdot \Delta v = F + \sum_{e \in E} \tau_e \cdot x_e \\
\Rightarrow \quad (\pi_s - \pi_t + 5n\epsilon) \cdot (v + \Delta v) &\geq F + \sum_{e \in E} \tau_e \cdot x_e (\pi_s - \pi_t + n\epsilon) \cdot \Delta v \\
\Rightarrow \quad \pi_s - \pi_t + 7n\epsilon &\geq F + \sum_{e \in E} \tau_e \cdot x_e (\pi_s - \pi_t + 7n\epsilon) \cdot \Delta v \\
\Rightarrow \quad \pi_s - \pi_t &\geq \frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v} - 7n\epsilon
\end{align*}\]

Because we do not relabel node $s$ in the inner loop of Subroutine Reduce-Gap, and the sink node $t$ is never relabeled in Reduce-Gap, we have $\pi_s = \pi^*_s$, and $\pi_t = \pi^*_t$. Therefore we have $\gamma(x', \pi') \leq 7n\epsilon$. \[\Box\]

Proof. [Proof of Theorem 3.1] The first condition is satisfied because when the algorithm terminates, any cycle $C$ in the residual network has cost $\sum_{e \in C} e'_v \geq -n\epsilon > -1/8$. Since all costs are integers, there does not exist any negative cycle in the residual network. By Theorem 2.1, the flow is a min-cost flow.

Now we show that the optimality condition (1.3) is satisfied when the final subroutine Saturate ends. Assume that the flow and node potentials are $x$ and $\pi$ respectively when the last scaling phase ends, and $x^*$ and $\pi^*$ are the flow and node potentials when the final subroutine Saturate ends. The termination condition of the subroutine Reduce-Gap and Lemma 3.3 indicate that when the last scaling phase ends, we have
\begin{equation}
\pi_s - \pi_t + 5n\epsilon \leq \frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v} \leq \pi_s - \pi_t + 7n\epsilon
\end{equation}
Using the left inequality of (3.19) in Lemma 3.2, we have
\[\begin{align*}
dts - n\epsilon &\leq \pi_s - \pi_t - \frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v} - 5n\epsilon
\end{align*}\]
Or equivalently,
\begin{equation}
dts \leq \frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v} - 4n\epsilon < \frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v}
\end{equation}
Therefore, the left inequality of the optimality condition (1.3) is satisfied. For the right inequality of (1.3), we use the right inequality of (3.19) in Lemma 3.2:
\[\begin{align*}
\frac{F + \sum_{e \in E} \tau_e \cdot x_e}{v} - 7n\epsilon &\leq \pi_s - \pi_t \leq dts + n\epsilon
\end{align*}\]
Or equivalently,
\[
F + \frac{\sum_{e \in E} \tau_e x_e}{v} \leq d_{st} + 8n\epsilon
\]
When the scaling loop ends, we have \(8n\epsilon < 1\). Therefore,
\[
(B.26) \quad \frac{F + \sum_{e \in E} \tau_e x_e}{v} < d_{st} + 1
\]
If \( (F + \sum_{e \in E} \tau_e x_e) / v \leq d_{st} \), the optimality condition (1.3) is met, and we do not need to execute the step of sending extra flow in the final subroutine \textit{Saturate}.

On the other hand, if \( (F + \sum_{e \in E} \tau_e x_e) / v > d_{st} \), the subroutine \textit{Saturate} invokes a maximum flow algorithm to saturate all the paths with cost \(d_{st}\) in the residual network \(G(x)\). Note that all the extra flow we send are through the shortest \(st\)-paths whose costs are all \(d_{st}(x)\) in the residual network \(G(x^*)\). The reverse of these paths exist in the residual network \(G(x^*)\). Therefore we have \(-d_{st}(x^*) = d_{st}(x)\), because otherwise we could find a negative cycle in the residual network \(G(x)\), which contradicts the fact that \(x\) is a min-cost flow.

Assume the extra amount of flow sent from \(s\) to \(t\) is \(\Delta v\), we have
\[
\frac{F + \sum_{e \in E} \tau_e x^*_e}{v^*} = \frac{F + \sum_{e \in E} \tau_e x_e + d_{st}(x) \cdot \Delta v}{v + \Delta v} \geq d_{st}(x) = -d_{st}(x^*)
\]
On the other hand, \(d_{st}\) strictly increases when all the shortest paths are saturated. Since all costs are integer, we have \(d_{st}(x^*) \geq d_{st}(x) + 1\). From (B.26), we have:
\[
\frac{F + \sum_{e \in E} \tau_e x^*_e}{v^*} = \frac{F + \sum_{e \in E} \tau_e x_e + d_{st}(x) \cdot \Delta v}{v + \Delta v} < d_{st}(x) + 1 \leq d_{st}(x^*)
\]
Now both the left and right inequalities of the optimality condition (1.3) are satisfied, and there is no negative cycle in the residual network. Therefore any temporarily-repeated flow of the final solution \(x^*\) is an optimal solution to the quickest flow problem. \(\square\)

\textit{Proof.} [Proof of Lemma 3.7] The arguments we use here are similar to the ones used in Lemma 5.7 and 5.8 by Goldberg and Tarjan [11], where they showed that any node potential \(\pi_w\) could increase by at most \(3n\epsilon\) in the Subroutine \textit{Refine}.

Let \(x\) and \(\pi\) be the flow and node potentials when Subroutine \textit{Reduce-Gap} starts in the \(\epsilon\)-scaling phase, and \(x'\) and \(\pi'\) be the flow and node potentials at any step within Subroutine \textit{Reduce-Gap}.

Consider any node \(w\) that is to be relabeled. According to Subroutine \textit{Reduce-Gap}, \(w\) must have a positive flow access \(e(w)\), and all the excess must have came from the source node \(s\). If we perform a flow decomposition on flow \(x' - x\) in the residual network \(G(x)\), we can obtain a set of paths from \(s\) to \(t\) and from \(s\) to all the nodes with positive access, including \(w\). This implies that there exists a path \(p_{sw}\) from \(s\) to \(w\) in the residual network \(G(x)\) and its reverse path \(p'_{ws}\) from \(w\) to \(s\) in the residual network \(G(x')\). Since both \((x, \pi)\) and \((x', \pi')\) are \(\epsilon\)-optimal, we have:
\[
-n\epsilon \leq \sum_{e \in E_{p_{sw}}} c^\pi_e = \sum_{e = (w', w'') \in E_{p_{sw}}} (\pi_{w''} + \hat{\tau}_e - \pi_{w'}) = \pi_w - \pi_s + \sum_{e \in E_{p_{sw}}} \hat{\tau}_e
\]
and
\[
-n\epsilon \leq \sum_{e \in E_{p'_{ws}}} c^{\pi'}_e = \sum_{e = (w', w'') \in E_{p'_{ws}}} (\pi'_{w''} + \hat{\tau}_e - \pi'_{w'}) = \pi'_s - \pi'_w + \sum_{e \in E_{p'_{ws}}} \hat{\tau}_e
\]
where \(\hat{\tau}_e\) is the cost of arc \(e\) in the residual networks.

Because path \(p'_{ws}\) is the reverse path of \(p_{sw}\), we have \(\sum_{e \in E_{p'_{ws}}} \hat{\tau}_e = -\sum_{e \in E_{p_{sw}}} \hat{\tau}_e\). Summing (B.27) and (B.28), we obtain
\[
(B.29) \quad \pi'_w - \pi_w \leq \pi'_s - \pi_s + 2n\epsilon \leq 13n\epsilon
\]
The last inequality holds because by Lemma 3.6, the source node \(s\) is relabeled at most \(11n\) times, and each time it is relabeled, \(\pi_s\) increases by \(\epsilon\).

Since each relabeling increases the node potential of \(w\) by \(\epsilon\), we conclude that any node \(w \in V\) is relabeled at most \(13n\) times. \(\square\)

\textit{Proof.} [Proof of Theorem 3.3] We can apply the same analysis by Goldberg and Tarjan [11] here. Note that in their analysis, all computational costs are either attributed or amortized to the number of relabeling operations at each node in each scaling phase. Since we have shown that in our cost-scaling algorithm the number of relabeling operations on each node in each scaling phase is also bounded by \(O(n)\), we can apply the same analysis and show that the computational time can be improved to \(O(nm \log (n^2/m) \log (nC))\) by using the dynamic tree data structure. Since the analysis is the same as in [11], it is omitted here. \(\square\)