Better bounds for minimizing SONET ADMs

Leah Epstein* Asaf Levin †

Abstract

SONET add-drop multiplexers (ADMs) are the dominant cost factor in SONET /WDM rings. The number of SONET ADMs required by a set of traffic streams is determined by the routing and wavelength assignment of the traffic streams. Following previous work, we consider the problem where the route of each traffic stream is given as input, and we need to assign wavelengths so as to minimize the total number of used SONET ADMs. This problem is known to be NP-hard, and the best known approximation algorithm for this problem has a performance guarantee of $\frac{3}{2}$. We improve this result, and present a $\frac{98}{69} \approx 1.42029$-approximation algorithm. We also study some of the previously proposed algorithms for this problem, and give either tight or tighter analysis of their approximation ratio.

1 Introduction

WDM (Wavelength Division Multiplexing)/SONET (Synchronous Optical NETworks) rings form a very attractive network architecture that is being deployed by a growing number of telecom carriers. In this architecture each wavelength channel carries a high-speed SONET ring. The key terminating equipments are optical add-drop multiplexers (OADM) and SONET add-drop multiplexers (ADM). Each vertex is equipped with exactly one OADM. The OADM can selectively drop wavelengths at a vertex. Thus, if a wavelength does not carry any traffic from or to a vertex, its OADM allows that wavelength to optically bypass the vertex. Therefore, in each SONET ring a SONET ADM is required at a vertex if and only if it carries some traffic terminating at this vertex. In this paper we study the problem of minimizing the total cost incurred by the SONET ADMs.

Formally, we are given a set E of circular-arcs over the vertices $0, 1, \ldots, n-1$, where the vertices are ordered clockwise. A pair of arcs $(i, j), (k, l)$ is *non-intersecting* if the clockwise path along the cycle $0, 1, \ldots, n-1, 0$ that connects i to j and the clockwise path that connects k to l do not share any arc of the cycle. A set of arcs is non-intersecting if each pair of arcs from this set is non-intersecting. A feasible solution is a partition of E into non-intersecting subsets of arcs $E_1, E_2, \ldots, E_p$. The cost of E_i is the number of different vertices of the ring that are end-points of the arcs of E_i. The cost of the solution is the sum of costs of E_i for all i. The goal is to find a minimum cost feasible solution.

For an arc (i, j), we define its *length* as $\ell(i, j) = j - i \mod n$. For a subset of arcs, the length of the subset is the total length of its arcs. Throughout the paper we often use vertex numbers x where $x \geq n$ to denote the vertex $x \mod n$. We omit the $\mod$ operation to simplify notations.

A *chain* is an open directed path of length at most $n-1$, and a *cycle* is a closed directed path of length exactly n. W.l.o.g. we can assume that the arcs in each E_i form a connected component (either a chain or a cycle). This is so because if the arcs in E_i are disconnected, then we can

*School of Computer Science, The Interdisciplinary Center, Herzliya, Israel. lea@idc.ac.il. Research supported by Israel Science Foundation (grant no. 250/01).

†Faculty of Industrial Engineering and Management, The Technion, Haifa, Israel. levinas@tx.technion.ac.il.

partition E_i to its connected components without increasing its total cost. Therefore, we ask for a partition of E into cycles and (open-)chains. The cost of a feasible solution equals the sum of $|E|$ and the number of chains in the solution.

Liu, Li, Wan and Frieder [3] proved that this problem is NP-hard. They also considered a set of heuristics, and tested them empirically. Gerstel, Lin and Sasaki [2] also designed some heuristics for this problem. Wan, Calinsecu, Liu and Frieder [4] proved that any nontrivial heuristic is a 7/4-approximation algorithm. I.e., any algorithm such that none of its chains can be united to form a larger chain (a local optimum) is a 7/4-approximation algorithm. Calinsecu and Wan [1] provided a 3/2-approximation algorithm, and analyzed the worst-case performance of the previously studied heuristics. Below, we further describe the results of [1].

Let OPT be a given optimal solution to our problem with cost *opt*. Assume that for $i = 2, 3, \ldots$, OPT has CY_i cycles with i arcs, and for $i = 1, 2, \ldots$, OPT has CH_i chains with i arcs. We further assume that CY_2 is maximized among all optimal solutions, and as noted in [1] that no feasible solution can have a higher value of CY_2. For an algorithm A, we also use A to denote the cost of its returned solution. We sometimes use APX to denote the cost of a solution returned by an approximation algorithm.

A feasible solution SOL induces a partition of the arcs into an Eulerian subgraph and a set of *mega-chains* as follows: We consider the set of cycles and chains used by SOL as a set of arcs in directed auxiliary graph over $\{0, 1, \ldots, n-1\}$ where cycles are loops and a chain is a directed arc from its starting vertex to its end vertex. In this directed graph we find a maximal subgraph in which the in-degree of each vertex equals its out-degree. The remaining arcs define a minimal set of chains such that each such chain is directed from a vertex whose out-degree is greater than its in-degree, towards a vertex whose in-degree is greater than its out-degree. Each such chain in the auxiliary graph corresponds to a *mega-chain* in the original graph (by replacing each arc in the auxiliary graph by its corresponding chain). Therefore, each mega-chain is composed of chains. The remaining arcs in the original graph are the Eulerian subgraph. Note that the Eulerian subgraph need not be connected. Note that the number of mega-chains in SOL is independent of SOL, and is common to all feasible solutions.

We use the following auxiliary definition. *The deficiency of a vertex* v, $def(v)$, is defined as follows. Let $in(v)$ be the number of ingoing arcs of v, and let $out(v)$ be the number of outgoing arcs of V. Then, $def(v) = \frac{1}{2}|in(v) - out(v)|$.

We now formalize Algorithm Iterative Matching (IM) (see [1]). The algorithm maintains a set of valid chains of arcs $\mathcal{P}$ that covers E throughout its execution. Initially, $\mathcal{P}$ consists of chains each of which is an arc in E. The fit graph $\mathcal{F}(\mathcal{P})$ is defined as follows: its vertex set is $\mathcal{P}$, and two of its vertices are connected by an edge if the two corresponding chains have a common end-point, and they can be concatenated to form a valid chain. The algorithm constructs $\mathcal{F}(\mathcal{P})$, and if its edge set is not empty, then it finds a maximum matching $\mathcal{M}$ in $\mathcal{F}(\mathcal{P})$. Then, it merges each matched pair of chains of arcs in $\mathcal{M}$ into a longer chain. When the edge set of $\mathcal{F}(\mathcal{P})$ is empty, $\mathcal{P}$ is the valid chain generation that is given as output. Calinescu and Wan [1] showed that the approximation ratio of Algorithm IM is at most $\frac{5}{3}$, and provided a negative example for the algorithm that shows that its approximation ratio is at least $\frac{3}{2}$. We improve the negative examples of the algorithm by presenting an example where the approximation ratio of the algorithm is at least 1.6. For a variant PPIM of algorithm IM with a preprocessing step that removes all cycles with two arcs each, we present a negative example that shows that the approximation ratio of PPIM is at least $\frac{14}{9}$. We further show that the approximation ratio of IM is strictly less than 5/3.

Calinescu and Wan considered a variant of Algorithm IM: Algorithm Preprocessed Iterative Matching (PIM) defined as follows:

1. Preprocessing phase: repeatedly remove cycles consisting of remaining arcs until no more cycle can be obtained.

2. Matching phase: apply Algorithm IM to the arcs remaining after the first phase.

They showed that Algorithm PIM has an approximation ratio of at most $\frac{3}{2}$, and gave a negative example for PIM that shows that its approximation ratio is at least $\frac{4}{3}$. We show that the $\frac{3}{2}$ bound is tight. We also provide a better analysis of the approximation ratio of the algorithm. This improved analysis in Section 3 does not improve the worst case performance of the algorithm, but together with our Algorithm GPTS defined in Section 4 it provides Algorithm COMB, and the main result of this paper (shown in Section 5) is that Algorithm COMB is a $98/69 \approx 1.420289855$-approximation algorithm. We show that the approximation ratio of algorithm COMB is at least $\frac{4}{3}$.

Algorithm Preprocessed Cut and Merge (PCM) is defined as follows:

1. Remove all cycles of two arcs each.

2. Choose a cycle's arc $(i, i+1)$ with minimum load, and let B_i denote the subset of E that pass through $(i, i+1)$. Partition $E \setminus B_i$ into an optimal set of chains using a greedy procedure. Let $\mathcal{P}$ be the obtained chains.

3. Construct a weighted bipartite graph with sides B_i and $\mathcal{P}$ as follows: if $a \in B_i$ can be merged with $P \in \mathcal{P}$, add an edge between a and P with weight equal to the number of their common end-vertices. Find a maximum-weight matching in the resulting graph. Merge each pair of arc and a chain into a larger chain. This step is repeated until no further merging can be obtained.

Calinescu and Wan [1] proved that the approximation ratio of PCM is between 3/2 and 5/3. In this paper we close this gap, and improve the lower bound on the approximation ratio of PCM by showing that the approximation ratio of PCM is exactly 5/3.

Algorithm Preprocessed Eulerian Tour-Trail Splitting (PET-TS) is defined as follows:

1. Remove all cycles of two arcs each.

2. Eulerian tour phase: add a minimum size set of fake arcs E' to make the directed graph with arc set $E \cup E'$ Eulerian (and if it is disconnected, each connected component is Eulerian). Find an Eulerian tour in this graph, and remove all the fake arcs from this tour to obtain a set of trails.

3. Trail decomposing phase: Decompose each trail into simple paths and circuits.

4. Chain split phase: split each (open) path into valid chains by walking along the path from its first arc, and generating a valid chain whenever overlap occurs; split each invalid circuit into valid chains by walking along the circuit from each arc, generating a valid chain whenever overlap occurs, and then choose the one with the smallest number of open chains.

5. Chain merging phase: Repeatedly merge any pair of open chains into a larger valid chain until no more merging can occur.

Calinescu and Wan [1] proved that the approximation ratio of PET-TS is between 3/2 and 7/4. We narrow this gap by showing that the approximation ratio of PET-TS is at least 5/3.

The paper [1] also considered MCC-TS that is a variation of PET-TS in which the Eulerian tour phase is replaced by the following: define a weighted directed graph $H(E)$ with vertex set E

as follows. For any pair of non-intersecting arcs $e_1, e_2 \in E$, such that $e_1 = (i, j)$ and $e_2 = (k, l)$, add an arc from e_1 to e_2 and an arc from e_2 to e_1. If e_1, e_2 do not share any end-vertices, then the weights of both arcs are set to two. If $j = k$, then the weight of the arc from e_1 to e_2 is set to one, and the weight of the arc from e_2 to e_1 is set to two. Otherwise, the weight of the arc from e_1 to e_2 is set to two, and the weight of the arc from e_2 to e_1 is set to one. Now, find a minimum weight circuit cover of $H(E)$. Remove from it all the arcs of weight two to obtain a set of paths and circuits. Calinescu and Wan [1] proved that the approximation ratio of MCC-TS is between 3/2 and 8/5. We close this gap by showing that the approximation ratio of MCC-TS is exactly 14/9. If the pre-processing step of two arcs cycles removal is not performed, we show that the bound 8/5 is tight (we call this algorithm NMCC-TS).

Note that although we consider the absolute approximation ratio in this paper, all results are valid for the asymptotic approximation ratio as well. All negative examples can be easily magnified by taking multiple copies of each input arc, to form arbitrary large negative examples.

2 Negative examples

In this section we give negative examples where the approximation ratio is at least 3/2. This will show that the upper bound of 3/2 on the performance of PIM given in [1] is tight i.e. that the following theorem holds.

Theorem 1 *The approximation ratio of PIM is exactly $\frac{3}{2}$. The approximation ratio of any algorithm that removes cycles in an arbitrary order (even if it removes the two arc cycles first) and then solves the remaining instance, is at least $\frac{3}{2}$*

If we are interested in the design of a better approximation algorithm, the negative examples in this section exclude the option that a better analysis of PIM or a design of a similar algorithm that replaces the matching phase may be the answer.

Proof. We start with a very simple example showing that an algorithm which removes cycles in an arbitrary way cannot perform better than 3/2. Let $n = 3$ and the input arcs be $(0,1)$,$(0,2)$,$(1,2)$,$(1,0)$, $(2,0)$,$(2,1)$. Clearly, OPT consists of three two arc cycles which are $(i, i+1), (i+1, i)$ for $i = 0, 1, 2$, and therefore $opt = 6$. However, if the algorithm removes the cycle $(0,1), (1,2), (2,0)$, then it is left with three arcs of length $2 > n/2$ that cannot be combined. Therefore, we have $APX = 9$. This gives approximation ratio of at least 3/2.

The above input consists of two arcs cycles only. As it was already noticed in [1], it is easy to remove such cycles before processing any algorithm, and prevent the situation above. In the next example we show that even if there are no two arc cycles in the input, still an arbitrary removal procedure cannot reach smaller performance ratios. Moreover, we consider the following **exponential-time** algorithm: first, remove all cycles of two arcs, next, remove cycles one after the other until the remaining arcs do not contain a cycle, and finally solve optimally (in exponential time) the remaining arcs. We show that this algorithm has an approximation ratio of at least 3/2. Since this algorithm outperforms PIM, we conclude that it is a 3/2-approximation algorithm (however, not a polynomial-time).

For a given integer parameter $\alpha \geq 2$, consider $n = 2\alpha^2 - 4\alpha + 4$, and the arc set (with the optimal solution) is given by: for every $0 \leq i \leq \alpha - 3$ and every $0 \leq j \leq \alpha - 3$, we have the arcs $(\alpha j+i, \alpha j+i+1)$, $(\alpha j+i+1, n-\alpha i-j-5)$, $(n-\alpha i-j-5, n-\alpha i-j-4)$ and $(n-\alpha i-j-4, \alpha j+i)$. For every $0 \leq i \leq \alpha-4$, we have the arcs $(n-\alpha(i+1)-4, n-\alpha(i+1)-3)$, $(n-\alpha(i+1)-3, n-\alpha(i+1)-2)$ and $(n-\alpha(i+1)-2, n-\alpha(i+1)-4)$. For every $1 \leq j \leq \alpha-3$, we have the arcs $(\alpha j-2, \alpha j-1)$, $(\alpha j-1, \alpha j)$ and $(\alpha j, \alpha j-2)$. Finally, we have the twelve arcs of the following four triangles:

$(\alpha^2-2\alpha-2,\alpha^2-2\alpha-1,\alpha^2-2\alpha)$, $(\alpha^2-2\alpha,\alpha^2-2\alpha+1,\alpha^2-2\alpha+2)$, $(n-4,n-3,n-2)$ and $(n-2,n-1,0)$. Then, OPT has $(\alpha-2)^2$ cycles of four arcs and $2(\alpha-2)+4$ cycles of three arcs, and its total cost is exactly $4(\alpha-2)^2+3[2(\alpha-2)+4]=4\alpha^2-10\alpha+16$.

We now argue that the instance contains the arc $(t,t+1)$ for every t. The arcs $(0,1),\ldots,(\alpha^2-2\alpha-3,\alpha^2-2\alpha-2)$ are given by the arcs $(\alpha j+i,\alpha j+i+1)$ for $0\leq i\leq\alpha-3, 0\leq j\leq\alpha-3$. In this set there is a gap of two arcs every $\alpha-2$ arcs which is filled by the arcs $(\alpha j-2,\alpha j-1),(\alpha j-1,\alpha j)$ for $1\leq j\leq\alpha-3$. Similarly the arcs $(\alpha^2-2\alpha+2,\alpha^2-2\alpha+3),\ldots,(n-5,n-4)$ are given by the arcs $(n-\alpha i-j-5,n-\alpha i-j-4)$ for $0\leq i\leq\alpha-3, 0\leq j\leq\alpha-3$. In this set there is again a gap of two arcs every $\alpha-2$ arcs which is filled by the arcs $(n-\alpha(i+1)-4,n-\alpha(i+1)-3),(n-\alpha(i+1)-3,n-\alpha(i+1)-2)$ for $0\leq i\leq\alpha-4$. The remaining eight arcs $(\alpha^2-2\alpha-2,\alpha^2-2\alpha-1),\ldots,(\alpha^2-2\alpha+1,\alpha^2-2\alpha+2)$ and $(n-4,n-3),\ldots,(n-1,0)$ are given by the arcs of the last four triangles (except for the last arc of each triangle). Assume that the cycle which consists of n arcs is exactly the cycle that our algorithm removes.

Note that in the remaining arc set S each arc has length at least 4. We next show that in the optimal solution for the remaining arcs each arc consists of its own chain. To see this it is enough to show that if there is a pair of arcs in S with a common end-vertex v, then their total length is at least $n+1$ (this claim also shows that the original instance does not contain two arcs cycles). First, note that if one of the arcs incident at v occurs in one of the triangles of OPT, then its length is exactly $n-2$, the other arc has length at least 4, and therefore their total length is greater than $n+1$. Therefore, we can assume that the pair of arcs incident at v are from the four arcs cycles of OPT. Let $0\leq i,j\leq\alpha-3$:

- Assume that $v=\alpha j+i$. Then, the arcs incident at v are $(n-\alpha i-j-4,v)$ and $(v,n-\alpha(i-1)-j-5)$, and their total length is $n+\alpha-1>n$ for all values of $\alpha\geq 2$.
- Assume that $v=n-\alpha i-j-5$. Then, the arcs incident at v are $(\alpha j+i+1,v)$ and $(v,\alpha(j+1)+i)$, and their total length is $n+\alpha-1>n$ for all values of $\alpha\geq 2$.

Therefore, our optimal solution for S is a chain for each arc. Since $|S|=2(\alpha-2)^2+1[2(\alpha-2)+4]$ (S contains two arcs from each cycle of four arcs in OPT, and one arc from each triangle of OPT), we conclude that the cost of the approximation algorithm is $|E|+|S|=6(\alpha-2)^2+4[2(\alpha-2)+4]=6\alpha^2-16\alpha+24$. Therefore, the approximation ratio of the algorithm approaches 3/2 as α goes to infinity (also n grows to infinity). ■

3 A better analysis of the algorithm PIM

In this section we assume that the Preprocessing phase of Algorithm PIM first removes cycles with two arcs, and only if such cycles do not exist, other cycles are removed.

The proof of the next theorem is similar to the proof of Lemma 19 in [1].

Theorem 2 *Algorithm PIM returns a solution whose cost is at most* $1\cdot 2CY_2+\frac{4}{3}\cdot 3CY_3+\frac{7}{5}\cdot 5CY_5+1\cdot(2CH_1+3CH_2)+\frac{5}{4}(4CH_3+5CH_4)+\frac{3}{2}\cdot 4CY_4+\frac{3}{2}\left(\sum_{i=6}^{n} iCY_i+\sum_{i=5}^{n-1}(i+1)CH_i\right)$.

Proof. To prove the claim we assign the cost of the solution obtained by Algorithm PIM to the arcs, such that the following properties hold:

1. The total cost assigned to the arcs that belong to a cycle in OPT of two arcs is exactly the cost paid by OPT to this cycle, i.e. 2.

2. The total cost assigned to the arcs that belong to a cycle in OPT of three (resp. five) arcs is at most $\frac{4}{3}$ (resp. $\frac{7}{5}$) times the cost paid by OPT to this cycle, i.e. 4 (resp. 7). The total cost assigned to the arcs that belong to a cycle in OPT of four arcs or at least six arcs is at most $\frac{3}{2}$ times the cost paid by OPT to this cycle.

3. The total cost assigned to the arcs that belong to a chain in OPT of at most two arcs is exactly the cost paid by OPT to this chain. The total cost assigned to the arcs that belong to a chain in OPT of three or four arcs is at most $\frac{5}{4}$ times the cost paid by OPT to this chain. The total cost assigned to the arcs that belong to a chain in OPT of at least five arcs is at most $\frac{3}{2}$ times the cost paid by OPT to this chain.

To prove property 1, note that the preprocessing phase take out exactly all cycles of OPT of exactly two arcs, and therefore their cost in the solution obtained by PIM is exactly their cost in OPT.

We prove the other properties by considering not the solution obtained by PIM, but an alternative solution that is no better than PIM in terms of cost. We replace the solution of PIM with the solution obtained by PIM after the first iteration of the matching phase. This is clearly an upper bound on the cost of PIM. We further replace the solution by a solution that does not create an optimal matching, but some feasible matching that we construct. The matching is constructed by uniting matchings that may be created from the remaining arcs (after cycle removal by PIM) from each component of OPT (cycle or chain) separately. Note that a remaining path of s arcs contributes a matching of size $\lfloor s/2 \rfloor$.

We now prove property 2. Each cycle of OPT loses at least one arc in the preprocessing phase. Consider a cycle of OPT which contains k arcs ($k \geq 3$). Let $\ell \geq 1$ be the number of arcs that the cycle loses in the preprocessing phase. There is a matching of the arcs from this cycle of size $\lfloor \frac{k-1}{2} \rfloor - (l-1)$ (after the first arc from this cycle is removed, we have a matching of size $\lfloor \frac{k-1}{2} \rfloor$. Every other arc that is removed destroys at most one arc in this matching). The cost of the cycle depends on the number of chains created from it. This number is at most $\lceil \frac{k-1}{2} \rceil \leq \frac{k}{2}$. I.e., for three arcs cycles we have at most one chain, and five arcs cycles we have at most two chains. This gives the costs 4 and 7 for cycles of three and five arcs (respectively). For a cycle of k arcs we get at most $k/2$ chains, which proves the case of four arc cycles, and cycles of at least six arcs.

Next, we prove property 3. For a chain in OPT of k arcs such that l arcs are taken out in the preprocessing phase, there is a matching of the arcs from this chain of size $\lfloor \frac{k}{2} \rfloor - l$ (before the preprocessing phase we have a matching of size $\lfloor \frac{k}{2} \rfloor$, and every arc that is removed, destroys at most one arc in the matching). Therefore, the number of chains created after the first matching step equals the number of chains OPT has if the number of arcs in the chain is one or two. If the number of arcs in the chain is three or four, then the number of chains after the first matching phase is at most two (whereas OPT has one chain). Therefore, in this case we have a ratio of at most $\frac{5}{4}$ between the cost OPT pays for this chain and the cost PIM pays for this chain. For longer chains, we get at most $\lceil \frac{k}{2} \rceil - 1$ new chains (compared to OPT), and so PIM pays at most $\frac{3}{2}$ times the cost OPT pays.

Taking the optimal matching instead of the matching we describe above, only decreases the cost of the solution, and therefore the claim holds also for the solution obtained by PIM. ■

4 Algorithm GPTS

In this section we study a different approximation algorithm for the problem. Given a set of input arcs we apply a certain greedy clean-up preprocessing phase that is composed of six steps. To be

able to analyze the algorithm, we need to know the exact number of mega-chains in OPT which consist of a single arc. Since we do not have this information, we apply the algorithm for every possible such value (between 0 and $|E|$), and choose the best solution we get. Therefore, in the analysis, we can assume that this number, which we denote MC_1, is known. We denote by L_1 the total length of mega-chains in OPT which consist of a single arc. First, we remove all cycles of two arcs (the number of the cycles that we remove in this step is exactly CY_2). Then, we remove a set of MC_1 mega-chains each of them has a single arc so that the total length of the arcs that we remove in this step is maximized. Then, we remove greedily certain subgraphs (cycles or chains with a certain number of arcs and certain length), by removing each time a single such subgraph as long as the remaining graph contains a subgraph with the desired property. E.g., in Step 3 we remove cycles of three arcs each one at a time until the remaining arc-set does not contain a cycle with exactly three arcs. W.l.o.g. we assume that n is divisible by 4 (otherwise, we can add dummy vertices to the ring such that none of the traffic streams terminate at these vertices).

Algorithm Greedy-Preprocessing Trail-Split (GPTS):

1. Remove all cycles of two arcs.
2. Construct the following bipartite graph $B = (R_B, L_B, E_B)$: The right hand side, R_B, contains the set of vertices whose in-degree in (V, E) is greater than its out-degree, and the left hand side L_B contains the set of vertices whose out-degree in (V, E) is greater than its in-degree. For an arc $(u, v) \in E$, such that both $u \in R_B$ and $v \in L_B$, we add an edge to E_B between the two corresponding vertices. The weight of an edge is simply the length of the corresponding arc. Among all possible b-matchings of cardinality MC_1, we find a maximum weight b-matching in B where the degree bound of a vertex u is twice the deficiency of its corresponding vertex in (V, E). For each edge in the optimal b-matching, we remove the arc between its corresponding vertices.
3. Remove greedily cycles of three arcs until there are no such cycles.
4. Remove greedily cycles of four arcs until there are no such cycles.
5. Remove greedily mega-chains of exactly two arcs with length in the intervals $[\frac{3n}{4}, n-1]$ and $[\frac{5n}{4}, 2n-1]$ until there are no such mega-chains.
6. Remove greedily mega-chains of exactly three arcs with length in the intervals $[\frac{7n}{4}, 2n-1]$ and $[\frac{5n}{2}, 3n-1]$ until there are no such mega-chains.
7. Cover the rest of the arcs with chains in the following way:
   (a) Find a set of mega-chains (with arbitrary lengths) that connect the vertices whose out degree is greater than its in-degree to vertices whose in-degree is greater than its out-degree, and remove them. For each such mega-chain of length greater than n, decompose it into chains of length at most n.
   (b) Partition the rest of the arcs (these are from the Eulerian subgraph) into chains (or cycles) of length at most n.

Observation 3 *Consider a mega-chain with length in the interval* $[kn, (k+1)n-1]$ *that is created in step 7a. Then, the number of chains that result from it is at most* $2k+1$.

Proof. We perform a greedy partition that chooses a maximum length chain at each step. Therefore, the total length of each pair of consecutive chains in a mega-chain is at least n (otherwise, they can be united). ■

Observation 4 *The chains obtained in step 7b have an average length of at least* $\frac{n}{2}$.

Proof. The claim follows because the total length of each pair of consecutive chains resulting from the Eulerian subgraph is at least n (otherwise, they can be united). ■

Notations: consider OPT. Partition it into mega-chains and an Eulerian subgraph. There may be several options to do that, therefore we fix an arbitrary partition. Assume that OPT has exactly MC_1 mega-chains each of them has a single arc. Denote by:

- CY - the number of cycles in OPT that contain at least five arcs.
- MC - the total number of mega-chains.
- MC_2^1 - the number of mega-chains of two arcs with length at most $\frac{3n}{4} - 1$.
- MC_2^2 - the number of mega-chains of two arcs with length in the interval $[\frac{3n}{4}, n-1]$.
- MC_2^3 - the number of mega-chains of two arcs with length in the interval $[n+1, \frac{5n}{4} - 1]$.
- MC_2^4 - the number of mega-chains of two arcs with length at least $\frac{5n}{4}$.
- MC_3^1 - the number of mega-chains of three arcs with length at most $n-1$.
- MC_3^2 - the number of mega-chains of three arcs with length in the interval $[n+1, \frac{7n}{4}]$.
- MC_3^3 - the number of mega-chains of three arcs with length in the interval $[\frac{7n}{4}, 2n-1]$.
- MC_3^4 - the number of mega-chains of three arcs with length in the interval $[2n+1, \frac{5n}{2} - 1]$.
- MC_3^5 - the number of mega-chains of three arcs with length at least $\frac{5n}{2}$.
- MC^i - for $i \geq 1$ the number of mega-chains with at least four arcs and length in the interval $[(i-1)n+1, in-1]$.
- CH_E^i - the number of chains in OPT of exactly i arcs that belong to the Eulerian subgraph of OPT.

Note that a mega-chain in OPT with total length in the interval $[(i-1)n+1, in-1]$ consists of at least i chains (in OPT).

The total length of all arcs is at most

$$\begin{aligned} UB_L &= (CY_2 + CY_3 + CY_4 + CY)n + L_1 + \sum_{i=1}^{\infty} \left(CH_E^i \cdot n\right) \\ &+ MC_2^1 \cdot \frac{3n}{4} + MC_2^2 \cdot n + MC_2^3 \cdot \frac{5n}{4} + MC_2^4 \cdot 2n \\ &+ MC_3^1 \cdot n + MC_3^2 \cdot \frac{7n}{4} + MC_3^3 \cdot 2n + MC_3^4 \cdot \frac{5n}{2} + MC_3^5 \cdot 3n + \sum_{i=1}^{\infty} \left(MC^i \cdot in\right). \end{aligned}$$

Consider Algorithm GPTS, and denote by $A = CY_2$ the number of cycles removed in step 1, $B = MC_1$ - the number of mega-chains removed in step 2, and by L_B their total length, C - the number of cycles removed in step 3, D - the number of cycles removed in step 4, F - the number of mega-chains removed in step 5 and G - the number of mega-chains removed in step 6.

Then, by the optimality of step 2, we conclude that $L_B \geq L_1$. We next argue that $B+3C \geq CY_3$. To see this inequality note that we remove three-arc cycles as long as such a cycle exists, and

therefore as long as there is a three-arc cycle of OPT such that none of the arcs of the cycle is removed, this step is not completed. Since each mega-chain that we remove in step 2, removes one arc and therefore can destroy at most one cycle of OPT with three arcs, and each cycle that we remove in step 3 removes three arcs and therefore can destroy at most three cycle of OPT, we conclude that $B + 3C \geq CY_3$. Similarly $B + 3C + 4D \geq CY_3 + CY_4$ as each arc that we remove in steps 2, 3 and 4 destroy at most one cycle of OPT of at most four arcs, and we continue to remove cycles as long as there is a cycle with at most four arcs.

Next, we argue that $3B+3C+4D+4F \geq CY_3+CY_4+MC_2^2+MC_2^4$. To see this constraint, note that each arc that we remove during the preprocessing can destroy at most one structure of OPT (where a structure of OPT is one of the following. A cycle of at most four arcs, or a mega-chain of two arcs with length either in $[\frac{3n}{4}, n-1]$ or at least $\frac{5n}{4}$). In addition to this each mega-chain that we remove in steps 2 and 5 can destroy the deficiency of its end-vertices and therefore decrease the number of mega-chains that we remove in step 5 by additional two mega-chains (i.e., a total decrease of three mega-chains for a mega-chain that we remove in step 2 and four mega-chains for a mega-chain that we remove in step 5).

Last, similarly to the previous inequality we have $3B + 3C + 4D + 4F + 5G \geq CY_3 + CY_4 + MC_2^2 + MC_2^4 + MC_3^3 + MC_3^5$. This holds since a mega-chain removed in the last step can destroy at most five other mega-chains that could be removed at this step, either by decreasing deficiency or by using one of its arcs. The other amounts calculated in the previous case still apply here.

We next consider a linear combination of the last four inequalities with non-negative coefficients $\alpha, \beta, \gamma, \delta \geq 0$, that satisfies the following inequalities: $\alpha + \beta + \gamma + \delta \leq \frac{2}{3}$, $\beta + \gamma + \delta \leq \frac{1}{2}$, $\gamma + \delta \leq \frac{3}{8}$, $\delta \leq \frac{3}{10}$. We obtain the following inequality.

$$\begin{aligned}&\alpha \cdot (B + 3C) + \beta \cdot (B + 3C + 4D) + \gamma \cdot (3B + 3C + 4D + 4F) + \delta \cdot (3B + 3C + 4D + 4F + 5G)\\ &\quad \geq CY_3 \cdot (\alpha + \beta + \gamma + \delta) + CY_4 \cdot (\beta + \gamma + \delta) + (MC_2^2 + MC_2^4) \cdot (\gamma + \delta) + (MC_3^3 + MC_3^5) \cdot \delta \ .\end{aligned}$$

The left hand size of the above expression is

$$\begin{aligned}&\alpha \cdot (B + 3C) + \beta \cdot (B + 3C + 4D) + \gamma \cdot (3B + 3C + 4D + 4F) + \delta \cdot (3B + 3C + 4D + 4F + 5G)\\ &\quad = B \cdot (\alpha + \beta + 3\gamma + 3\delta) + 3C \cdot (\alpha + \beta + \gamma + \delta) + 4D \cdot (\beta + \gamma + \delta) + 4F \cdot (\gamma + \delta) + 5G \cdot \delta \ .\end{aligned}$$

Therefore, we obtain the following inequality, using the restrictions on $\alpha, \beta, \gamma, \delta$, and $MC_1 = B$,

$$\begin{aligned}&MC_1 \cdot (\alpha + \beta + 3\gamma + 3\delta) + 2C + 2D + \frac{3}{2}F + \frac{3}{2}G\\ \geq\ & CY_3 \cdot (\alpha + \beta + \gamma + \delta) + CY_4 \cdot (\beta + \gamma + \delta) + (MC_2^2 + MC_2^4) \cdot (\gamma + \delta) + (MC_3^3 + MC_3^5) \cdot \delta \ .\end{aligned}$$

Let F_s be the number of mega-chains of two arcs with length in $[\frac{3n}{4}, n-1]$ removed by the algorithm and F_ℓ the number of mega-chains of two arcs with length in $[\frac{5n}{4}, 2n-1]$ removed by the algorithm. We have $F = F_s + F_\ell$.

Let G_s be the number of mega-chains of three arcs with length in $[\frac{7n}{4}, 2n-1]$ removed by the algorithm and G_ℓ the number of mega-chains of three arcs with length in $[\frac{5n}{2}, 3n-1]$ removed by the algorithm. We have $G = G_s + G_\ell$.

Denote by UB'_L the total length of the arcs that remain at the end of step 5. By Observations 3 and 4, the total number of chains obtained by our algorithm is at most $\frac{UB'_L}{n/2} + MC + F_\ell + 2G_s + 2G_\ell$ (note that this includes also the chains from step 5 and step 6).

We next want to upper bound $UB'_L + \frac{nF_\ell}{2} + Gn$. Since we have an upper bound on UB_L, we can get this from a lower bound on the length of removed structures, subtracting the term $\frac{nF_\ell}{2} + Gn$.

The result is at least $An + L_B + Cn + Dn + F_\ell \cdot \frac{5n}{4} + F_s \cdot \frac{3n}{4} + G_\ell \cdot \frac{5n}{2} + G_s \cdot \frac{7n}{4} - \frac{nF_\ell}{2} - Gn \geq CY_2 \cdot n + L_1 + C \cdot n + D \cdot n + F \cdot \frac{3n}{4} + G \cdot \frac{3n}{4}$.

Therefore,

$$
\begin{aligned}
& UB'_L \leq UB_L - (CY_2 \cdot n + L_1 + C \cdot n + D \cdot n + F \cdot \frac{3n}{4} + G \cdot \frac{3n}{4}) \\
\leq \quad & (CY_3 + CY_4 + CY)n + \sum_{i=1}^{\infty} \left(CH_E^i \cdot n\right) \\
+ \quad & MC_2^1 \cdot \frac{3n}{4} + MC_2^2 \cdot n + MC_2^3 \cdot \frac{5n}{4} + MC_2^4 \cdot 2n \\
+ \quad & MC_3^1 \cdot n + MC_3^2 \cdot \frac{7n}{4} + MC_3^3 \cdot 2n + MC_3^4 \cdot \frac{5n}{2} + MC_3^5 \cdot 3n + \sum_{i=1}^{\infty} \left(MC^i \cdot in\right) \\
- \quad & \frac{n}{2} CY_3 \cdot (\alpha + \beta + \gamma + \delta) - \frac{n}{2} CY_4 \cdot (\beta + \gamma + \delta) \\
+ \quad & -\frac{n}{2}(MC_2^2 + MC_2^4) \cdot (\gamma + \delta) - \frac{n}{2}(MC_3^3 + MC_3^5) \cdot \delta + \frac{n}{2} MC_1 \cdot (\alpha + \beta + 3\gamma + 3\delta) \\
\leq \quad & n \cdot CY_3(1 - \frac{\alpha}{2} - \frac{\beta}{2} - \frac{\gamma}{2} - \frac{\delta}{2}) + n \cdot CY_4(1 - \frac{\beta}{2} - \frac{\gamma}{2} - \frac{\delta}{2}) + n \cdot CY + \sum_{i=1}^{\infty} \left(CH_E^i \cdot n\right) \\
+ \quad & MC_2^1 \cdot \frac{3n}{4} + (1 - \frac{\gamma}{2} - \frac{\delta}{2}) MC_2^2 \cdot n + MC_2^3 \cdot \frac{5n}{4} + (2 - \frac{\gamma}{2} - \frac{\delta}{2}) MC_2^4 \cdot n \\
+ \quad & MC_3^1 \cdot n + MC_3^2 \cdot \frac{7n}{4} + (2 - \frac{\delta}{2}) MC_3^3 \cdot n + MC_3^4 \cdot \frac{5n}{2} + (3 - \frac{\delta}{2}) MC_3^5 \cdot n + \sum_{i=1}^{\infty} \left(MC^i \cdot in\right) \\
& + \frac{n}{2} MC_1 \cdot (\alpha + \beta + 3\gamma + 3\delta)
\end{aligned}
$$

Therefore, we obtained the following corollary:

Corollary 5 *The total cost of the solution returned by Algorithm GPTS is at most* $|E| + MC + CY_3 \cdot (2 - \alpha - \beta - \gamma - \delta) + CY_4 \cdot (2 - \beta - \gamma - \delta) + 2CY + 2 \sum_{i=1}^{\infty} CH_E^i + \sum_{i=1}^{\infty} [MC^i \cdot 2i] + \frac{3}{2} MC_2^1 + (2 - \gamma - \delta) MC_2^2 + \frac{5}{2} MC_2^3 + (4 - \gamma - \delta) MC_2^4 + 2MC_3^1 + \frac{7}{2} MC_3^2 + (4 - \delta) MC_3^3 + 5MC_3^4 + (6 - \delta) MC_3^5 + MC_1 \cdot (\alpha + \beta + 3\gamma + 3\delta)$.

We assign the cost of GPTS among the structures of OPT (where a structure is either a cycle or a chain). We initialize the assigned cost of a structure to the number of arcs in the structure. Then, we increase the assigned cost of a structure according to the following:

- For a three arc cycle we increase the assigned cost by $2 - \alpha - \beta - \gamma - \delta$.
- For a four arc cycle we increase the assigned cost by $2 - \beta - \gamma - \delta$.
- For a cycle with at least five arcs we increase the assigned cost by 2.
- For a chain of OPT that belongs to the Eulerian subgraph we increase the assigned cost by 2.
- For a mega-chain of OPT with exactly two arcs, we act as follows (we write the increase for the complete mega-chain and not for each resulting chain).
  - If its length is at most $\frac{3n}{4} - 1$, we increase the assigned cost by $\frac{5}{2}$.

- If its length is in the interval $[\frac{3n}{4}, n-1]$, we increase the assigned cost by $3-\gamma-\delta$.
- If its length is in the interval $[n+1, \frac{5n}{4}-1]$, we increase the assigned cost by $\frac{7}{2}$.
- If its length is at least $\frac{5n}{4}$, we increase the assigned cost by $5-\gamma-\delta$.

- For a mega-chain of OPT with exactly three arcs, we act as follows (we again write the increase for the complete mega-chain and not for each resulting chain).
  - If its length is at most $n-1$, we increase the assigned cost by 3.
  - If its length is in the interval $[n+1, \frac{7n}{4}]$, we increase the assigned cost by $\frac{9}{2}$.
  - If its length is in the interval $[\frac{7n}{4}, 2n-1]$, we increase the assigned cost by $5-\delta$.
  - If its length is in the interval $[2n+1, \frac{5n}{2}-1]$, we increase the assigned cost by 6.
  - If its length is at least $\frac{5n}{2}$, we increase the assigned cost by $7-\delta$.

- For a mega-chain of OPT which has length in $[(i-1)n+1, in-1]$ and at least four arcs, we increase its cost (of the complete mega-chain) by $2i+1$. Note that the number of arcs in it, t is at least $\max\{4, i\}$.

- For a one-arc mega-chain of OPT we increase the assigned cost by $1+\alpha+\beta+3\gamma+3\delta$.

By Corollary 5 we conclude that the total assigned cost is at least the total cost of the solution returned by Algorithm GPTS.

5 A $\frac{98}{69}$-approximation algorithm: Algorithm COMB

In this section we design a new approximation algorithm COMB. Algorithm COMB combines the two algorithms: PIM and GPTS. It simply applies both PIM and GPTS, and picks the better solution.

Theorem 6 *Algorithm COMB is a $\frac{98}{69}$-approximation algorithm.*

Proof. Since $COMB = \min\{PIM, GPTS\}$, we conclude that $COMB \leq \frac{25}{69} \cdot PIM + \frac{44}{69} \cdot GPTS$. We will use the following values for α, β, γ and δ. $\alpha = \gamma = \frac{1}{11}$, $\beta = \frac{3}{22}$ and $\delta = \frac{3}{11}$. Note that these values satisfies our earlier assumptions on them.

We next bound the approximation ratio of Algorithm COMB. To do so, we will use the assigned costs of each of the algorithms (PIM and GPTS) to each of the structures of OPT (either a cycle or a chain), and show that the convex combination of the assigned costs of the two algorithms is at most $\frac{98}{69}$ the cost OPT paid for this structure. We then conclude that our algorithm is a $\frac{98}{69}$-approximation algorithm.

- For a two arc cycle of OPT, both GPTS and PIM assigned a cost of two and OPT pays also two. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 2+\frac{44}{69}\cdot 2}{2} = 1 \leq \frac{98}{69}$.
- For a three arc cycle of OPT, GPTS assigned a cost of $5-\alpha-\beta-\gamma-\delta = \frac{97}{22}$ and PIM assigned a cost of 4, whereas OPT pays three. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 4+\frac{44}{69}\cdot\frac{97}{22}}{3} = \frac{98}{69}$.
- For a four arc cycle of OPT, GPTS assigned a cost of $6-\beta-\gamma-\delta = \frac{11}{2}$, and PIM assigned a cost of 6 whereas OPT pays four. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 6+\frac{44}{69}\cdot\frac{11}{2}}{4} = \frac{98}{69}$.

- For a cycle with $i \geq 5$ arcs of OPT, GPTS assigned a cost of $i+2$, PIM assigned a cost of $\frac{3i}{2}$ for $i \geq 6$ and 7 for $i = 5$ whereas OPT pays i. Therefore, if $i = 5$ both GPTS and PIM assigned a cost of 7 and the ratio is $\frac{7}{5} \leq \frac{98}{69}$. Otherwise, i.e., $i \geq 6$ and the ratio is $\frac{\frac{25}{69}\cdot\frac{3i}{2}+\frac{44}{69}\cdot(i+2)}{i} = \frac{163}{138} + \frac{88}{69i} \leq \frac{489}{414} + \frac{88}{414} = \frac{577}{414} \leq \frac{98}{69}$.

- For a chain of OPT with i arcs that belongs to the Eulerian subgraph of OPT, GPTS assigned a cost of $i+2$ and PIM assign a cost of $\frac{3i+1}{2}$, whereas OPT pays $i+1$. Therefore in this case the ratio is $\frac{\frac{25}{69}\cdot\frac{3i+1}{2}+\frac{44}{69}\cdot(i+2)}{i+1} = \frac{75i+25+88i+176}{138(i+1)} = \frac{163}{138} + \frac{38}{138(i+1)} \leq \frac{182}{138} = \frac{91}{69} \leq \frac{98}{69}$.

- For a mega-chain of OPT with exactly two arcs and length at most $\frac{3n}{4} - 1$, GPTS assigned a cost of $\frac{9}{2}$, PIM assigned a cost of 3 and OPT also pays 3. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 3+\frac{44}{69}\cdot 4.5}{3} = \frac{273}{207} < \frac{98}{69}$.

- For a mega-chain of OPT with exactly two arcs and length in the interval $[\frac{3n}{4}, n-1]$, GPTS assigned a cost of $5-\gamma-\delta = \frac{51}{11}$, PIM assigned a cost of 3 and OPT also pays 3. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 3+\frac{44}{69}\cdot\frac{51}{11}}{3} = \frac{279}{207} < \frac{98}{69}$.

- For a mega-chain of OPT with exactly two arcs and length in the interval $[n+1, \frac{5n}{4} - 1]$ (which is actually two chains of OPT), GPTS assigned a cost of $\frac{11}{2}$, PIM assigned a cost of 4 and OPT also pays 4. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 4+\frac{44}{69}\cdot\frac{11}{2}}{4} = \frac{342}{276} < \frac{98}{69}$.

- For a mega-chain of OPT with exactly two arcs and length at least $\frac{5n}{4}$ (which is again two chains of OPT), GPTS assigned a cost of $7-\gamma-\delta$, PIM assigned a cost of 4 and OPT also pays 4. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 4+\frac{44}{69}\cdot\frac{73}{11}}{4} = \frac{392}{276} = \frac{98}{69}$.

- For a mega-chain of OPT with exactly three arcs and length at most $n-1$, GPTS assigned a cost of 6, PIM assigned a cost of 5 and OPT pays 4. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 5+\frac{44}{69}\cdot 6}{4} = \frac{389}{276} < \frac{98}{69}$.

- For a mega-chain of OPT with exactly three arcs and length in the interval $[n+1, \frac{7n}{4}-1]$ (which is actually at least two chains of OPT), GPTS assigned a cost of 7.5, PIM assigned a cost of at most 6 and OPT pays at least 5. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 6+\frac{44}{69}\cdot 7.5}{5} = \frac{480}{345} < \frac{98}{69}$.

- For a mega-chain of OPT with exactly three arcs and length in the interval $[\frac{7n}{4}, 2n-1]$ (which is at least two chains of OPT), GPTS assigned a cost of $8-\delta = \frac{85}{11}$, PIM assigned a cost of at most 6 and OPT pays at least 5. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 6+\frac{44}{69}\cdot\frac{85}{11}}{5} = \frac{490}{345} = \frac{98}{69}$.

- For a mega-chain of OPT with exactly three arcs and length in the interval $[2n+1, \frac{5n}{2} - 1]$ (which is actually three chains of OPT), GPTS assigned a cost of 9, PIM assigned a cost of at most 6 and OPT pays 6. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 6+\frac{44}{69}\cdot 9}{6} = \frac{546}{414} < \frac{98}{69}$.

- For a mega-chain of OPT with exactly three arcs and length in at least $\frac{5n}{2}$ (which is three chains of OPT), GPTS assigned a cost of $10-\delta = \frac{107}{11}$, PIM assigned a cost of at most 6 and OPT pays 6 too. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 6+\frac{44}{69}\cdot\frac{107}{11}}{6} = \frac{578}{414} < \frac{98}{69}$.

- For a mega-chain of OPT with t arcs that has length in $[(i-1)n+1, in-1]$, where the number of arcs in it, t is at least $\max\{4, i\}$. We assigned a cost of $t+2i+1$ for GPTS. Clearly OPT pays at least $i+t$. If $i = 1$, PIM pays at most $\frac{3t+1}{2}$. If $i = 2$, PIM can build chains of two

arcs from all arcs but at most three, and so it pays at most $\frac{3t+3}{2}$. Otherwise, i.e. $i \geq 3$, PIM pays at most $2t$.

In the first case we get using $t \geq 4$, $\frac{\frac{25}{69}\cdot\frac{3t+1}{2}+\frac{44}{69}\cdot(t+3)}{t+1} = \frac{163t+289}{138(t+1)} = \frac{163}{138} + \frac{126}{138\cdot 5} = \frac{941}{690} < \frac{98}{69}$.

In the second case, using $t \geq 4$, we get, $\frac{\frac{25}{69}\cdot\frac{3t+3}{2}+\frac{44}{69}\cdot(t+5)}{t+2} = \frac{163t+515}{138(t+2)} = \frac{163}{138} + \frac{189}{138\cdot 6} = \frac{1167}{828} < \frac{98}{69}$.

In the last case we get, using $t \geq 4$, $t \geq i$ and $i \geq 3$, $\frac{\frac{25}{69}\cdot 2t+\frac{44}{69}\cdot(t+2i+1)}{i+t} = \frac{94t+88i+44}{69(i+t)} \leq \frac{98}{69} - \frac{2}{69(i+t)} < \frac{98}{69}$.

- For a one-arc mega-chain of OPT, PIM and OPT pay two and GPTS assigned a cost of $2+\alpha+\beta+3\gamma+3\delta = \frac{73}{22}$. Therefore, the ratio in this case is $\frac{\frac{25}{69}\cdot 2+\frac{44}{69}\cdot\frac{73}{22}}{2} = \frac{98}{69}$.

■

We note that example 15 in [1] provides a negative example for algorithm COMB. Let $n = 6$, and the input consist of the nine arcs $(0,2),(2,5),(5,0),(0,3),(3,4),(4,0),(1,2),(2,4),(4,1)$. Clearly OPT can construct three cycles and have $opt = 9$. If PIM removes the cycle $(0,2),(2,4),(4,0)$ first, it is left with six arcs that cannot form a valid cycle. Then, PIM can match the three pairs that give chains of length 4, $(2,5)(5,0)$, $(0,3)(3,4)$ and $(4,1)(1,2)$, and terminate. Consequently, $PIM = 12$. Algorithm GPTS may remove the same cycle of three arcs. There are no mega-chains, therefore the six arcs that are left, are partitioned into three chains of length four, similarly to the output of PIM.

Therefore, we establish the following theorem.

Theorem 7 *The approximation ratio of Algorithm COMB is at most $\frac{98}{69}$, and at least $\frac{4}{3}$.*

6 Other algorithms

In this section we consider several previously known algorithms, and give tight or tighter bounds on their performance.

6.1 PCM

In [1] it was shown that the approximation ratio of PCM is in the interval $[3/2, 5/3]$. We show that the upper bound is tight by giving a class of examples with performance ratio that approaches $5/3$ for large n.

Theorem 8 *The algorithm PCM has approximation ratio of exactly $5/3$.*

Proof. We need to show that the approximation ratio of PCM is at least $5/3$. Let n=8k. The input consists of the following $6k+3$ arcs. For every i such that $k \leq i \leq 3k$, we have the three arcs $(i,4k)$. $(4k,i+4k)$ and $(i+4k,i)$. OPT consists of $2k+1$ cycles of the form $(i,4k),(4k,i+4k),(i+4k,i)$, for $k \leq i \leq 3k$. Therefore, $opt = 6k+3$.

Since OPT consists of cycles, all arcs have the same load and it may be the case that arc $(0,1)$ of the cycle is removed. We are left with arcs of the types $(i,4k)$ and $(4k,i)$. These arcs can only be combined into chains of length two (all the end-points except $4k$ are distinct). We may get the chains $(i,4k),(4k,i+4k+1)$ for $k \leq i \leq 3k-1$, and the chain $(3k,4k),(4k,5k)$. These are $2k+1$ chains. Each chain among the first $2k$ chains has length $4k+1 > n/2$. All the removed arcs have length $4k = n/2$. The only possible way to combine to chains is to combine $(3k,4k),(4k,5k)$ with one chain of length $4k$ (either with $(5k,k)$ or with $(7k,3k)$).

We are left with $2k$ chains of two arcs, one chain of three arcs, and $2k$ chains of a single arc. Therefore, $APX = 6k+4+4k = 10k+4$, and the ratio is $(10k+4)/(6k+3)$. As k grows the ratio approaches $5/3$. ■

6.2 PET-TS

In [1], it is shown that algorithm ET-TS has performance of exactly 1.75 (this is the version of PET-TS that does not remove two arc cycles as a preprocessing step). However, it was shown that algorithm PET-TS has performance in the interval $[1.5, 1.75]$. We narrow the gap by showing the following.

Theorem 9 *Algorithm PET-TS has approximation ratio of at least* $5/3$.

Proof. Let $k \geq 2$ be an integer, and let $n = 4 \cdot 3^{4k}$. For every $0 \leq i < n$, we have the following arcs $(i, i+n/2-11), (i+n/2-11, i+n/2-11+3^{k+1})$. We also have for $3 \leq s \leq k$ and for every $0 \leq i < n$ the arcs $(i, i+3^s), (i+3^s, i+n/2-2), (i+n/2-2, i)$. Clearly, OPT can use cycles from this last class of arcs (the triples), and chains of two arcs from the first class of arcs. Thus $opt \leq 3n + 3(k-2)n = 3kn - 3n = n(3k-3)$. Algorithm PET-TS builds cycles which consist of the arcs of length $n/2+2$, each cycle consists of the same number of arcs, and gets decomposed into chains consisting of a single arc each. For $3 \leq s \leq k+1$, PET-TS constructs cycles (two cycles for each value of s) that alternate between arcs of lengths 3^s and arcs of length $n/2 - 3^{s-1} - 2$. Since $s \geq 3$, we have $3^s > 2(3^{s-1}+2)$, so when a cycle is partitioned, each arc of length $n/2 - 3^{s-1} - 2$ is in a separate chain. Each cycle is therefore decomposed into chains of length larger than $n/2$ except possibly one last chain. All lengths of arcs are odd, therefore a cycle is closed having the same number of arcs of both lengths. For a given s, all the arcs are combined together into at most two cycles. This can be seen in the following way. We need to solve the equation $(3^s + n/2 - 3^{s-1} - 2)i = 0 \mod n$. Using the value of n we get that the smallest possible solution is $i = 2 \cdot 3^{4k} = n/2$. This means that a cycle consists of $n/2$ pairs of arcs.

Chains of length larger than $n/2$ do not get combined later, we count such chains resulting from the alternating cycles. For each value of s we get at least $2(n/2-1)$ such chains. We get that the cost of the algorithm is more than $APX > 2(k-2)n + 3(k-1)2(n/2-1)) = 5nk - 7n - 6k + 6 = n(5k-7) - 6(k-1)$. As k grows the ratio APX/opt approaches $5/3$. ■

6.3 IM

In [1], it is shown that algorithm IM has performance in the interval $[3/2, 5/3]$. We narrow the gap by showing the following.

Theorem 10 *The approximation ratio of Algorithm IM is at least* 1.6*, if the two arc cycles removal step is not performed.*

Proof. Let $n = 10$. The arcs are $(0,8), (8,9), (9,0), (4,9), (9,4)$. OPT has the cycles $(0,8), (8,9), (9,0)$ and $(4,9), (9,4)$.

The matching chooses the pairs $(4,9)$,$(9,0)$ and $(8,9)$,$(9,4)$. The arc $(0,8)$ is unmatched.

We get chains of length larger than $n/2$ which cannot be combined, so the next matching step matches nothing and $APX = 8$ whereas $opt = 5$. ■

Recall that PPIM removes two arc cycles before applying matchings.

Theorem 11 *Algorithm PPIM has approximation ratio of at least* $14/9$.

Proof. Let $n = 10$. The input consists of nine arcs, $(0,1)$, $(1,6)$, $(6,0)$, $(0,2)$, $(2,7)$, $(7,0)$, $(0,3)$, $(3,8)$, $(8,0)$. OPT constructs three cycles from the first three arcs, the next three arcs and the last three arcs. However, a maximum matching consists of four pairs, and a possible matching is $(7,0)$ with $(0,3)$, $(6,0)$ with $(0,2)$, $(3,8)$ with $(8,0)$, $(0,1)$ with $(1,6)$ and $(2,7)$ is left unmatched. In the second round of finding a maximum matching there are no pairs to match, since all chains but $(2,7)$ have length larger than $n/2 = 5$. Therefore, we get $APX = 14$ and $opt = 9$. ■

We can also show that the upper bound 5/3 is not tight

Theorem 12 *Algorithms IM and PPIM have approximation ratios of strictly less than* 5/3*.*

Proof. (Sketch.) The upper bound proof [1] which yields an upper bound of 5/3 reaches this bound as a result of two assumptions. The first is that the iterative matching performs a single matching step. The second is that OPT consists solely of triangles, and one arc of each triangle remains unmatched. However, if OPT indeed consists of triangles only, it is the case that either on average more than two arcs from each triangle were matched (it is possible to create three matched pairs from two triangles that have a vertex in common), or the second phase of IM creates some complete triangles (this happens if there are few cases of arcs of different triangles of OPT that were matched). The proof distinguishes between two cases. One case where there are relatively few triangles in OPT, and therefore the worst case ratio 5/3 may be reached for only a fraction of the arcs. Another case where there are many triangles, and we analyze two iterations of IM. This gives an upper bound of slightly less than 5/3. ■

6.4 MCC-TS

In [1] algorithm MCC-TS has a preprocessing step of two arcs cycles removal. However, the algorithm can be easily adapted to work without this step, and the analysis still works. While building the auxiliary graph the option of two arcs that form a cycle should be taken into account, and the arcs between those arcs both get weight one. It was shown [1] that the performance ratio for this algorithm is in the interval $[1.5, 1.6]$. We show that the upper bound is tight. To distinguish between the two versions we call them MCC-TS (the version with pre-processing) and NMCC-TS (without pre-processing).

Theorem 13 *Algorithm NMCC-TS has approximation ratio of exactly* 1.6*.*

Proof. We need to show that the approximation ratio is at least 1.6.

Let n=12, the input consists of sixty arcs. For $0 \leq i \leq 11$, we have twice the arc $(i, i+5)$, once the arc $(i, i+2)$, and twice the arc $(i, i+6)$. OPT builds 24 cycles which are $(i, i+5), (i+5, i+10), (i+10, i)$ and $(i, i+6), (i+6, i)$ for $0 \leq i \leq 11$. Therefore, $opt = 60$.

Running the algorithm we get the (invalid) cycles. $(i, i+5), (i+5, i+11), (i+11, i+4), (i+4, i+10), (i+10, i)$. Checking the five options to partition such a cycle we get that each one has 3 chains. So we get 36 chains, and $APX = 96$ whereas $opt = 60$. ■

For algorithm MCC-TS (with two arc cycles removal), we can show a tight bound of 14/9. We prove it using the next two lemmas.

Lemma 14 *Algorithm MCC-TS has approximation ratio of at least* 14/9*.*

Proof. Let $n = 24m^4$ for an integer $m > 1$. The input arcs are described in Table 1. The input consists of five families of arcs. Each family has certain amount of parallel copies of arcs (this amount appears in the column Amount). The arc set of each family is parameterized by i or by

Amount	Index range	Arcs
$12m^3$	$0 \le i < n$	$(i, i+n/2), (i+n/2, i-2m^2), (i-2m^2, i)$
$24m^2$	$0 \le i < n, 1 \le s \le m$	$(i, i+n/3-sm^2)$
$12m^2$	$0 \le i < n, 1 \le s \le m$	$(i, i+n/3+2(s+1)m^2)$
$12m^3$	$0 \le i < n$	$(i, i+2), (i, i+3)$
$6m^3$	$0 \le i < n$	$(i, i-4), (i, i-6)$

Table 1: Input arcs

i, s. For each value of the parameters in the Index range (that appears in the second column) we have the amount of parallel copies of the arcs that appear in the Arcs column.

We give an upper bound on *opt* by the cost of the following solution. The solution has for every $0 \le i < n$, $12m^3$ cycles which are $(i, i+n/2), (i+n/2, i-2m^2), (i-2m^2, i)$. For $0 \le i < n$ we have $6m^3$ cycles of $(i, i+2), (i+2, i+4), (i+4, i)$ and $6m^3$ of $(i, i+3), (i+3, i+6), (i+6, i)$ for $0 \le i < n$. Finally, for every $0 \le i < n$ and for every $2 \le s \le m$ there are $12m^2$ identical cycles: $(i, i+n/3-sm^2), (i+n/3-sm^2, i+2n/3-2sm^2), (i+2n/3-2sm^2, i)$. The arcs $(i, i+n/3-m^2)$ and $(i, i+n/3+2(m+1)m^2)$, are not combined into cycles but into paths, $12m^2$ copies of $(i, i+n/3-m^2), (i+n/3-m^2, i+2n/3-2m^2)$ for every $0 \le i < n$ and $6m^2$ of $(i, i+n/3+2(m+1)m^2), (i+n/3+2(m+1)m^2, i+2n/3+4(m+1)m^2)$ for every $0 \le i < n$. Since $m > 1$, $4(m+1)m^2 < n/3$.

The MCC solution may consist of the following cycles (it manages to combine all arcs into long cycles). Note that each pair of consecutive arcs in each cycle is indeed valid for MCC as their combined length is less than n. This will hold due to the choice of $n = 24m^4$ which gives $(m+2)m^2 < n/3$ and $2(m+1)m^2 < n/6$. We have $12m^2$ copies of the following cycle $(i, i+n/2), (i+n/2, i+5n/6-sm^2), (i+5n/6-sm^2, i+n/3-(s+2)m^2), (i+n/3-(s+2)m^2, i+2n/3-(2s+2)m^2), (i+2n/3-(2s+2)m^2, i)$, for every $0 \le i < n$ and for every $1 \le s \le m$. These cycles can be decomposed into three chains, no matter which arc is chosen to be first. We have the following cycle $6m^3$ times for every $0 \le i \le 4m^2-1$. The number of arcs in a cycle is $48m^2$, and no vertex is repeated until the cycle is closed. The cycle consists of $6m^2$ phases of eight arcs. For $0 \le q \le 6m^2-1$, we have the eight arcs $(i+4qm^2, i+2+4qm^2), (i+2+4qm^2, i+2+(4q+2)m^2), (i+2+(4q+2)m^2, i+4+(4q+2)m^2), (i+4+(4q+2)m^2, i+(4q+2)m^2), (i+(4q+2)m^2, i+3+(4q+2)m^2), (i+3+(4q+2)m^2, i+3+(4q+4)m^2), (i+3+(4q+4)m^2, i+6+(4q+4)m^2), (i+6+(4q+4)m^2, i+(4q+4)m^2)$. Since $m > 1$ is an integer, $2m^2 \ge 8$, and so vertices with different residues (indices $\mod 2m^2$) cannot coincide. Vertices with the same residue are distinct due to the different coefficients of m^2. The decomposition of each cycle creates $24m^2$ chains. We get that $opt \le n(36m^3+18m^3+18m^3+36m^2(m-1)+54m^2) = n(108m^3+18m^2)$. $APX = (12nm^3) \cdot 8 + 48m^2 \cdot 6m^3 \cdot 4m^2 \cdot 1.5 = nm^3(96+72) = 168nm^3$. This gives a ratio of $168m/(108m+18)$ which tends to $14/9$ for large m. ■

Lemma 15 *Algorithm MCC-TS has approximation ratio of at most* $14/9$.

Proof. For every arc e, define a weight $w(e)$ in the following way. $w(e) = 1/3 + 2\ell(e)/(3n)$. We show the following properties.

1. The total sum of weights of arcs is at most $(5/9)$opt.
2. The number of new chains caused by decomposition is at most the total sum of weights.

The total cost for original chains and valid cycles constructed by MCC is bounded by *opt*, so the result of proving the properties would be $APX \le 14opt/9$.

We start with proving property 1. Consider a cycle C in OPT which consists of k arcs. The total cost paid by OPT for C is k. The total weight of the arcs of C is exactly $k/3 + 2/3$, and $k/3 + 2/3 \leq 5k/9$ for $k \geq 3$. Consider a chain created by OPT which consists of k arcs. The total cost paid by OPT for this chain is $k + 1$. The total weight of the arcs of this chain is at most $k/3 + 2/3$, and $(k/3 + 2/3) \leq 5(k + 1)/9$ for $k \geq 1$. Next, we prove property 2. Consider a (not necessarily valid) cycle of $2k + 1$ arcs constructed by MCC which is of length sn for some integer s. Every such cycle can be split into at most $2s - 1$ chains in the following way. Let i be the starting vertex of an arc of the cycle, then i would be the first end-point of the first chain and the last end-point of the last chain (it can be the case where those two chains are combined into one). The distance to go from the first end-point to the last is sn. The length of two consecutive chains along the cycle is at least $n + 1$ (otherwise, they can be merged). If there are $2s$ chains, this means that the distance between the first and the last is more than sn, and therefore there are at most $2s - 1$ chains. On the other hand any pair of successive arcs can be combined in a chain due to the construction of the MCC graph, so $k + 1$ chains are always possible. We get that the number of new chains is at most $\min(k + 1, 2s - 1)$. The weight for these $2k + 1$ arcs is $(2k + 1)/3 + 2s/3 = (2k + 2)/3 + (2s - 1)/3 \geq (2/3 + 1/3) \min(k + 1, 2s - 1)$. Therefore, the weight of the cycle is at least the amount of additional cost caused by the decomposition.

Consider a cycle of $2k$ arcs which is of length sn for an integer s. We can get that the number of new chains is at most $\min(k, 2s-1)$. The weight for these $2k$ arcs is $2k/3+2s/3 > 2k/3+(2s-1)/3 \geq (2/3 + 1/3) \min(k, 2s - 1)$.

Consider a chain of $2k$ or $2k + 1$ arcs with length in the interval $[sn, (s + 1)n)$. Note that the original connected component built by MCC is already a chain and not a cycle. There are at most $\min(k, 2s)$ new chains. The weight of the chain is at least $2k/3 + 2s/3 \geq \min(k, 2s)$. ■

Summarizing we proved the following theorem.

Theorem 16 *Algorithm MCC-TS has approximation ratio of exactly* 14/9.

7 Conclusion

We introduced an approximation algorithm COMB for the problem of minimizing the number of SONET ADMs. COMB is a combination of two algorithms, one of them was introduced in this paper and the other was previously studied. Algorithm COMB is the current best approximation algorithm for this problem. We showed that it is a 98/69 approximation algorithm, and we provided a lower bound on its worst-case performance of 4/3. Closing this gap, and finding a better approximation algorithm is left for future research. We also raise the following question: Is there a good approximation algorithm whose preprocessing step consists of cycle removal solely (without removal of chains)?
A summary of the results in the paper can be found in Table 2.

References

[1] G. Calinescu and P.-J. Wan, "Traffic partition in WDM/SONET rings to minimize SONET ADMs," *Journal of Combinatorial Optimization*, **6**, 425-453, 2002.

[2] O. Gerstel, P. Lin and G. Sasaki, "Wavelength assignment in a WDM ring to minimize cost of embedded SONET rings," *Proc. INFOCOM 1998*, **1**, 94-101, 1998.

[3] L. Liu, X. Li, P.-J. Wan and O. Frieder, "Wavelength assignment in WDM rings to minimize SONET ADMs," *Proc. INFOCOM 2000*, **2**, 1020-1025, 2000.

Heuristic	Lower bound on the approximation ratio in [1]	Lower bound on the approximation ratio (this paper)	Upper bound on the approximation ratio (this paper)	Upper bound on the approximation ratio in [1]
COMB	–	4/3	$\mathbf{98/69 \approx 1.4203}$	–
PIM	4/3	3/2	–	3/2
PCM	3/2	5/3	–	5/3
MCC-TS	3/2	14/9	14/9	8/5
NMCC-TS	3/2	8/5	–	8/5
PET-TS	3/2	5/3	–	7/4
IM	3/2	8/5	$< 5/3$	5/3
PPIM	3/2	14/9	$< 5/3$	5/3

Table 2: Summary of results

[4] P.-J. Wan, G. Calinescu, L. Liu and O. Frieder, "Grooming of arbitrary traffic in SONET/WDM BLSRs," *IEEE Journal on Selected Areas in Communications*, **18**, 1995-2003, 2000.