WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

AN ARC FLOW FORMULATION FOR THE TRAVELLING SALESMAN PROBLEM

A. CLAUS
ECOLE SUPERIEURE DE COMMERCE (ESCAE)
MONTPELLIER, FRANCE

and

* SLOAN SCHOOL OF MANAGEMENT
MASSACHUSETTS INSTITUT OF TECHNOLOGY
WP # 1949-87 Oct /22/1987

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139
AN ARC FLOW FORMULATION FOR THE TRAVELLING SALESMAN PROBLEM

A. CLAUS
ECOLE SUPERIEURE DE COMMERCE (ESCAE)
MONTPELLIER, FRANCE

and

SLOAN SCHOOL OF MANAGEMENT
MASSACHUSETTS INSTITUT OF TECHNOLOGY
WP # 1949-87   Oct /22/1987

Visiting Fall 1987
ABSTRACT. Multicommodity formulations of the travelling salesman problem (TSP) use usually node commodities that have to be send from a given source to the various nodes of the problem. An alternative multicommodity formulation for the TSP is to consider the arcs of a solution to the TSP as commodities.

1. INTRODUCTION Many formulations for the travelling salesman problem have been developed. These include n-commodity formulations (Wong[6],Claus[1]) and 2-commodity formulations (Finke,Claus,Gunn[3],Finke[4]).

The formulation of the TSP described in this paper uses first n2 commodities to describe the problem and then n6 to proof certain theorems about the solutions satisfying the required constraints. It will for example be shown that a convex combination of two assignment problems will only be feasible if the solution is also a convex combination of hamiltonian paths.

2. THE PROBLEM AND FORMULATION. The problem we consider is one of finding a minimum cost Hamiltonian path from one vertex s to another t in a given weighted directed graph.

The programming approach involves introducing a variable Xij for each pair of nodes, that takes on value if the edge i to j is on the path corresponding to the given choice of variables and 0 if it is not. (In the corresponding polytope these variables are constrained to be between zero and one.)

The continuity constraints, for all i,

\[ \text{SUM}(i) \ Xij = 1, \ \text{SUM}(j) \ Xij = 1, \]
\[ \text{SUM}(j) \ Xsj = 1, \ \text{SUM}(j) \ Xjt = 1, \]
\[ \text{SUM}(i) \ Xis = 0, \ \text{SUM}(j) \ Xtj = 0, \]

insure that, for an integral solution, exactly one arc enters and leaves each vertex.
There are several ways to add additional constraints in order to guarantee that an integral solution corresponds to a Hamiltonian tour.

The Dantzig, Fulkerson and Johnson[2] approach was to require that for every nontrivial subset \( V \) of the nodes (other than \( s \)), we have

\[
X_{ij} \geq 1
\]

for all \( i \) not in \( V \) and \( j \) in \( V \). This insures that one can get from the source to any set of nodes in a tour corresponding to an integral solution.

In (Claus[1]) \( n \)-commodities one corresponding to each node were used to define the subtour elimination polytope, and a more compact two-commodity versions of the TSP can be found in (Finke,Claus,Gunn[3],Finke[4]).

We now introduce a different approach. The commodities \( X \) now to be send from the source are the arcs \( (X_{kp}) \).

Thus for each arc \( X_{kp} \) we define variables \( X_{ij} \) (which correspond to the flow of commodity \( X \) on the arc \( i,j \)) with the following properties:

1.1 Continuity at \( i \): \( \sum(j) X_{ji} = \sum(j) X_{ij} \) for \( i \) other than \( s \) or \( t \);

1.2 Flow from \( s \) to \( k \): \( \sum(j) X_{sj} = X_{kp}, \sum(i) X_{ik} = X_{kp} \);

1.3 Flow from \( k \) to \( t \): \( \sum(j) S_{pj} = X_{kp}, \sum(i) S_{it} = X_{kp} \);

1.4 Continuity at \( i \): \( \sum(j) \{ S_{ij} - S_{ji} \} = 0 \);

1.5 Identity: \( A_{ij} = X_{ij} + S_{ij} \);

1.6 Capacity: \( 0 \leq A_{ij} \leq X_{ij} \);

1.7 Symmetry: \( A_{kp} = A_{qr} \).
A second level of commodities (variables) \( A \) is introduced with similar types of constraint as the above except that the quantities leaving the source and transiting each node are now equal to \( A_{qr} \) instead of \( A_{kp} \), and

2.1 Continuity at \( i \): \( \sum(j) A_{ji} = \sum(j) A_{ij} = A_{qr} \)

for \( i \) other than \( s \) or \( t \);

2.2 Flow from \( s \) to \( k \): \( \sum(i) A_{si} = A_{qr} \), \( \sum(i) A_{ik} = A_{qr} \);

2.3 Capacity:

\[
0 \leq A_{ij} \leq A_{ij},
\]

2.4 Symmetry:

\[
A_{ij} = A_{ij};
\]

Before considering the flow behaviour for a convex combination of two assignment solutions let us define the following terms:

A 'Perfect Exchange Node' (PEN) is a node where all flows coming into a node on one arc must leave the node together on another arc.

A 'Regular Flow' (RF) is a flow from the source such that along an RF path all arc-commodities progress together along such a path Fig 1). A 'Regular Flow Path' (RFP) is a path with RF.

\[
\begin{array}{cccc}
  s1 & s1 & s1 & s1 \\
  A_{s1} > 0 & A_{12} > 0 & A_{23} > 0 & A_{34} > 0 \\
  12 & 12 & 12 & 12 \\
  A_{s1} > 0 & A_{12} > 0 & A_{23} > 0 & A_{34} > 0 \\
\end{array}
\]

(s) \( \rightarrow (1) \rightarrow (2) \rightarrow (3) \rightarrow (4) \)

\[
\begin{array}{cccc}
  23 & 23 & 23 & 23 \\
  A_{s1} > 0 & A_{12} > 0 & A_{23} > 0 & A_{34} > 0 \\
  34 & 34 & 34 & 34 \\
  A_{s1} > 0 & A_{12} > 0 & A_{23} > 0 & A_{34} > 0 \\
\end{array}
\]

Fig 1
3. TWO ARC THEOREMS AND PROOFS

Theorem 1: If a solution has an RFP of length $h < n$ then we cannot have two arcs back from $h$ towards the RFP.

Proof: Consider the RFP shown below (Fig 2) and assume the contrary i.e node $h$ has two arcs back towards the RFP. In this case the commodity $A_{pqr}$ (were $hp$ and $hr$ are the two arcs leaving node $h$) cannot leave node $h$, violating constraints 2.1.

![Diagram Fig 2](image)

(s) $\rightarrow$ (k) $\rightarrow$ (p) $\rightarrow$ (q) $\rightarrow$ (r) $\rightarrow$ (h)

Theorem 2: If we have an RFP of length $m$ to node $h$ and node $h$ is not the sink then we have an RFP of length $m+1$.

Proof: Consider the RFP in Fig 3. Now assume that one of the arcs leaving node $h$ $(h,f)$ has a flow value of zero for commodity $(k,p)$ $A_{hf}=0$. This requires that this commodity as well as the other commodities on the RFP are positive on the arc $(h,g)$ $(A_{hg} > 0, ed \text{ RFP})$. Due to constraints 1.7 we also have $A_{ed} > 0$, $hg$ RFP, thus proving the theorem.

![Diagram Fig 3](image)
Theorem 3: If we have a regular flow of length n-1 then we have a regular flow of length n, where n is the number of nodes of the graph.

Proof: Consider an RFP of length n-1 (Fig 4). Node n must have two arcs to this path. But this would require that the path is not an RFP.

\[(s)\rightarrow(k)\rightarrow(p)\rightarrow(q)\rightarrow(r)\rightarrow(n-1)\rightarrow(t)\]

\[kp\]
\[Anp = 0\]
\[Anr = Xkp\]

Fig 4

Theorem 4: A path from s to t cannot be an RFP unless it is part of an assignment solution.

Proof: Assume we have an RF path from s to t of length h<n. Due to the flow constraints (1.2) we must have at least one arc connected to the other nodes not on the RFP (Fig 5). Assume that this node is node r. Then the other inward arc into node r must have positive values for all the commodities on the RFP. Thus we have an RFP path going through arc (q,r). This path cannot go back to any nodes on the original RFP path and is thus a RF cycle. If there are more nodes, then one of those must have an inward arc from the RFP or the RF cycle etc..

\[(s)\rightarrow(l)\rightarrow(2)\rightarrow(3)\rightarrow(4)\rightarrow(5)\rightarrow(6)\rightarrow(t)\]

\[34\]
\[A_{912} = X_{34}\]
\[34kp\]
\[thus \ A_{912} > 0\]

Fig 5
The previous theorem showed that RFP's are only possible on assignment solutions. This implies that a feasible solution to the arc formulation model is a convex combination of RFP's.

Theorem 5: A convex combination of two RFP's must also be a convex combination of two Hamiltonian Paths, otherwise constraints (1.2) cannot be satisfied.

Proof: Assume the contrary i.e. the two RFP's consist of zero or only one Hamiltonian path. Since the RFP carry only arcs from their own solution an RFP can only be a Hamiltonian path in order to satisfy constraints (1.2).

\[(s)\rightarrow(1)\rightarrow(2)\rightarrow(3)\rightarrow(4)\rightarrow(5)\rightarrow(6)\rightarrow(t)\]

\[A = 0\rightarrow(9)\rightarrow(10)<\rightarrow\rightarrow A = 0\]

\[A = 0\rightarrow(12)<\rightarrow(11)<\rightarrow A = 0\]

Fig 6

Theorem 6: If n-1 is even and the solution for n-1 nodes has to be a convex combination of Hamiltonian paths then it is also true for n nodes.

Proof: Since n is odd we must have at least one alternating cycle (Fig 7) of odd length implying that we have at least one PEN, thus reducing the problem to an n-1 node problem.

\[\text{PEN} > (1)\rightarrow(2)<\rightarrow(3)\]

\[(5)\rightarrow(4)\]

Fig 7
4. THREE ARC THEOREMS AND PROOFS

In order to prove certain theorems about three arc assignment solutions we introduce a third level of commodities (variables) $A_{ij}$ with similar types of constraints as for the other commodities except that the quantities leaving the source and transiting each node are now equal to:

$$\text{SUM}(j) A_{ij} = A_{vw}.$$

3.1 Continuity at i: $\text{SUM}(j) \{ A_{ji} - A_{ij} \} = 0$

for $i$ other than $s$ or $t$;

3.2 Flow from $s$ to $k$: $\text{SUM}(i) A_{is} = A_{vw},$

$$\text{SUM}(i) A_{ik} = A_{vw};$$

3.3 Capacity:

$$0 \leq A_{ij} \leq A_{ij},$$

$$0 \leq A_{ij} \leq A_{ij},$$

$$0 \leq A_{ij} \leq A_{ij},$$

3.4 Symmetry:

$$A_{ij} = A_{vw},$$

etc. for all permutations.

Theorem 7: If we have an RFP of length $m < n$ then we also have an RFP of length $m+1.$

Proof: Consider the graph of Fig 9. Assume none of the three arcs has all $A > 0$ for $(k,p)$ on the RFP from $s$ to $k.$
This would prohibit commodity $A_{kpvw}$ from transiting node $h$, thus contradicting the flow constraint for this commodity:

$$
\begin{align*}
kpqrvw & \quad kpqr \\
Ahj & = Avw;
\end{align*}
$$

Theorem 8: The RFP's can only be assignment solution.

Proof: The proof is similar to the proof of the theorem for two arcs. Consider an RFP path from $s$ to $t$ that does not belong to an assignment solution. Due to the flow constraints (1.2) this implies that we must have at least one arc on one of the remaining nodes $l$. Due to theorem 9 this node requires at least one arc with the same RFP as on the path. Since it cannot go back to the RFP it has to form an RF circuit, etc...

Theorem 9: A solution satisfying the AFP constraints, assuming that it is a convex combination of at most three RFP's, must also be a convex combination of three hamiltonian paths.

Proof: It was shown above that we can only have RFP's along assignment solutions. If the solution is not a convex combination of hamiltonian paths then this would violate at least one of the flow constraints (1.2).
5. APPLICATIONS

The above formulation has been applied to a limited number of problems and the initial computational results are encouraging.

I. Peterson Graph

Applying the above constraints to the Peterson graph (Fig. 10) we can prove with a relative small linear program (64 variables, 54 constraints) that this graph does not have a Hamiltonian path in its support, since it does not satisfy the AFP constraints.

Consider commodity (1,2) and assume that $A_{23} > 0$.

This requires that commodity $A_{38} + A_{34} > 0$. 
Assume $A_{38} = \alpha$. This implies that for each node $j$ we have

\[ \sum_{i} A_{ij} = \alpha. \]

Maximizing $\alpha$ it was found that its optimal value is zero.

The same holds for the commodity $A_{34}$ as well as for the commodities $(1,2)(2,7)$. Thus $X_{12}$ has to equal zero and since the graph is symmetric we can conclude that we also must have $X_{15} = 0$.

Since a feasible solution to the subtour elimination constraints would require that $X_{12} + X_{15} > 0$ we have a contradiction and thus proof that the Peterson graph does not have a Hamiltonian path in its support.

This result is encouraging since most known linear programming approaches to the Traveling Salesman Problem have failed at the Peterson graph.

II. A six node problem (Murty[7])

The traveling salesman problem in [7] has the following distance matrix:

\[
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
1 & - & 27 & 43 & 16 & 30 & 26 \\
2 & 7 & - & 16 & 1 & 30 & 25 \\
3 & 20 & 13 & - & 35 & 5 & 0 \\
4 & 21 & 16 & 25 & - & 18 & 18 \\
5 & 12 & 46 & 27 & 48 & - & 5 \\
6 & 23 & 5 & 5 & 9 & 5 & - \\
\end{array}
\]

As reported in [3] the optimal assignment solution has a value of 54 units. The optimal solution using the subtour elimination constraints (Fig 11) has a value of 57.9. Using a branch and bound procedure the optimal tour can be found
From Fig. 11 we notice for example that it is not feasible to send 0.7 units of commodity X35 from s to node 3, and 0.7 units of slack from node 5 to t.

Introducing flow requirements for commodity A35:

\[(3,5)\]

\[\text{i.e. } \sum_{i} X_{si} = X_{35},\]

\[(3,5)\]

\[\sum_{i} S_{it} = X_{35},\]

Continuity at node i:

\[\sum_{j} S_{ij} - \sum_{j} S_{ji} = 0, \quad \sum_{j} X_{ij} - \sum_{j} X_{ji} = 0,\]

\[\sum_{j} S_{ij} + \sum_{j} X_{ij} = X_{35};\]

The linear programming solution is integer valued and is thus the optimal tour (Fig. 12). It has a length of 63 units.
6. PROPOSED ALGORITHM

As mentioned above, numerical results are quite limited at this time, but the following algorithm appears to be promising:

**STEP 1** Find the optimal solution using the assignment procedure. If we have a tour STOP, if not go to STEP 2.

**STEP 2** Find an optimal solution using the two-commodity subtour elimination constraints [3]. If integer valued STOP, if not go to STEP 3.

**STEP 3** Find an arc with violated flow requirements; (If none can be found go to STEP 4) introduce the necessary constraints and find the optimal solution. If the solution is integer valued STOP the optimal tour has been found, else we have a lower bound. Introduce this lower bound (i.e. the objective function value has to be $\geq$ this bound) and discard the previous flow requirements. Go to STEP 3.

**STEP 4** Find a two-arc commodity with violated flow requirements.

**STEP k+2** Find a $k$ arc commodity
7. CONCLUSIONS

Although the number of potential variables is extremely large the above outlined algorithm would only need 4-commodities at a time (and is thus computationally feasible even for very large problems), if at each STEP we get an increase in the objective function value (No alternative optimal solutions) and probably not too many commodities if we do have alternative optimal solutions.

At each STEP we introduce subtour elimination constraints for the commodities introduced in the previous STEP. The interesting theoretical question arises as to whether we have a reasonable worst case upper bound on the number of STEPS we do have to execute. If this upper bound would be a constant then this would imply that \( P=NP \). Thus it appears likely that this upper bound is an exponential function of the number of nodes and or arcs of the graph.

Past computational experience with the two-commodity approach [3],[5] has shown that the larger the problem size the more likely the optimal solution to the two-commodity problem will be integer valued; thus we may expect that for large problems we rarely need more than 4-commodities and that in view of the results for the Peterson graph we probably don't have to go very often beyond STEP 4 or 5.

Extensive numerical experimentation is under way at this time, and will be reported on soon.

8. ACKNOWLEDGEMENTS The author would like to thank D.J.Kleitman for editorial assistance and a lot of advice.
REFERENCES
