A History of the Traveling Tournament Problem

Kelly Easton[1] , George L. Nemhauser[2] and Michael Trick[3]

[1] Kansas State University, Manhatten, KS 66506
[2] Georgia Institute of Technology, Atlanta, GA 30332
[3] Carnegie Mellon University, Pittsburgh, PA 15213

1 Before Publication

The genesis of the 2001 paper "The Traveling Tournament Problem: Description and Benchmarks" (denoted TTP hereafter) [1] occurred more than five years earlier when Trick was approached by Doug Bureman, an executive of the local Major League Baseball (MLB) team, the Pittsburgh Pirates. Bureman had sat for many years on the scheduling committee for MLB. He was about to leave the Pirates and wanted to embark on a consulting career, with scheduling a key component of his business. Bureman knew what MLB wanted in a schedule, but did not know how to create schedules. Trick was referred to him by a student in Carnegie Mellon's MBA class as someone with knowledge of optimization methods. Perhaps those methods could help create schedules.

Scheduling for MLB at the time was done by a husband and wife team named Henry and Holly Stephenson. Most of the work they did was by hand, which was truly amazing given the size and complexity of the MLB scheduling problem. The MLB schedule consists of 2430 games played by 30 teams over approximately 180 days. Each team plays 162 games (half at their home venue and half away), at most one per day, so there are only 18 off days per team across the season. Teams play almost all the time! But teams are spread across the entire United States (and a bit of Canada), so being efficient with team travel is paramount. It would not do to have, for instance, the Los Angeles team play in New York, then next in San Francisco, then in Miami before returning home. The problem is slightly simplified since teams do not play a single game against an opponent but a series of two to four games. Even with this simplification, there are still 780 series to be scheduled.

Travel is only one issue that a scheduler faces for MLB. Teams cannot play too many consecutive home series before local interest wanes; teams cannot play too many consecutive away series without fans losing track of the team. Weekend dates lead to larger attendance, so those need to be balanced among the teams. Summer dates work similarly: no team wants to have less than its share of home games during good weather. Then there are a host of team specific requirements: this team doesn't want to have a home game during the first day of deer hunting season, and that team has to be home for the 4th of July celebrations, and so on. And television partners have strong views of what would be appealing matchups throughout the season. This is just a sampling of the constraints on a schedule: there are many, many more.

Bureman understood all of these constraints and, more importantly, could provide tradeoffs among them. For instance, nearly all teams wanted to play at home on Father's day, but Bureman understood for which teams this was critical and for which other issues were more important.

Trick, as a specialist in the operations research technique of integer programming, promptly embedded this problem into a large IP with predictable results: the resulting model could not be solved with the optimization codes of the time. In fact, this formulation, based on binary variables of the form x_{ijk} (where a 1 denotes team *i* plays team *j* on day *k*, and a 0 denotes not playing) is too inefficient for use to this day.

Trick then reformulated this problem using a more complicated set of variables. In this formulation, there is a binary variable for every possible home stand (consecutive games played in the team's stadium) and road trip (consecutive games played away from that stadium). While this formulation had many more variables (approximately 2 million), the constraints were much easier and the optimization worked better, though still not well.

The approaches Trick took were not anywhere near enough to compete with the schedules created by the Stephensons, who continued to provide the schedules for MLB. But, being an academic, this did not stop Trick from giving talks at other universities and academic conferences on the issue of scheduling MLB.

In 1997, Trick gave one of these talks at his alma mater, Georgia Tech, and the talk was attended by George Nemhauser, one of the founders of the field of integer programming. Nemhauser is also extremely engaged with sports (his first book on integer programming was dedicated to the basketball team, the NY Knicks) and was the faculty representative for his university's athletic conference, the Atlantic Coast Conference (ACC). The ACC was and is one of the very best college basketball conferences in the country, and they too had a scheduling problem. But instead of scheduling 2430 games or 780 series, it only involved 72 games.

Nemhauser and Trick got to work scheduling the ACC and were successful in scheduling the league, starting with the 1998/99 schedule. Their approach involved, fundamentally, the x_{ijk} variables, but decisions were made in three phases: first a set of home/away patterns were found that the teams might play, then games were assigned consistent with those patterns, then teams were assigned to the patterns. The first two steps were done with integer program, while the third was done through complete enumeration. The description of this approach was published in *Operations Research*[2]. and has been about as influential as the TTP paper, as measured by Google Scholar references.

While Nemhauser and Trick successfully scheduled numerous college sports leagues, Bureman and Trick continued to struggle with Major League Baseball. It was about this time that Easton enters the story as a doctoral student at Georgia Tech under Nemhauser's supervision. Nemhauser invited Trick to co-supervise, and Easton began working on a dissertation on optimization approaches to sports scheduling. Shortly thereafter, Easton, Nemhauser, and Trick combined to do all their sports scheduling work, both academic and in practice (with Bureman), together.

The scene is now set for TTP.

2 The Traveling Tournament Problem

In early 1999, Trick decided to write up some of his work on MLB scheduling. Work up to then has always been on solving the full MLB problem. But certain aspects of that work were confidential (team preferences most particularly), so Trick could not simply describe the whole problem. Instead he decided to start with the simplest set of constraints: simply find a minimum travel schedule with limits on the length of home-stands and road-trips and no other constraints. Once this was solved, additional constraints could be added until the reason for the difficulty in solving could be identified. Presumably those could eventually be overcome and the MLB problem could be solved.

To Trick's enormous surprise, his methods, based on the stronger home-stand/road-trip formulation were not sufficient to solve even the base model. Taking 30 teams and finding a minimum travel distance schedule did not seem doable. Even taking 16 or 14 teams (at the time MLB had two leagues with only weak interactions between them) seemed difficult.

Trick looked at smaller and smaller instances of this simplified problem before realizing the only size instance he could solve had just 4 teams! It may have been small, but on May 1, 1999, the first instance of what would be the Traveling Tournament Problem was solved. It involved teams from Atlanta, New York, Philadelphia and Montreal (baseball fans can date the data sets based on the fact that Montreal lost its team after the 2004 season). Here is the solution found (the schedule for each team is read down in each column, so Atlanta begins at home against Philadelphia, New York, and Montreal, before traveling to Philadelphia, New York, and Montreal):

```
  ATL  NYM  PHI  MON
  ---  ---  ---  ---
  PHI  MON @ATL @NYM
  NYM @ATL  MON @PHI
  MON @PHI  NYM @ATL
 @PHI @MON  ATL  NYM
 @NYM  ATL @MON  PHI
 @MON  PHI @NYM  ATL
```

Trick sent the problem definition and data to Easton who had more computing power and more advanced methods, including constraint programming approaches. She was quickly able to confirm the optimality of Trick's four team solution and provide the optimal solution to the six team instance. But the 8 team instance could not be solved to optimality with the methods of the time. In fact, it was not until 2002 that Easton finally found what was to be the optimal solution to the 8 team instance, and not until 2008 that it was confirmed optimal by the lower bounding method of Irnich and Schrempp.

At this point, the problem did not have a name. Trick suggested something like "The minimum travel for a sports league problem". He was wisely overruled by his coauthors and the Traveling Tournament Problem was born.

3 The CP Paper

After working on the problem for some time, we were convinced that the Traveling Tournament Problem was quite a computational challenge, so we decided to introduce the problem at CP 2001. Trick was planning to attend the conference in any case (the location of Cyprus may have had something to do with it).

While we prepared a longer paper, the CP paper ended up being a short, five-page paper. In this paper, we began by formalizing the Traveling Tournament Problem:

Input: n, the number of teams; D an n by n integer distance matrix; L, U integer parameters.
Output: A double round robin tournament on the n teams such that
– The length of every home stand and road trip is between L and U inclusive, and
– The total distance traveled by the teams is minimized.

Teams were considered to begin and end the tournament at their home base.

In subsequent definitions of the problem, a "no repeater" constraint was added, prohibiting consecutive games of team A at B followed by team B at A. It is not clear that this additional requirement added much to the problem, since it seems about as easy to solve the problem with and without the constraint. But the constraint is not irrelevant: for some instances the addition of the no-repeater constraint increases the minimum travel.

For the work we presented, and for much of the work afterwards, L=1 and U=3, matching MLB's preferred range.

We then proposed two classes of instances. In the first, the "Circle Instances", the goal was to have distance matrices where the tour aspects of the problem were easy to solve. In these instances, the venues are arranged around a circle with distance 1 between adjacent teams. The distance between any pair of teams is the shortest of the two paths around the circle between the teams. Given any set of teams, the best cycle through them simply goes around the circle, skipping teams not in the set.

The second set of instances was based on MLB teams from the National League. Instances with 4, 6, 8, 10, 12, 14, and 16 cities were given.

In the paper, we presented optimal solutions for these instances up through size 6, and non-optimal solutions for CIRC8, NL8, and NL16, found through a combination of integer programming and constraint programming methods.

Since the paper was accepted as a short paper, Trick gave a brief presentation on the work at the conference (while being threatened with a water pistol from Toby Walsh if he went overtime), and then stood by (with a poster?) for questions.

4 Aftermath

We feared that the Traveling Tournament Problem would turn out to be too easy. We had nightmares that one of the smart CP people would say "Here are the solutions to all your instances" and we would be embarrassed.

Fortunately, that did not happen. These instances were and are difficult to solve to provable optimality. Trick has kept a web page at http://mat.tepper.cmu.edu/TOURN that has faithfully kept track of every improved solution ever found. Here is the entry for NL12:

```
NL12. 12 teams Data set
Feasible Solution: 143655 (Rottembourg and Laburthe May
2001), 125803 (Cardemil, July 2 2002), 119990 (Dorrepaal
July 16, 2002), 119012 (Zhang, August 19 2002), 118955
(Cardemil, November 1 2002), 114153 (Anagnostopoulos,
Michel, Van Hentenryck and Vergados January 14, 2003),
```

113090 (Anagnostopoulos, Michel, Van Hentenryck and Vergados February 26, 2003), 112800 (Anagnostopoulos, Michel, Van Hentenryck and Vergados June 26, 2003), 112684 (Langford February 16, 2004), 112549 (Langford February 27, 2004), 112298 (Langford March 12, 2004), 111248 (Anagnostopoulos, Michel, Van Hentenryck and Vergados May 13, 2004), 110729 (Van Hentenryck and Vergados, May 30 2007).

Lower Bound: 107483 (Waalewign August 2001), 107494 (Melo, Ribeiro, and Urrutia July 15 2006), 107548 (Mitchell, Trick and Waterer July 31 2008), 108244 (Uthus, Riddle, and Guesgen, Feb 11 2009), 108629 (Uthus, Riddle, and Guesgen January 6, 2010)

Summarizing, the current best solution is 110,729 and the lower bound is 108,629. While it has been almost a decade since the last best solution or bound, people are still working on this problem, as shown by its inclusion in a number of subsequent papers that could not improve on best values.

At this point, the largest of the NL instances solved to optimality is NL10, with the solution first found by Langford in 2005 and proved optimal by Uthus, Riddle, and Guesgen in 2009. The CIRC instances appear no easier, with CIRC10 being the largest solved (references and years as for NL10). We hypothesize that many of the current "best solutions" are, in fact, optimal, but that lower bounding methods have not improved enough to show that optimality.

5 Extensions

While there are a number of papers that directly addressed the Traveling Tournament Problem, our TTP paper would not have been so well-cited if only those were the only follow-ons. The Traveling Tournament Problem and our presentation has been has been enhanced, modified and adapted in dozens of ways. This is not meant to be a full bibliographic survey, but here are a few of the directions the field has gone.

- Different distance matrices. People have looked at constant distances, distances from other leagues, even distances between galaxies!
- Different trip bounds. Particularly well studied is having U be $n-1$, removing any upper bounds on the length of homestands and road trips. There has almost no work where L is anything but 1.
- Additional constraints. The most well studied of these is the "mirror constraint" where the double round-robin occurs in two phases, with the second phase reversing the venue for each game in the first phase.

- Varying the game count. Some leagues have divisions where teams play, say, 2 games within a division and 1 game outside the division, leading to an unbalanced tournament. This can be extended to an arbitrary game count matrix.
- Non-compact scheduling. The schedule length can be extended to more than $2n - 2$ time slots, allowing off-days. This more closely resembles schedules for leagues like the National Hockey League and National Basketball Association.
- Complexity results. While it is clear that the Traveling Tournament Problem is computational difficult, its NP-completeness was not proved until 2011 by Thielen and Westphal. This leads to a search for approximation schemes and other approaches.
- Practical applications. The Traveling Tournament Problem has been used directly to schedule Argentinian Volleyball and a small number of other leagues.
- Other aspects of sports management. It is possible to add umpires/referees and fans to the sports league, leading to the Traveling Umpire Problem and Traveling Fan Problem.

It is truly gratifying to see all the interest and activity inspired by our little five page paper.

6 Back to Major League Baseball

Eventually our methods and computing resources were powerful enough to surpass the Stephenson's (who truly were remarkable schedulers), and we provided MLB with their schedules for 2005-2006 and 2008-2017.

References

1. Easton K., Nemhauser G., Trick M. (2001) The Traveling Tournament Problem Description and Benchmarks. In: Walsh T. (eds) Principles and Practice of Constraint Programming — CP 2001. CP 2001. Lecture Notes in Computer Science, vol 2239. Springer, Berlin, Heidelberg
2. Nemhauser G, Trick M. (1998) Scheduling a Major College Basketball Conference, *Operations Research*, vol 46, 1-8.