Youth Sports Leagues Scheduling

Douglas Moody, Graham Kendall and Amotz Bar-Noy

City University of New York Graduate Center and School of Computer Science (Moody,Bar-Noy), University of Nottingham, UK (Kendall)

dmoody@citytech.cuny.edu, gxk@cs.nott.ac.uk, amotz@sci.brooklyn.cuny.edu

1 Introduction

Youth sports are administered by governing bodies that determine sportsmanship rules, promote the sport, and organize youth participation. Organizations within these bodies may be towns, high schools, sports clubs with international affiliations (e.g., FIFA -Federation International de Football Federation), and religious groups. Each of the organizations sponsor teams in leagues, and provides a venue or fixture. For example youth leagues in the United States include: junior soccer leagues, Little League baseball, inter-scholastic high school basketball, and the Catholic Youth organization (CYO). Youth sports leagues are played worldwide. For example, Little League Baseball is played in 72 countries worldwide within 7,170 leagues, comprising over 2 million players [7].

A Youth Sports League (YSL) consists of divisions (see Appendix A for terms we use in this paper), which are sets of teams grouped by age, gender, and/or level of play. The number of teams in a division can vary, ranging from 4 to 20 teams. Each participant registers with the league to play the same number of games, regardless of division. The schedule for a division is often a round robin tournament followed by additional games against selected opponents from the division in order to meet each team's required number of games. This type of schedule is referred to as "unbalanced" since a team may play one opponent once more than another. The sharing of the organization's venue by its sponsored teams creates a dependency between the division schedules. Two of an organization's teams, possibly from different divisions, cannot host a game at the same time. Hence, the administrator must consider the schedule of all divisions, when creating the master league schedule.

The scheduling of youth sports leagues differs from the professional sports league problem, widely studied in scheduling literature and surveyed by Kendall et al. [6]. Professional sports involve a balanced schedule, with guaranteed availability of the venue. Youth sports leagues play unbalanced schedules, and teams from all divisions must share a venue. A YSL venue can support several games a day, whereas professional sports teams' venues typically only host one game in a day. This venue sharing creates a schedule dependency among all divisions. A professional league with 4 divisions comprising 12 teams, can be viewed as four separate and distinct round-robin tournaments. The same league structure in the YSL must be viewed as one schedule with 48 teams playing an unbalanced schedule. Real world instances of youth sports can include 400 divisions involving 3500 teams, for example the Long Island CYO youth basketball.

In the following section, we will informally define the youth sports league problem presenting its hard and soft constraints, and the objective function. The subsequent section discusses related research in this area, followed by our tiling approach to address the problem. The last section discusses the availability of real-world problem instances and related future work.

2 Problem Definition

The YSL involves scheduling multiple divisions, each containing teams sponsored by a common set of member organizations, across a set of venues. All games are intra-divisional. A season consists of a specific number of games to be played by all teams in the league, regardless of the number of teams in a particular division. If the number of games in a season is 12, then teams in a five team division and an eight team division will all play exactly 12 games. Some divisions are required to play an unbalanced round-robin tournament, where a team will play various opponents a different number of times.

Each organization makes a venue available to the league. The venue has a daily capacity or number of games that can be played on one day. A league may allow games to be played on more than one day of the week, particularly weekends. A season consists of a set of consecutive weeks, to be played on a given number of days per week. A venue's season capacity is the number of games per day that can be hosted by the venue, multiplied by the available days in the season. Divisions also have an associated referee level, indicating the minimum level of the referee's certification. This level is referred to as the referee category. A referee must have the proper level of certification to officiate the game. Referees may choose to officiate games at a lower level than their certification based upon their personal preference. This preference may stem from a goal of providing more on-field rule instruction to the younger age groups. However, in this instance the referee would be assigned a lower level category for the season. Each referee is assigned a category based upon their certification and personal requests. Often, a referee is assigned two games at a venue on a given day. The referee prefers to have the games follow each other to minimize his or her time at the venue. Hence it is desirable to schedule games, requiring the same referee level in succession. A venue's daily schedule should have an even number of games for each referee category to support this concept. This will enable the scheduler to schedule for each referee. A referee may officiate games consecutively from different divisions, if the divisions' referee category for the consecutive games are identical.

The schedules for the individual divisions are combined into a master schedule. The master schedule must address all venue sharing and referee assignment constraints.

A master schedule is to be created for all divisions such that:

1. Teams must play exactly *g* games within *w* weeks, where *g* is the league wide number of games per team, and *w* is the number of weeks of play during the season.
2. Each division will play a multiple round-robin schedule. For divisions requiring unbalanced schedules, teams may play an opponent only one additional time than other opponents.
3. Teams can play multiple games in a week, but only one game per day.
4. All teams from the organization will play at their organization's venue, when designated as the home team.
5. Each venue has a fixed capacity of *s* games per day. This value differs by venue.

1.

The quality of the schedule is evaluated by calculating the number of penalty points within the schedule. Schedules with lower penalty points, are of higher quality. The following section lists the penalty points, also referred to as soft constraints.

1. A team playing more than one game in a week (4 penalty points per game). Teams may play only one game per day, but are permitted to play on multiple days. For example, a team having games on both weekend days would be assigned 4 penalty points.
2. A venue hosting an odd number of games with the same referee category during one day (2 penalty points per referee category). Venues with an odd number of games within the same referee category will prohibit a referee from officiating successive games at the venue on that day.
3. A team playing home (or away) games in consecutive weeks (1 penalty point per week). Teams prefer to alternate home and away games.

The YSL solution must be a feasible solution satisfying all the hard constraints. The objective function of the YSL is to minimize the penalty points of the soft constraints above. Regin [10] shows that for a round-robin tournament, it is not possible to eliminate all breaks, defined as consecutive home or away games in the schedule. Hence, the third soft constraint renders a schedule of zero penalty points for the YSL impossible.

3 Example YSL Instance

Our example problem instance involves three towns, referred to as A, B, C, each sponsoring one or more boys and girls teams, within the league. Table A shows the input parameters necessary for the YSL in our sample league. The boys division contains two teams from town A, and one team from B and C. The girls division has three teams from B, two teams from A and one team from C. Each town maintains one venue with a capacity of two games per day, on two days each week. The season should contain 8 games per team, over a 7 week season. Both divisions have the same referee category. Note that the boys division is a double round-robin tournament followed by additional games causing an unbalanced schedule. The girls division has a single round robin tournament followed by additional games resulting in an unbalanced schedule. The two divisions share a venue provided by each town.

Table A provides the set of inputs needed to define a YSL problem instance. Each row contains an input parameter, or a set of information about the league's structure.

Input Parameter	**Value**	**Definition**
G	8	Nubmer of games to be played by all teams
W	7	Number of weeks in the season
Y	2	Number of days per week for play
V	3	Number of venues
D	{b,g}	Values identifying each division.
R	L1	Referee categories
T*v*	A(b-2,g-2) B(b-1,g-3) C(b-1,g-1)	Organizations and number of teams sponsored by division. For each organization, the number of teams per division identified in ***D*** is provided. The value is specified by Organization, with the division-team number for each team in parentheses
C*v*	{2,2,2}	Capacity for each venue. All venues can host 2 games per day.

Table A – Sample problem input parameters and data

Table B provides the rounds necessary to schedule the master schedule for this youth sports league. A team is referred to by its town (A,B or C), followed by the sequence number of the team within the town. The team reference of "A2" would indicate this team is the second team sponsored by town A. For the Boys division, the first round shown (A1-B1, A2-C2) has been placed in week 2 of the master schedule. Each team plays an unbalanced schedule. Team A1, in the boys division, plays B1 and C1 three times while only playing A2, twice. Each box of the master schedule represents a venue's games for a day, whose maximum is two in our example instance. In weeks, where only one day is used by a venue, the choice of the day is not consequential.

The penalty point calculations are shown at the bottom of Table C. Since the number of games in a season is less than the number of elapsed weeks, it is guaranteed that penalty points will be earned for each team due to the soft constraint of playing multiple games in a week. Since both divisions have the same referee category, venues hosting an even number of games, will not incur a penalty point for non consecutive games by referee category. For example, in week 2, A2-C1 was scheduled on the same day as A4-C2 to avoid having an odd number of games on each day.

Boys	**A1-B1** **A2-C1** **-2-**	**A1-A2** **B1-C1** **-4-**	**C1-A1** **B1-A2** **-1-**	**B1-A1** **C1-A2** **-7-**	**A2-A1** **C1-B1** **-7-**	**A1-C1** **A2-B1** **-3-**	**A1-B1** **A2-C1** **-5-**	**B1-A2** **C1-A1** **-6-**
Girls	*A3-A4* *B3-B4* *B2-C2* *-7-*	*B4-A3* *B2-B3* *C2-A4* *-2-*	*A3-B3* *A4-B2* *C2-B4* *-3-*	*B2-A3* *A4-B4* *B3-C2* *-6-*	*B4-B2* *B3-A4* *A3-C2* *-1-*	*A4-A3* *B4-B3* *C2-B2* *-4-*	*A3-B4* *B3-B2* *A4-C2* *-2-*	*B4-A3* *B2-A4* *C2-B3* *-5-*

Table B – Tiles (Rounds) for the example league divisions

Venue /Week	1		2		3		4		5		6		7	
A	**C1-A1** **B1-A2**	*B3-A4*	*B4-A3* *C2-A4*				**A1-A2** *A4-A3*		*B4-A3* *B2-A4*		**B1-A2** **C1-A1**	*B2-A3*	**B1-A1** **C1-A2**	**A2-A1** *A3-A4*
B	*B4-B2*		**A1-B1** *B2-B3*	*A3-B4* *B3-B2*	*A3-B3* *A4-B2*	**A2-B1** *C2-B4*	*B4-B3* *C2-B2*		**A1-B1** *C2-B3*			*A4-B4*	**C1-B1**	*B3-B4*
C		*A3-C2*		**A2-C1** *A4-C2*	**A1-C1**		**B1-C1**		**A2-C1**		*B3-C2*			*B2-C2*

Penalty Points Total = 95 points

1) Multiple Games in a week: all Boys Teams week 7, all Girls Teams week 2 10 occurrences (40 points)
2) Days with odd number of Games in Referee Category: 12 days (24 points)
3) Playing Consecutive Home or Away games A1(4),A2(2),A3(3),A4(3), B1(1),B2(4),B3(3),B4(4), C1(4),C2(3) = (31 points)

Table C – Sample Problem scheduling grid

4 Related Work

A great deal of research in sports scheduling has been aimed at professional sports leagues. Kendall et al. [6] provides a broad survey of various sports leagues and their scheduling challenges. Leagues are often structured as single or double round-robin tournaments. The scheduling challenges revolve around the quality of the schedule, to achieve certain objectives. Norhona et al. [8], Ribeiro and Urrutia [11], consider fairness in South American football leagues, Kendall [5] seeks to minimize travel distances in the English Football league, and Rasmussen [9], Goossens and Spieksma in [4] consider various venue availability constraints. The instances in these problems can be reduced to tournament schedules of 20 to 24 teams, with consideration for some unique constraints and minimizing travel distance. Only Kendall [5] introduces a relationship between different divisions due to the pairing requirement (sets of venues that cannot host a game on the same day). In all these professional sports problems, a venue can be used for only one game in a day.

YSL are most closely associated with the Travelling Tournament Problem (TTP) described by Easton et al. [3]. The TTP and YSL share constraints, with the following exceptions:

1. The YSL may schedule multiples games per week, on separate days.
2. A YSL team is not required to play in every week.
3. Teams share home site venues, which support several games per day.
4. A YSL division may play an unbalanced tournament schedule.

These differences have a profound effect on the instance size, within real-world applications. The TTP instances documented on the problem definition web site [12] involve a maximum of 32 teams, albeit there is no real-world example for that size of tournament. The YSL has common occurrences of leagues with 50 divisions, comprising over 500 teams. The CYO of Westchester-Putnam, near New York City, has 68 divisions, comprising 582 teams sponsored by 58 parishes.

5 Proposed Approach

Our approach is a two phase approach based upon "tiling". Each tile represents a round of games for a division, requiring a total number of tiles equaling the number of divisions multiplied by the games per season. Each tile is one round of a tournament for a division. The tiles are placed in the master schedule in a greedy fashion. Each tile is placed in a week that results in the least number of penalty points being added to the objective function. Tiling continues until a tile cannot be placed without causing a hard-constraint violation. At this point, the venue capacity constraint is relaxed, and the remaining tiles are placed in the schedule. These final placements continue to use greedy approach to minimize total penalty points.

Once all tiles are placed, we analyze each game's contribution to the penalty points. Games violating the venue capacity hard constraint are temporarily assigned an artificially large penalty point value to prioritize them, within our local search phase. The schedule is now ready for the second phase, a local search. This phase consists of a set of swaps to reduce the current penalty points of the schedule. The following sections discuss the steps of these two phases in more detail.

5.1 Tiling Phase

For each division, we use a single round-robin tournament to produce n-1 rounds of opponent pairings, where n is the number of teams within a given division. We use one of a variety of single round-robin tournament method generators discussed in [2]. The YSL requires a number of teams, g, for each team to play. Hence, g / (n-1) is the number of round-robin tournaments needed for each division, The YSL will frequently have an unbalanced schedule. This occurs when g / (n-1) is not an integer. These divisions require using a partial round-robin tournament to complete the season. When divisions require more than a round-robin tournament, a mirrored round-robin tournament is used for the second tournament. We may use a round-robin tournament definition and its mirror several times within a division's schedule. Table D provides examples of divisions with unbalanced schedules:

Teams in the Division	Games per season	Full Round-Robin Tournament usage	Additional Tournament rounds
10	9	1 Round-Robin	None
6	12	1 Round Robin 1 Mirrored Round-Robin	2 rounds from round-robin
7	20	2 Round-Robin 1 Mirrored Round-Robin	2 rounds from Mirrored round-robin

Table D – Unbalanced Schedule in terms of tournaments

Each round of a tournament for a division is a tile. We apply a cost to each tile based upon the percentage of venue capacity used by each game. Tiles that contain several games with home teams in highly constrained venues will have a higher cost. A highly constrained venue is a venue where the usage rate ((number of teams/2) / daily slots), over the season, is higher than other venues. For example a venue sponsoring 6 teams, which can host 4 games a week, has a 75% usage rate. The rate suggests that the venue will be using an average 75% of its capacity each week. A venue usage rate over 100% indicates that a feasible schedule is not possible due to venue capacity.

Tiles with the highest costs are the first tiles to be placed in the schedule. For each week the tiles placement cost is calculated by determining the number of penalty points created by the tile's assignment for a given week. The week with the smallest placement cost, is chosen for the tile's placement. Tiling continues until the remaining tiles cannot be placed without causing a hard constraint violation of venue capacity.

The venue capacity constraint in relaxed to allow "over booking" of a venue. An artificial penalty point value of 9999 is used to enable the tile placement to continue. This temporary penalty assignment will highlight these games during the remaining tile placements and our local search phase.

5.2 Local Search Phase

The second phase involves a local search and seeks to remove hard constraint violations and reduce the number of penalty points in the existing schedule. The swaps in the YSL are similar to those discussed by Anagnostopoulos et al. [1] for solving the TTP. All swaps are done between teams in the same division. Each swap maintains a feasible schedule for each division. The swaps are:

Home / Away Swap – We swap two games, involving the same two teams, where each team is home in one game and away in the other. The swap is done between two weeks. The impact of the swap is a reduction in the number of games for one venue and an increase for the other venue. This swap is only effective in the YSL for teams playing their home games at different venues.

Round Swap – All games for two rounds of a division schedule are moved between two weeks. All venues hosting a game in either week are affected by the swap.

Partial Round Swap - The partial round swap moves a set of connected games between two weeks. We begin this operation by selecting two games in different weeks. The teams in these games create our swap set. We then add teams that are playing a team in the team-swap set to that set. The process continues until all teams, in both weeks, are in the swap set or its complement. We then move all games involving teams in the swap set between weeks. We can also choose to move all teams in the complement of the swap set between weeks.

We perform the local search in a structured fashion to reach the local minima. In the first step, we consider all games at the venue during the week that has a hard constraint violation, indicated by the artificially high penalty points. We calculate the reduction in penalty points, if any; of performing a Home/Away swap for each game. We also calculate the reduction in performing a round swap for each round. The most beneficial swap is chosen and executed.

In the second step, we analyze all games, involved in producing penalty points. Each possible home and away swap, and each possible round swap for each game is analyzed for its potential reduction in penalty points. All partial round swaps are identified for every round. The partial round swap considers both the swap set described above, and its complement set. The most beneficial swap among the three types of swaps is chosen for execution. The swap is performed and the resulting Master Schedule will have a lower penalty point total. If no improvement is found, processing is halted as the schedule has reached its local minimum. Figure 1 below presents the algorithm as pseudo-code.

1. *Tile creation*
    a. For each division generate a round-robin tournament and its mirrored image
    b. Create a tile for each round of the table until G tiles are generated, alternating the tournament table with its mirror.
    c. Assign a cost to each tile based upon the venue capacity usage of each game.
2. *Tile Placement*
    a. For each tile sorted by cost, find all weeks where the tile may be scheduled with no hard constraint violations. If no week exists, relax the venue capacity constraint and assign artificially high penalty point value for that week.
    b. Place all games in the tile within the week causing the fewest penalty points.
3. *Local Search: Home / Away swap*
    a. Set S′ to be the schedule of swapping home and away assignments for two games with the same opponents.
    b. Consider all games with the same opponents and find S′ with the minimum penalty points.

c. If S′ reduces penalty points, then swap home and away assignments for the games and continue step 3.

4. *Local Search: Round swap*
   a. Set S′ to be the schedule of swapping all games in one week with all games in another week.
   b. For each pair of weeks, calculate S′
   c. Find the S′ that minimizes the number of penalty points
   d. If S′ improves the schedule, swap rounds and repeat step 4.
5. *Local Search: Partial Round swap*
   a. For each pair of weeks, create sets of opponents who play each other in week 1 and / or week 2.
   b. Set S′ to the schedule of swapping each set of teams between the weeks.
   c. Find the S′ that minimizes the number of penalty points
   d. If S′ improves the schedule, swap rounds and repeat step 5.

Figure 1 – Tiling by Round Algorithm

6. **Results**

We will present our results at the conference. The results will include solutions to the sample instances shown in Table A, as well as the results from real-world instances of youth sports leagues, comparing the schedules actually used with our approach.

7. Future Work

In this paper, we have described a sports scheduling problem widely faced by thousands of organizations. This problem has not been reported before in the scientific literature. A set of professional-sports scheduling problems, including the TTP, have similarities to the YSL. However, key differences arise such as: sharing venues among several teams, significantly larger problem instances, and unbalanced schedules. These add real-world complexities that will challenge current approaches. We look to document problem instances and results for the community similar to the TTP [12]. We are currently working with several youth leagues to help in defining these world problem instances.

References

1. Anagnostopoulos A., Michel L., Van Hentenryck P., Vergados Y. A simulated annealing approach to the traveling tournament problem. Journal of Scheduling 2006;9:177–93.
2. Dinitz J.,E.Lamken, and Wallis, W.D. Scheduling a tournament. Handbook of Combinatorial Designs, pages 578-584. CRC Press, 1995.

3. Easton K., Nemhauser G.L., Trick M.A. The travelling tournament problem: description and benchmarks. In: Walsh T, editor. Principles and practice of constraint programming. Lecture notes in computer science, vol. 2239. Berlin: Springer; 2001. p. 580–5.
4. Goossens D, Spieksma F. Scheduling the Belgian soccer league. Interfaces 2009; 39:109–18.
5. Kendall G. Scheduling English football fixtures over holiday periods. Journal of the Operational Research Society 2008; 59:743–55.
6. Kendall G., Knust S., Ribeiro C. C. and Urrutia S. Scheduling in Sports: An Annotated Bibliography. Computers & Operations Research (2009), 37: 1-19
7. Little League Home website. Little League history. Online documentation at: http://www.littleleague.org/Learn_More/about/historyandmission/aroundtheworld.htm
8. Noronha T.F., Ribeiro C.C., Duran G., Souyris S., Weintraub A. A branch-and-cut algorithm for scheduling the highly-constrained Chilean soccer tournament. Practice and theory of automated timetabling VI. Lecture notes in computer science, vol. 3867. Berlin: Springer; 2007. pp. 174–186.
9. Rasmussen R.V. Scheduling a triple round robin tournament for the best Danish soccer league. European Journal of Operational Research 2008; 185:795–810.
10. Regin J.C. Minimization of the number of breaks in sports scheduling problems using constraint programming. DIMACS series in discrete mathematics and theoretical computer science, vol. 57, 2001. p. 115–30.
11. Ribeiro C.C., Urrutia S. Scheduling the Brazilian soccer tournament with fairness and broadcast objectives. Practice and theory of automated timetabling VI. Lecture notes in computer science, vol. 3867. Berlin: Springer; 2007. p. 147–57.
12. Trick, M.A., Challenge Traveling tournament instances. Online document at http://mat.gsia.cmu.edu/TOURN/.

Appendix A. Glossary

Capacity: The number of games that can be played at a given venue on any day. The capacity may vary by venue.

Division: A set of teams where every team in the division plays against every other team a set number of times. Subsets of these teams may play one additional game amongst them to reach the required number of games.

Game: A sports contest between two teams.

League: A set of divisions, comprised of teams, which share a common set of venues.

Organization: An entity that maintains a venue and sponsors teams across many divisions.

Round-robin tournament: A round-robin tournament involves *n* teams playing *n*-1 games. Each game is played against a different opponent in the tournament. Also, every team plays a game in each round.

Slot: A time period during a given day to be used to play a game at a venue. The YSL schedules a fixed number of slots per day at each venue, depending upon the venue's daily capacity. The actual clock time of the slot is not considered in the problem.

Unbalanced Schedule: A schedule where a team will play one opponent more often than another opponent. The YSL requires unbalanced schedules in divisions where the required number of games in a season is not a multiple of the number of n 1, for a division.

Venue: A physical location for a game to played.