CORNERSTONE
MINNESOTA STATE UNIVERSITY MANKATO

Proceedings of the National Developmental Conference on Individual Events

Volume 5 | Issue 1 Article 9

September 2020

Towards a Novel Tournament Scheduling Algorithm and Statistical Measure of Team Equity in Large Scale Forensic Tournaments

Mark Kokoska
Ohio University

Towards a Novel Tournament Scheduling Algorithm and Statistical Measure of Team Equity in Large Scale Forensics Tournaments

Mark Kokoska
Ohio University

Abstract

Business items raised at recent national forensics tournaments regarding scheduling seem to be based around issues of an increased number of competitors and a trend for some schools to have a disproportionate amount of competitors in a single event. This paper examines the stated goals in scheduling a tournament as a device that measures the ability of competitors. Stemming from these goals, this paper proposes a novel random scheduling algorithm capable of scheduling a large number of competitors in an individual event. After implementing this algorithm, its performance is measured in relation to its ability to schedule a tournament comparable with given national level competitions. This paper further suggests that there may be a need to establish a method for measuring the fairness of a schedule. This paper concludes with the recommendation that the means of scheduling the most important tournaments be reexamined and that tournaments describe the means by which they will be scheduled.

At the 2008-2009 National Forensics Association (NFA) spring business meeting a motion was made to discuss limiting the entries per event for each team. While those speaking against the motion indicated the commitment of NFA to inclusiveness, those in favor suggested that the size of the tournament made it difficult to schedule. While the motion to discuss was eventually tabled, discussion about the act of scheduling a tournament, especially one as daunting in size as a national level competition, revealed that scheduling was an inherently conflicted task.

Scheduling a large tournament forces the tournament staff to find a balance between catering to the individual and catering to the team. A tournament simultaneously functions as an assessment tool to find the best competitor in each event, the best competitor overall and the best team overall. It is possible to see how the ability to assess individuals and the ability to assess teams might conflict by examining the decision to break brackets and stop competitors from hitting their own teammates. In general breaking brackets measures team success more accurately but measures individual success with less precision.

Additionally any schedule is a balance between time and efficacy. While there are many means of scheduling a tournament, any method of scheduling is improved the more time the tournament staff spend. Some methods are improved simply by checking and double checking and time spend by tournament staff ensures the basic goals of the tournament as an assessment tool are met. However, if large tournaments had clearly stated goals for fairness and balance, even a fully double checked tournament schedule could be completely rescheduled from a different starting point. Comparing two possible schedules for the same event and choosing the better one insures that more time could always create a better schedule. This means that the tournament staff are always finding a balance between the amount of time they can afford and creating a fair and balanced schedule.

My personal interpretation of the 2008-2009 business meeting and the general disposition of the forensics community suggests that fairness of scheduling is interpreted as having three key components. The first component of fair preliminary rounds is that competitors should not hit competitors from their own team. It is worth mentioning that this criteria cannot be met if any individual school has more students entered in an event than the total number of rooms of competition in the event. The second component is that an individual should not hit another individual more than once in the preliminary rounds of any given event. While at a national tournament this criteria can easily be met, there is the potential that a small tournament will prevent this type of criteria from being met. The third component is that all other decisions, after satisfying the first two rules, should be as random as possible to avoid human intervention. The first two criteria are well established norms within the forensics community that are considered to be best practices. The third criteria stems from the fact that the people scheduling the tournament are members of the forensics community, inevitably bringing with them to the tab room their own expectations and bias, and the desire to make a fair schedule depends on the ability of the scheduling process to isolate the decisions of the staff from the process.

These criteria are well established, frequently voiced by both competitors and tournament staff as valuable, and work well within the forensics community. However, it is already clear that there are several difficulties that intrinsically present themselves in tournament scheduling. For example, it seems paradoxical that tournament staff could decide how to establish the balance between time and efficacy while completely isolating themselves from the decision in the name of randomness. This paradox shadows similar concerns that, "scheduling ... is problematic because judgment calls, peer scheduling" (Littlefield, 1986). I propose a new means of scheduling that makes the job of placing competitors in rooms of six over a period of three or four rounds that makes an effort to resolve the problem of the level to which tournament staff are involved and cater to the ideal of a random schedule. While I admit that scheduling a large tournament takes a lot more than simply figuring out the ordering of competitors in each round, for example the scheduling of which rooms to use and which judges to use, the arrangement of all of the competitors at the tournament seems significantly daunting in large numbers and is the focus of the method presented here. In addition to proposing a new scheduling algorithm this paper attempts to makes suggestions to help complete create what Littlefield and Sellnow (1992) call healthy competition and to help "create a shared vision of what a tournament experience should include for healthy competition (i.e., well scheduled, well managed)" (Hatfield, Hatfield & Carver, 1989).

An Example Current Scheduling Method

The most current published description of the methods for scheduling a large forensics tournament is Peters (1983) description of the NFA grid scheduling system. The NFA scheduling system is described as revolving around sets of six by six grids. Students are first ordered and anonymously transformed into numerical codes independent of any identifying information. Codes are then placed in multiple six by six grids in such a way that students from the same team follow a preset pattern. If these patterns are followed then four rounds of six individuals, in which competitors do not hit the same competitor and do not hit their own teammates, can be scheduled by using each group of thirty-six's rows, columns, diagonally left right and diagonally right left groups. This method clearly meets the first two scheduling constraints, that students cannot hit their own team and cannot hit competitors twice in the same event. Additionally the anonymous transformation and use of numerical codes at the starts attempts to scramble the individual competitors and prevent human intervention.

This method is surprisingly effective at scheduling a large tournament efficiently, in terms of both time and successfully meeting the preestablished scheduling criteria. An interesting interpretation of the amount of time needed to schedule a tournament illustrates the value of a by hand method like this. One way to consider the amount of time to schedule a tournament is to ask how much it would take to schedule if one additional competitor was added. Because the grid system simply works in independent interchangeable blocks of thirty-six, scheduling each separate grid should take the same amount of time as the previous. Thus while each additional competitor adds a burden of an identical amount of time. It is possible to imagine a hypothetically more complex method in which the entire tournament was considered at once and each competitor added exponentially more time to the equation instead of a steady increase. This analysis is analogous to the means of describing complexity frequently found in computer science and referred to as big O notation (Knuth, 1976), the advantage of the grid system is that for n competitors it has a linear complexity O(n).

However, while the grid system swiftly and efficiently meets the first two scheduling criteria, it also illustrates an interesting problem with the issue of randomness. While the grid system may appear to be random it is a partially imperfect system. The grid system can never be truly random because initial placements and team dependent placements in the grid rely on some human intervention. Additionally the process that occurs to create four preliminary rounds insures that the schedules for each round are not independent of each other, a requirement for true randomness. While the process of using arbitrary numbers for individuals should cut down on human intervention, and other dangers of a non-random schedule, this is not the same as completely meeting the communities expectations for a truly random tournament.

Verbal Slippage and a Random Schedule

A significant portion of the issue relating to the scheduling of a tournament stems from the fact that the term random, especially as used by the community in this instance, has multiple meanings. For example, in the 2008-2009 business meeting, some individuals referred to the grid system as being a random method because it had a random initial condition while described the structured process that scheduled round one and round two. It seems contradictory to be able to predict what will happen from round one to round two successfully and for the schedule to be considered completely random. In order to illustrate what I believe the true goal of the community is, complete randomness, I will examine four words which I believe are all being used interchangeably as the definition for randomness in the discussion of forensics tournament scheduling.

The first definition to consider is that random means any situation that has a probabilistic outcome. For example, rolling a fair die clearly is random under this definition because before the roll the exact outcome is not known and it ultimately will be somewhere between one and six. However, this interpretation also seems to refer to systems in which different outcomes have different probabilities. For example, rolling two dice is still "random" because the sum of the faces is not predetermined but there is a higher chance that a seven will be rolled and a lower chance that an eleven or two will be rolled. Even though these might seem to be two different situations, in both situation the outcome is undetermined prior to the rolling the dice and leads to an interpretation of random as anything where probabilities determine the result. In terms of a schedule a probabilistic schedule would be any schedule in which some kind of shuffling or randomizing process was used at any point in the scheduling regardless of what tools the rest of the process employed.

The second term the community frequently seems to employ as the definition for the word "random" is better referred to as pseudo-randomization. Pseudo-randomization is best thought of as having the appearance of being random regardless of what the actual underlying methods of determining outcomes are. For example, instead of rolling a ten sided die one hundred times in order to to choose random numbers, it might be quicker to simply use the first one hundred digits of the number pi. To an individual who didn't have the first one hundred digits of pi memorized, this would appear to be the result of a random process, as the numbers in pi are fairly well scrambled. It is easy to see why this definition of a pseudo-random process is frequently used for the word "random" in casual conversation because it is based on appearance to the observer. In the terms of scheduling a forensics tournament if the tournament looks scrambled to the competitors and coaches than it is pseudo-random and in casual conversation might be referred to as "random."

Obfuscated, or in conjunction the presence of too much complexity to grasp, is the third interpretation that is sometimes substituted as a definition of the "random" in the phrase a "random schedule." Obfuscated simply means that the underlying process, regardless of what the result looks like, is hidden from the observer. A classic example of obfuscation is referred to as a black box, whatever happens in the black box is obscured from the outside world and any numbers this mysterious box might produce could be the result of a die, a coin toss or one hundred monkeys at typewriters. Frequently this interpretation is employed if things appear too complex or difficult to understand, and thus are made as if a black box to the viewer. Scheduling a large tournament involves arranging a huge number of individuals into multiple rooms over multiple rounds, a process this complex is almost automatically dubbed "random" under this interpretation.

However, I believe that the best interpretation for "random," and the definition that best meets the needs of the community, is a uniformly random distribution. This interpretation is best thought of in contrast to the first interpretation which said that any outcome that is based on probability is random. Uniformly random refers only to probabilistic events that have an equal chance of occurring. For example rolling one die is uniformly random, as one through six are equally likely, but rolling two dice is not uniformly random, because seven has a higher probability of occurring. A uniformly random forensic tournament scheduling process would have an equal chance of arriving at any possible schedule that met the criteria. Examining the grid system again, while it clearly is probabilistic to some extent, looks scrambled to observers and is both complex and happens behind closed doors, it is clear that it does not meet the criteria of uniformly random. Many possible schedules are excluded having a decidedly unequal probability of occurring. For example there can be no schedule where competitors A and B are in the first grid of thirty six students and competitors X and Y are in the second grid of thirty six students and the two pairs compete against each other. I believe that uniformly randomness is the interpretation that the community should embrace as it intrinsically creates the most balance by allowing every possible outcome to occur.

Abstract Scheduling Process

I next developed an ideal scheduling mechanism based on two central ideas, that constraints of the tournament must always be met and that uniform randomness should be privileged as much as possible. The same constraints of the tournament, that no individual competes against an individual from their own team, and that no individual competes against the same person more than once in preliminary rounds of the same event were employed. These criteria are held paramount and the scheduling mechanism is designed to meet these constraints 100% of the time. Because the mechanism is designed to be automated by a computer, randomness is handled by the computers' internal processes. Uniform randomness is employed on the level of the individual, such that whenever an individual needs to be chosen to be placed into a room, every possible individual has an equal chance of being selected. This ensures that the third criteria of uniform randomness is met by the algorithm.

A generalized description of the process is represented by a decision tree (Fig 1) which represents actions as circles and decisions as diamonds. To schedule the algorithm gos through each room in each round, for each room an unconstrained individual is selected at random and placed into the room. If ever in the process an individual needs to be selected to fill a room but there are no individuals who can be placed into the room due to scheduling constraints all scheduled individuals are cleared and the process is restarted. An alternative to this process would be to remove the last individual scheduled backtrack through the schedule in an attempt to free up unconstrained individuals. However, it is unclear how the backtracking effects the uniformly random outcome of the schedule so I have opted to start over anytime there are irreconcilable conflicts.

In order to identify conflicts the process maintains a list of all individuals entered in the event and a corresponding list of blocked competitors for each individual. Thus the set of all constraints in a tournament can be thought of as a set of corresponding pairs of individuals and lists of blocked competitors. At the start of the scheduling process, every individual who shares a team with someone is placed on their blocked list. As the scheduling process continues, every time someone is added to a room they are added to each person's blocked list, and each person in the room is added to their list. Whenever the scheduler resets, the list is reverted the list that contains only the constraints due to school affiliations.

Given the collection of blocked individuals, the easiest way to perform a random selection is to maintain two lists, a list of all unscheduled individuals in the round, and a temporary list of unconstrained individuals for that room. Whenever starting to schedule a room a person is randomly chosen from the list of available individuals for that round, then they are removed from that list. Whenever adding people to a room that already has people in it, a temporary list is made that is the the list of available people in the round with all constrained people removed from it. A randomly selected person for a room that already has people scheduled in it is chosen from this temporary list. This second temporary list also provides a mechanism for testing if irreconcilable conflicts exist, if ever an individual needs to be entered in a room but the list of unconstrained individuals is empty, because all people left in the round have been struck from it, then the scheduling process must start fresh. An example of two steps in this decision process, and the correspondingly maintained and updated lists is included as figure 2.

Implementation

After designing the scheduling process it was implemented using the Java programming language. Java was chosen for both familiarity and computability as it can be run on all operating systems and even in many web applications. Because the scheduler can be thought of as a theoretical model of a tournament, I followed software design process borrowed from Gilbert and Troitzsch's *Simulation for the Social Scientist*, which included the following steps: definition, observation, verification, validation and sensitivity analysis. The definition and observation steps involve selecting the target, a successful tournament schedule, and observing its important elements, namely that it is uniformly random and meets the necessary constraints. After coding the scheduling algorithm I began be performing the process of verification. Verification is essentially debugging the program, I confirmed that all the lists were being created and maintained by printing them out at each time step. Additionally I verified that the program when given an input produced an output that looked like a schedule, the functional elements of the scheduler clearly passed visual inspection.

The process of validation was performed using data from the 2008-2009 National Forensics Association national tournament, the 2009-2010 National Forensics Association national tournament and the 2009-2010 American Forensic Association National Individual Events Tournament. These tournaments were selected as sample entry data because the results had been sent in a digital format making it relatively easy to create a list of all competitors entered in an event and because they present different situations across both time and tournament. Next I selected two events to serve as benchmarks for difficulty. In general across the selected data Prose was the largest event and Rhetorical Criticism/Communication Analysis was the smallest. Additionally these events tended to be those entered to a high level by specific schools mimicking the problem that initiated the entire discussion, individual schools with nearly as many entries as the number of rooms in the event. The 2008-2009 prose data was selected as the final by hand verification and was entered into the scheduling program. The resultant schedule was hand checked to confirm that it was complete and did not violate the given constraints based on team membership and previous rounds. The results of this process suggest that the implementation of the scheduling algorithm successfully schedules an event according to the rules that have been provided.

Finally I tested the sensitivity to initial conditions, in this case initial conditions are the set of all individuals, and their team affiliations, to be scheduled. To do this I began to track the number of times the scheduler reached a set of conditions that forced it to restart before it found a valid schedule and the approximate time taken for to reach a valid schedule. Because of the random method of the scheduler, given a set of individuals that can be placed into a valid schedule, the algorithm will eventually find it. So the measurement of restarts and time represent assessments of the amount of time needed to find a valid schedule. Once these measurements were established, the data from the selected tournaments was entered and scheduled such that one hundred valid schedules were found for each. For each valid schedule I recorded the amount of time in seconds and the number of times the algorithm had to start from scratch. It is worth noting that the actual time taken is dependent on both the computer being used and the other tasks the computer is performing. This being said these values represent a possible amount of time it might take to schedule an event. Additionally there is a linear relationship between the number of restarts and the time taken implying that number of restarts will correlate with time on any machine, and that we can consider either number to be a rough measure of the difficulty of scheduling an event. The mean of the times required to produce a single valid schedule for Prose and Rhetorical Criticism/Communications Analysis are represented below.

The results suggest several conclusions about the effectiveness of the algorithm in different conditions. First the significantly faster scheduling of AFA events, which typically have a smaller number of competitors per school due to tournament entry limits, suggests that the constraint of competitors per school is the most difficult to deal with. This is further illustrated by the generally increased difficulty of Rhetorical Criticism/Communication Analysis in comparison to Prose. RC/CA in general have fewer total competitors but more competitors per school creating difficult scheduling scenarios.

Tournament	Mean Time
NFA 2009 Prose	28.5 seconds
NFA 2009 RC	187.8 seconds
NFA 2010 Prose	9.8 seconds
NFA 2010 RC	186 minutes
AFA 2010 Prose	.9 seconds
AFA 2010 CA	.7 seconds

Suggestions

The most obvious suggestion from this analysis of tournament scheduling is that tournaments should be more open and transparent with their scheduling mechanisms. Not only will this help create more fair and well understood tournaments, this helps eliminate the illusion of both pseudo-randomness and obfuscated as being actually “random.” The movement of the entire forensics community towards a unified definition of random helps to create a single unified assessable goal. Once that goal is determined the best ways to meet it can be constructed. I argue that if the goal of a tournament is to be uniformly randomly scheduled than the process presented and tested here is equivalent to the best possible option.

The further suggestions of this paper are to consider creating mathematical models for measuring the randomness of the tournament. Quantifiable tournament metrics could take multiple forms, but I suggest that all should in some way measure the distribution of the number of times each team competes against each other team. This is partially because I believe that the first two scheduling constraints are designed to regulate the measurement of the success of individuals, but few constraints exist to protect the assessment of team quality. Measurement of the distribution of the number of times each team hits another team could be performed as simply as with a measure of variance. However, variance provides the a problem in that each team has a different number of competitors and thus it would not be expected that each team actually hit each other team the same number of times. The problem this produces is that it does not allow much in the way of comparison between schedules because the expected variances would actually be radically different.

Instead of variance, one possible measure of distribution would be to perform a chi-squared test for proportions on any schedules number of times any competitor from a team competed against a competitor from another team. While this produces a probability, and thus doesn't completely solve the problem of comparisons posed by variance, the community could arrive on a standard necessary for their tournaments. For example, to be a valid schedule it could be proposed that the collision of teams must have a greater than 95% chance of occurring by random chance.

The final suggestion of this paper is to explore more partially-deterministic, non-uniformly random, scheduling methods such as the grid system. For example, if the community decided that equalizing the number of team collisions was a top priority, a method of manual forcing teams to collide with each other team at the tournament, while still randomizing individual competitors, could be constructed. This could help meet the dual criteria of balancing the individual and the team in addition to balancing the criteria of time and human effort.

If a forced team collision model is not satisfactory to the community, but measurement criteria similar to variance or probability are determined, another option might be to employ a mass scheduling system. If one hundred or one thousand schedules were produced for an event and then mathematically compared to each other, the best produced schedule could be produced that could be interpreted as the most fair by the communities collective standards.

In conclusion, the discrepancies between interpretations of the word random and how it functions as a criteria of successfully scheduling large forensics tournaments has generated a useful and fruitful discussion regarding automated scheduling. The algorithm proposed and tested here randomizes every possible decision and successfully automates the scheduling process in a fraction of the time that is needed for traditional by hand, and less uniformly random, scheduling methods. I recommend tournament directors consider establishing the criteria they wish to meet in scheduling their tournaments, and if uniform randomness is a valued criteria, then I suggest the deployment of a system similar to the one discussed here.

References

Gilbert, N., & K. G. Troitzsch. (1999). Simulation for the Social Scientist. Open University Press, London, UK.

Hatfield, S. R., Hatfield, T. L. & Carver, C. (1989). Integrating wellness and forensics: Tournament management as a starting point. In L. Schnoor & V. Karns (Eds.), Perspectives on Individual Events: Proceedings of the First Developmental Conference on Individual Events (pp. 26-29, 32). Mankato, MN: Speech Communication Department of Mankato State University.

Knuth, D. (1976). Big omicron and big omega and big theta. *ACM SIGACT 8(2)*, 24

Littlefield, R. S. (1986). The use of the undergraduate student in tournament management: Someethical considerations. The Forensic of Pi Kappa Delta, 72(1), 1-5.

Littlefield, R. S. & Sellnow, T. L. (1992). Assessing competition and stress: The perceived effect of tournament atmosphere on students and coaches. National Forensic Journal, 10, 1-10.

Peters, D. F. (1983). Scheduling for individual events nationals: The N.F.A. experience in developing fairness safeguards. National Forensic Journal, 1(1), 33-42.