Student Admissions in Hungary as Gale and Shapley Envisaged

Péter Biró

Department of Computing Science
University of Glasgow
Glasgow G12 8QQ
UK

Technical Report
TR-2008-291
October 2008

University of Glasgow | Department of Computing Science

Student Admissions in Hungary as Gale and Shapley Envisaged

Péter Biró*

Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK.
Email: `pbiro@dcs.gla.ac.uk`.

Abstract. Student admissions, for both secondary and higher education in Hungary, are organised by centralised matching schemes. The program for secondary schools is based on the original model and algorithm of Gale and Shapley, which makes this application especially interesting. The program for higher education is more complex; the model has several special features, but the core of the algorithm is the same. Besides presenting these applications, we study one important aspect of the higher education scheme, the solution concept of score-limit. Here, we formulate appropriate stability criteria that take score-limit into account and then we prove that the existing algorithms that are used/have been used produce matchings that are stable according to these criteria. Therefore, we show that the results of Gale and Shapley apply for the generalised model as well, namely, the applicant/college-oriented algorithms produce stable score-limits, moreover these solutions are the best/worst possible stable score-limits for the applicants.

Introduction

The college admissions problem was introduced and studied by Gale and Shapley [6]. Later, Roth [8] described the history of the National Resident Matching Program (the centralised matching scheme that coordinates junior doctor recruitment in the US), where the same type of algorithm has been used since 1952. For further literature about two-sided matching markets see the book by Roth and Sotomayor [11] and a recent survey of Roth [9] .

Recently, the college admission problem again came into prominence. New centralised matching schemes have been implemented for high schools in Boston [2] and New York [1], and further applications were described by Abdulkadiroğlu and Sönmez [3]. There are also some studies of existing matching schemes in several other countries, including Spain [7], Turkey [4] and Germany [5] but still, there is a lack of information about the details of these schemes and about schemes in other countries.

In Hungary, the admission procedure of higher education institutions has been organised by a centralised matching scheme since 1985. The new national admission system for secondary schools was established in 2000. In both cases, the core of the program is the Gale-Shapley algorithm [6]. In fact, the secondary school admission program follows authentically the model and algorithm of Gale and Shapley.

Besides giving brief descriptions of these systems we study the major special feature of the higher education admission scheme, namely the way that ties are handled in the algorithm. We introduce natural stability criteria for score-limits that is equivalent to classical stability if no ties occur. The college-oriented score-limit algorithm, which was used in practice until 2007, and the applicant-oriented version, used since 2007, are generalisations of the Gale-Shapley algorithms. Therefore our results about the stability and the optimality of the resulting solutions also generalise the theorems of Gale and Shapley.

*Supported by EPSRC grant EP/E011993/1 and by OTKA grant K69027.

1 Secondary school admission system

The new system for secondary school admission was established in 2000. Here, the students, typically at the age of 14, can apply to public secondary schools, where they will study for four years. In fact, their application is to "special studies" offered by the schools. Each study has a quota, determined by the school.

As a first step of the process, the students submit their strict ranking lists over the studies they apply for, with no restriction on the length of the lists. In fact, the students are encouraged to submit long lists that contain also less preferred but still acceptable options. These lists are private, no school can be aware of the rankings of their applicants.

As a second step, the schools organise entrance exams that will provide their strict rankings over their acceptable applicants. Some schools may simply use the grades of the students from primary school and the results of centralised entrance exams, but other schools may organise special entrance exams for each special study they offer. After the exams, every school submits its ranking lists to the center, moreover, these ranking lists are published on the internet in an anonymous way, so every applicant is informed of her current position at each study she applied to.

One extra option is that the applicants can modify their ranking lists after receiving the results of their entrance exams. This is reasonable, because the true preference of a student can change during those months, especially as a result of her personal experiences in the exams.

After the center receives the revised preference lists from the applicants, a program, that uses **the applicant-oriented Gale Shapley algorithm**, provides the final matching. This matching is therefore optimal for the applicants. After seeing the number of students admitted to each study, the schools decide how they will form their new classes for the following September. A school with some unfilled studies can either announce their remaining places for an extra round or cancel those studies. Similarly, if an applicant is not admitted to any school by the program, then she can try to find an available place in the extra round, which is not assisted by the centre.

To make the process clearer from the point of view of the schools, let us describe an example. A school might offer a general study with quota 60 and two special studies, e.g. French and Maths of quotas 15 each.[1] For the general study the ranking list over the students may be determined by their grades from primary school and by results in the centralised entrance exams. The school may organise an oral exam on literature and grammar for the French study, and a high level written or oral maths exam for the Maths study. If all places are filled by the centralised admission scheme then this school would start with three new classes in September, each of size 30: two of them will consist of the students admitted to general studies and the third one will consist of the students admitted to special French and Maths studies.

In 2007, the number of applicants was 116,672, and the number of places in secondary schools was 149,033 in 5709 specialties at 1196 schools. The number of admitted students was 107,429. According to the representatives of the organising institution, the system is very successful and well accepted by the public. There are some complaints every year from parents who are not satisfied with the assignment of their children, but the answer, i.e. each preferred study on the student's list was filled by better applicants, is considered generally as a fair reason. The one controversial aspect of the system is that it is highly competitive at this early age. This is a question to be further studied by sociologists and educational researchers.

[1] A special French study can mean only some extra hours every week in that foreign language, but also can offer an extra year to learn that language and even to study some subjects in that language later. The detailed descriptions of the studies are available from the webpage of the center.

2 Higher education admission system

First we give a short description of the scheme. Then we introduce the notion of stable score-limit, we present the college-oriented and the applicant-oriented algorithms and show that they produce stable score-limits. Finally, we prove that the first solution is the maximal and the second solution is the minimal stable score-limit and therefore these are the worst and the best possible stable solutions for the applicants, respectively.

2.1 Brief description of the scheme

The Ministry of Education founded the Admission to Higher Education National Institute (OFI) in 1985 in order to create, operate and develop the admission system for higher education. First, we note that instead of colleges, in Hungary the higher education institutions have faculties, where education is organised in different fields of studies quite independently. Moreover, most of the studies can be either completely financed by the state or partly financed by the students. Therefore, the applicants have to indicate also in their rankings which kind of study they are applying for at each faculty.[2] So here, students apply for fields of studies of particular faculties. The institutions can admit a limited number of students to each of their fields, these *quotas* being determined by either the institutions or by the Ministry of Education. For simplicity, these fields are referred to as colleges henceforth in order to keep the original terminology of Gale and Shapley.

At the beginning of the procedure, students submit their ranking lists that correspond to their preferences over the colleges they are applying for. There is no limit on the length of the list, however applicants are charged for each item. The students receive scores at each college they applied for according to their final grades at secondary school, and entrance exams. Note, that the scores of a student can differ at two colleges. These scores are integer numbers, currently limited to 480 (this limit was only 144 until 2007, which resulted in massive ties).

The administration is conducted by a government organisation. After collecting the applicants' rankings and their scores, a centralised program computes the score-limits of the colleges as is described in the following. An applicant is admitted by the first college on her list where she is above the score-limit.

The implemented program was a generalised version of the college-oriented Gale-Shapley algorithm until 2007, and since 2007 the core of the program has been **the applicant-oriented Gale-Shapley algorithm**[3]. There are at least three special features in this program that required an extension of the original algorithm with some extra heuristics.

The first special feature of this system, which is the subject of this paper, is that ties must be handled, since applicants may have equal scores. It turns out that the system produces a matching that satisfies a stability condition based on score-limits that we formulise in the next subsection. This generalises the original notion of stability by

[2]An applicant may rank first a state-financed Political Science study at the faculty of Social Sciences in Corvinus University, Budapest. Her second choice may be another state-financed study in Business Administration at the faculty of Business and Economy at the University of Pécs. She may put the very same study in third place on her list as that in first place, only with the privately-financed option rather than the state-financed, and so on. Therefore, a student may prefer a particular study even with fees to another state-financed study, or to the option of not being admitted. So the fees are actually included in the preferences of the applicants in this way.

[3]We note that the change of the program in 2007 was because of some legislative changes. According to a representative of the organising institution, their tests showed that the applicant-proposing version really provided better results for the applicants, although the difference was less than 1%. This is similar to the effect of an analogous change in the National Resident Matching Program described by Roth and Peranson [10].

Gale and Shapley, since they are equivalent if no ties occur. We show that the score-limit algorithms that were implemented in the Hungarian application find solutions that are stable according to the generalised condition. Moreover, we prove that the applicant-oriented version provides the minimal stable score-limit, and therefore it is the best possible solution for the applicants, whilst the college-oriented version provides the maximal stable score-limit, therefore, it is the worst possible solution for the applicants.

The two other specialties are the presence of lower quotas and common quotas over some sets of colleges. These extra restrictions have also crucial effects on the program, which resolves these problems by special heuristics. The analysis of these questions will be the subject of a further study.

In 2007 the number of applicants was 108,854 in total, of whom 88,795 applied for state-financed studies. The number of admitted applicants was 81,563 in total, of whom 48,726 were admitted to state-financed studies. But in 2007 there was a significant drop in the number of applicants. In 2004, for example, the number of applicants was 167,082 in total, of whom 115,798 applied for state-financed studies, whilst the number of admitted applicants was 109,562 in total, of whom 59,641 were admitted to state-financed studies. [4]

This application is also well accepted by the public. The main advantage of the system is the tranparency and the fairness for both sides. Moreover, even bearing in mind that the total number of applicants varied significantly in the last few years as a result of the legislative changes, most of the institutions (and especially the popular institutions) could always admit the same number of students to each field in each year, which makes their education services economical. Again, a criticism of this system might be its competitiveness, since both the students and the institutions are under pressure to achieve good scores and to attract good students, respectively (the popularity of some universities is clearly indicated by their high score-limits). But the benefits of this performance based competition are very likely to outweigh all its drawbacks.

2.2 The definition of stable score-limit

Let $A = \{a_1, a_2, \dots, a_n\}$ be the set of applicants and $C = \{c_1, c_2, \dots, c_m\}$ be the set of colleges, where q_u denotes the quota of college c_u. Let the ranking of the applicant a_i be given by a preference list P^i, where $c_v >_i c_u$ denotes that c_v precedes c_u in the list, i.e. the applicant a_i prefers c_v to c_u. Let s^i_u be a_i's final score at college c_u.

The score-limit l is a nonnegative integer mapping $l : C \to \mathbb{N}$. An applicant a_i is admitted by a college c_u if she achieves the limit at college c_u, and that is the first such place in her list, i.e. $s^i_u \geq l(c_u)$, and $s^i_v < l(c_v)$ for every college c_v such that $c_v >_i c_u$. If the score-limit l implies that a college c_u admits applicant a_i, then we set the Boolean variable $x^i_u(l) = 1$, and 0 otherwise. Let $x_u(l) = \sum_i x^i_u(l)$ be the number of applicants allocated to c_u. A score-limit l is *feasible* if $x_u(l) \leq q_u$ for every college $c_u \in C$.

Let $l^{u,t}$ be defined as follows: $l^{u,t}(c_u) = l(c_u) - t$ and $l^{u,t}(c_v) = l(c_v)$ for every $v \neq u$. That is, we decrease the score limit by t at college c_u, but leave the other limits unchanged. We say that a score-limit l is *stable* if l is feasible but for each college c_u, $l^{u,1}$ is not feasible. This stability condition means that no college can decrease its limit without violating its quota (assuming that the others do not change their limits). We note that if no ties occur (i.e. every two applicants have different scores at each college), then this stability condition is equivalent to the original one defined by Gale and Shapley.

[4]The government introduced fees for every student in higher education in 2007, which was the main reason for the drop. However, this new law was annulled after a national referendum in March 2008, so the number of applicants is very likely to rise again in forthcoming years.

2.3 The score-limit algorithms and the optimality of their outputs

Both score-limit algorithms are very similar to the two versions of the original Gale–Shapley algorithm. The only difference is that now, the colleges cannot necessarily select exactly as many best applicants as their quotas allow, since the applicants may have equal scores. Here, instead each college sets its score-limit always to be the smallest value such that its quota is not exceeded. If the scores of the applicants are all different at each college then these algorithms are equivalent to the original ones.

College-oriented algorithm:
In the first stage of the algorithm, let us set the score-limit at each college independently to be the smallest value such that, when all applicants are considered, the number of applicants offered places does not exceed its quota. Let us denote this limit by l_1. Obviously, there can be some applicants who are offered places by several colleges. These applicants keep their best offer, and reject all the less preferred ones, moreover they also cancel their less preferred applications.

In the subsequent stages, the colleges check whether their score-limits can be further decreased, since some of their offers may have been rejected in the previous stage, hence they look for new students to fill the empty places. So each college sets its score-limit independently to be the least possible, considering their actual applications. If an applicant is admitted by some new, better college, then she accepts the best offer (at least temporarily), and rejects or cancels her other, less preferred applications.

Formally, let l_k be the score-limit after the k-th stage. In the next stage, at every college c_u, the largest integer t_u is chosen, such that $x_u(l_k^{u,t_u}) \leq q_u$. That is, by decreasing its score-limit by t_u, the number of applicants offered a place by c_u does not exceed its quota, supposing that all other score-limits remained the same. For every college let $l_{k+1}(c_u) := l_k^{u,t_u}(c_u)$ be the new score-limit. If some limits are decreased simultaneously, then some applicants can be offered a place by more than one college, so $x_u(l_{k+1}) \leq x_u(l_k^{u,t_u})$. Obviously, the new score-limit remains feasible.

Finally, if no college can decrease its limit, then the algorithm stops. The stability of the final score-limit is obvious by definition.

Applicant-oriented algorithm:
Let each applicant propose to her first choice in her list. If a college receives more applications than its quota, then let its score-limit be the smallest value such that the number of provisionally accepted applicants does not exceed its quota. We set the other limits to be 0.

Let the score-limit after the k-th stage be l_k. If an applicant has been rejected in the k-th stage, then let her apply to the next college in her list, say c_u, where she achieves the actual score-limit $l_k(c_u)$ (if there remains such a college in her list). Some colleges may receive new proposals, so if the number of provisionally accepted applicants exceeds the quota at a college, it sets a new, higher score-limit l_{k+1}. Again, this is the smallest score-limit such that the number of applicants offered a place by the college does not exceed its quota, supposing that all other score-limits remained the same. At the same time, it rejects all those applicants that do not achieve this new limit.

The algorithm stops if there is no new application. The final score-limit is obviously feasible. It is also stable, because after a limit is increased for the last time, then the rejected applicants get less preferred offers during the algorithm. So if the limit in the final solution were decreased by one for this college, then these applicants would accept the offer, and the quota would have been exceeded. The following result is therefore immediate.

Theorem 1 *Both the score-limit l_C, obtained by the college-oriented algorithm, and the score-limit l_A, obtained by the applicant-oriented algorithm, are stable.*

It is easy to give an example to show not only that some applicants can be admitted by preferred places in l_A as compared to l_C, but the number of admitted applicants can also be larger in l_A. We say that a score-limit l is *better* than l_* for the applicants if $l \leq l_*$, (i.e. $l(c_u) \leq l_*(c_u)$ for every college c_u). In this case every applicant is admitted by the same or by a preferred college at score-limit l than at l_*.

Theorem 2 *l_C is the worst possible and l_A is the best possible stable score-limit for the applicants, i.e. for any stable score-limit l, $l_A \leq l \leq l_C$ holds.*

Proof: Both proofs are based on indirect arguments, that are similar to the original one of Gale and Shapley.

Suppose first, that there exists a stable score-limit l_* and a college c_u such that $l_*(c_u) > l_C(c_u)$. During the college-oriented algorithm there must be two consecutive stages with score-limits l_k and l_{k+1}, such that $l_* \leq l_k$ and $l_*(c_u) > l_{k+1}(c_u)$ for some college c_u. Obviously, $l_k^{u,t_u}(c_u) = l_{k+1}(c_u)$ by definition, and $x_u(l_k^{u,t_u}) \leq q_u < x_u(l_*^{u,1})$ by the stability of l_*. So, on the one hand, there must be an applicant, say a_i who is admitted by c_u at $l_*^{u,1}$ but not admitted by c_u at l_k^{u,t_u}. On the other hand, the indirect assumption $l_k^{u,t_u}(c_u) = l_{k+1}(c_u) \leq l_*(c_u) - 1 = l_*^{u,1}(c_u)$ implies that a_i must be admitted at l_k^{u,t_u} by a college preferred to c_u (since a_i has a score of at least $l_k^{u,t_u}(c_u)$ there), and obviously also at l_k. That is impossible if $l_* \leq l_k$, a contradiction.

To prove the other direction, we suppose that there exists a stable score-limit l_* and a college c_u such that $l_*(c_u) < l_A(c_u)$. During the applicant-oriented algorithm there must be two consecutive stages with score-limits l_k and l_{k+1}, such that $l_* \geq l_k$ and $l_*(c_u) < l_{k+1}(c_u)$ for some college c_u. At this moment, the reason for the incrementation is that more than q_u students are applying for c_u with a score of at least l_*. This implies that one of these students, say a_i, is not admitted by c_u at l_* (however she has a score of at least $l_*(c_u)$ there). So, by the stability of l_*, she must be admitted by a preferred college, say c_v at l_*. Consequently, a_i must have been rejected by c_v in a previous stage of the algorithm, and that is possible only if $l_*(c_v) < l_k(c_v)$, a contradiction. ∎

Acknowledgment

I would like to thank to Gábor Varjasy, the representative of Higher Education National Institute, and György Salomvári, the representative of Educational Information Office for the detailed information about these programs. Furthermore, I acknowledge Tamás Fleiner and Rob Irving for their useful comments.

References

[1] A. Abdulkadiroğlu, P. A. Pathak, and A. E. Roth. The New York City high school match. *American Economic Review, Papers and Proceedings*, 95(2):364–367, 2005.

[2] A. Abdulkadiroğlu, P. A. Pathak, A. E. Roth, and T. Sönmez. The Boston public school match. *American Economic Review, Papers and Proceedings*, 95(2):368–371, 2005.

[3] A. Abdulkadiroğlu and T. Sönmez. School Choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003.

[4] M. Balinski and T. Sönmez. A tale of two mechanisms: Student placement. *J. Econom. Theory*, 84(1):73–94, 1999.

[5] S. Braun, N. Dwenger, and D. Kübler. Telling the truth may not pay off: An empirical study of centralised university admission in Germany. 2007. SFB 649 Discussion Paper 2007-070.

[6] D. Gale and L. S. Shapley. College Admissions and the Stability of Marriage. *Amer. Math. Monthly*, 69(1):9–15, 1962.

[7] A. Romero-Medina. Implementation of stable solutions in a restricted matching market. *Review of Economic Design*, 3(2):137–147, 1998.

[8] A. E. Roth. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of Political Economy*, 6(4):991–1016, 1984.

[9] A. E. Roth. Deferred Acceptance Algorithms: History, Theory, Practice, and Open questions. *Internat. J. Game Theory*, 36, 2008.

[10] A. E. Roth and E. Peranson. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review*, 89(4):748–780, 1999.

[11] A. E. Roth and M. Sotomayor. *Two-sided matching*, volume 18 of *Econometric Society Monographs*. Cambridge University Press, Cambridge, 1990. A study in game-theoretic modeling and analysis, With a foreword by Robert Aumann.