

Hardest Sudokus

May 27, 2011

What is the hardest sudoku ?

The basic problem with this question is, that there has never been a rating definition for hard puzzles, which has been commonly accepted. Interestingly there was a time (in 2006), where it could be answered nevertheless. When Arto Inkala published his AI Escargot, it was the only puzzle, which could not solved by the solution and rating programs of that time, which did not use guessing/backtracking. The common way for the programmers to fix that was to introduce a second level, e.g. nested contradiction chains. It did not last longer than a few weeks however, until Ocean and others from the Player's Forum found other puzzles with that property. And a year later hundreds of other puzzles were rated higher than AI Escargot in the hardest lists (beside of his own).

But also the hardest lists from different rating programs did not show much correlation. There have been only few private and public programs with some acceptance at all for hard puzzles – the majority of solvers simply use guessing and backtracking to get to a solution, a method, which cannot provide a serious rating.

The most popular and most accepted rating program is [Nicolat Juillerat's Sudoku Explainer](#). The solution techniques mainly used for hard puzzles are forcing and contradiction chains. The rating is based on the minimum length of such chains, needed to crack the hardest step in the solution path.

Its main drawback is, that it has not integrated some solution techniques (e.g. ALS and all its extensions), which have been developed later. But the rating still has a rather good correlation to others like the ones by [Denis Berthier](#), who also mentioned recently, that an improvement of a technique only lowers the rating of a small percentage of puzzles.

This is somewhat different with [champagne's](#) rating. His techniques seem to be basically different. He is able to solve a good part of ER 11+ puzzles without recursion, but fails on other puzzles with Explainer rating down to 10.6. One reason is, that he also uses the possible combinations of the candidates in 2 or 3 cells as starting points for derivations (SE only does that for one digit in region forcing chains). This method allowed some astonishing easy solutions of high rated puzzles, but normally its hardly easier to follow the solutions manually than it is for nested chains. The program is private and the ratings are given only verbally up to now. Unfortunately a rating program, which combines the techniques of these two programs, is not in sight.

There are two properties, a **good rating program** must have in my eyes:

- It must be stable for equivalent versions of the same puzzle
- It must be based on solution techniques, which can be applied manually and produce a solution, which can be followed manually.

Its a matter of taste for me, if only the hardest step or the overall solution determines the rating. Also backdoor solutions might be

accepted (there are extremely hard puzzles, which can be solved with singles, when you just try the right candidates), but i prefer to forbid them.

The first property is not fulfilled by all the programs, which use guessing/backtracking or which e.g. scan a puzzle left/right and top/down, until they find an elimination. To fulfill it, for each step you have to find an "optimal" solution. This leads to very slow algorithms.

The second property is the reason, why i see the popular programs by [gsf](#) (q1/q2 rating) and [dukuso](#) (suexrat) more as filters for extremely hard puzzles than as rating programs. dukuso's program does not use a technique other than singles, so its a kind of machine rating. gsf's program does not provide any readable solutions. He only gave a description, how they work. I noticed many cases, where puzzles were strongly underrated, and for a manual solver it is not possible to follow, why this puzzle should be harder than that. However, there is obviously a correlation to the Explainer ratings. The percentage of 11+ puzzles in sets with higher q2 rating is noticable higher. But there is also another correlation: In puzzles with more clues the percentage of 11+ puzzles in sets with high q2 is much lower than for puzzles with less clues. E.g. for 21 clue puzzles with q2 rating 96xxx i had about the same 10% rate as for 23 clues with q2 98xxx.

My search for hard sudokus

My interest in generating hard sudokus arised in autumn 2010. I had made a search for minimal puzzles with the maximum number of givens a year before and i wanted to know, if the same method also could give good results for such a vaguely defined problem. The method is

described in the player's forum thread "[The making of a gotchi](#)".

At that time the people with the top entries in the most popular top hardest lists (q1, ER, q2 and the two suex ratings) seemed to have finished their work at 2 years before. Concerning the Explainer rating, champagne has compiled a list of 488 puzzles with ER 11+ from various sources (public and private). I had the feeling, that all the generators were using a similar method, which was pattern oriented. They tried to find promising cell patterns, which partially were filled with givens, and then find hard puzzles in the neighbourhood using the $\{-n, +m\}$ method (remove n clues and add m others).

My own method was pattern independent, so i had some hope to find hard puzzles with patterns never investigated – and with a difficulty never discovered.

I started with a big set of random puzzles and tried to find harder puzzles in the neighbourhood.

With some breaks the search lasted about 6 months running as a background process (on a 3Gh PC). In this time i generated more than 15 mio "step1" puzzles, i.e. puzzles, which cannot be solved with cell forcing chains. Using strong filters i selected more than 42000 to rate them with the "[FIXED6](#)" version of Sudoku Explainer in batch mode. The rating alone took more than 2 months in sum, the average rating of these puzzles was 10.65.

There is a notable high percentage of symmetrical puzzles under the highest rated ones. While in the old list this obviously was due to the search methods (most high rated puzzles have diagonal or rotational symmetric puzzles up to ER 11.8), i did not prefer symmetrical patterns in any way and still found more than

10% symmetric puzzles under the hardest (also new ones up to ER 11.8).

What i found:

- o New hardest puzzles for the q1, q2, both suexrat and champagne's rating
- o Three more ER 11.9 puzzles ("[Imam Bayildi](#)", "[Kolk](#)", "[Patience](#)" and also the known one, "Golden Nugget" by tarek)
- o 201 new puzzles with ER 11.5+ (50 known before)
- o More than 7200 puzzles with ER 11+ (488 – 556minimals – known before)
- o Almost 90% of the known ER 11+ puzzles (independantly – only 92 of them passed my rating filters)
- o More than 26000 puzzles with ER 10.6+
- o A grid with only a single strong link
- o A puzzle, which cracked gsf's q2 rating ("no solution")
- o A puzzle with double diagonal symmetry and ER 11.6 ("[Pretty Biest](#)")

Note: Al Escargot has ER 10.5, Arto's "hardest" of 2010 10.6, but its solvable with cell forcing chains. Both did not pass one of my rating filters.

What i did not find:

- o A puzzle, which is not solvable with (level 2) nested singles (contradiction) chains, also denoted as T&E(singles, 2), so Denis Berthier's conjecture still holds.
- o A puzzle with ER higher than 11.9

Some stats

A comparison of ER distributions in the "others" list and mine:

ER		11.9	11.8	11.7	11.6	11.5			
	11.4	11.3	11.2	11.1	11.0				
Others			1		9	10	14	16	

34	103	135	131	38		
Eleven			3	20	29	44 105
241	1117	1541	2436	1704		
Sum			4	29	39	58 121
275	1220	1676	2567	1742		
ER 11.5+						
# clues			20	21	22	23 24
37						
Others			2	40	3	0 0 0
Eleven			0	41	118	42 0 0
Sum			2	81	121	42 0 0
ER 11.0+						
# clues			20	21	22	23 24
37						
Others			2	416	69	0 0 1
Eleven			9	1240	3645	1907 238 0
Sum			11	1656	3714	1907 238 1

Downloads

My list with **26012 ER 10.6+** puzzles can be downloaded [here](#). It contains all the puzzles in the "[elevens_ER11](#)" list posted on Mar 26, 2011. Probably none of the puzzles have been published by others before, at least i have also checked JPF's and gsf's list,

A list with the puzzles collected by champagne (with Mauricio's 37 clue exchanged by 60 minimals, and including 8 puzzles from the patterns game) can be downloaded [here](#) (state Mar 27, 2011). I saw that a few puzzles with ER up to 11.3 have been found in the meantime.

I had also uploaded my list of all [21 clue "step1" puzzles](#) (8.8MB, state Mar 27, 2011).

Contact

If you have any comments or questions, please post them in the new hardest thread of the New

Edit August 7, 2011

When champagne posted a [new top 11.9](#) puzzle 2 weeks ago, dobrichev noted, that it has the property, that you cannot derive any number with $n-1$ of the givens. Using his program i scanned my sets for other puzzles with this property and found 690 new 11+ puzzles this way (out of more than 60000 "dob-0"- and even more non minimal dob-0 puzzles). They can be found [here](#).

I also removed 103 non minimal puzzles from my hardest list, as well as 229, champagne pointed out, that they have been published before somewhere. My new list now is [elevens_hardest_V2](#).