A Pencil-and-Paper Algorithm for Solving Sudoku Puzzles

J. F. Crook

The puzzle Sudoku has become the passion of many people the world over in the past few years. The interesting fact about Sudoku is that it is a trivial puzzle to solve. The reason it is trivial to solve is that an algorithm exists for Sudoku solutions. The algorithm is a tree-based search algorithm based on backtracking in a tree until a solution is found.

If all a person needs to do is sit down at their personal computer, punch in the numbers given in the puzzle, and then watch a computer program compute the solution, we can reasonably ask why a person would bother to struggle to solve Sudoku puzzles. The reason is that people enjoy struggling with pencil and paper to work out Sudoku solutions. Herzberg and Murty (2007, p. 716) give two reasons for the enjoyment of this struggle:

First, it is sufficiently difficult to pose a serious mental challenge for anyone attempting to do the puzzle. Secondly, simply by scanning rows and columns, it is easy to enter the "missing colors", and this gives the solver some encouragement to persist.

This paper develops an algorithm for solving any Sudoku puzzle by pencil and paper, especially the ones classified as *diabolical*.

Definition of the Sudoku Board

Sudoku is played on a 9×9 board. There are eighty-one cells on the board, which is broken down into nine 3×3 subboards that do not overlap. We call these subboards *boxes* and number them from 1 to 9 in typewriter order beginning in the upper left-hand corner of the board, as displayed in Figure 1.

The notation for referring to a particular cell on the board is to give the row number followed by the column number. For example, the notation $c\,(6,7)$—where c denotes *cell*—denotes the cell at the intersection of row 6 and column 7.

The theory we develop in the next section uses the widely known concept of *matching numbers across cells*. Various authors, as suits their whim, name matching numbers differently. For example, Sheldon (2006, p. xiv) names them *partnerships*, whereas Mepham (2005, p. 9) names the concept *number sharing*. Here, we will use the name *preemptive sets*, which is more precise from a mathematical point of view. The theory developed here applies to Sudoku boards of all sizes.[1]

J. F. Crook is professor emeritus of computer science, Winthrop University, Rock Hill, SC. His email address is `crookj@winthrop.edu`.

I thank John Hohn, Kathryn Cooper, David Gill, and Libby Neely for reading my paper. I also thank Loretta Nethercot for giving me my first two Sudoku books—on Christmas Day, 2005, and on my seventieth birthday in 2007.

[1] *Sudoku boards can be classified into regular and nonregular boards. The formula for regular Sudoku boards is: Let m denote the width and height of a Sudoku subboard, where $m \geq 2$. Then the width and height of a regular Sudoku board is m^2. The sizes of regular subboards and boards are given for a few values of m in the following table:*

Subboard Width and Height m	Board Width and Height m^2	Number of Cells on the Board $m^2 \times m^2$
2	4	16
3	9	81
4	16	256
5	25	625
6	36	1296

The most common nonregular Sudoku board is the 6×6, which consists of six nonoverlapping 2×3 subboards. The newspaper USA Today *publishes 6×6 puzzles regularly.*

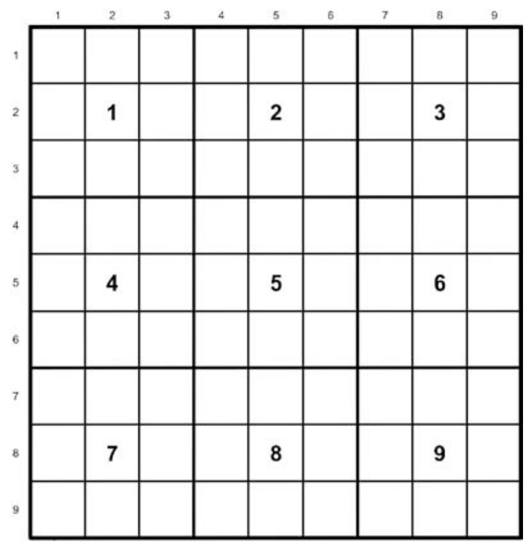

Figure 1. The Sudoku board.

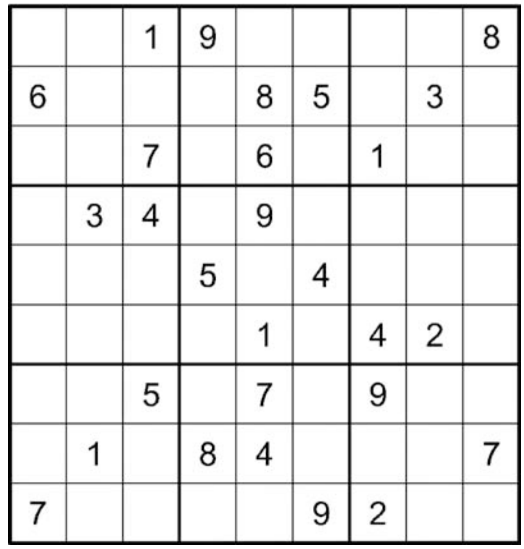

Figure 2. Example of a Sudoku puzzle.

Preemptive Sets and the Occupancy Theorem

The single most important tool for solving Sudoku puzzles is based on the definition of the *solution* of a Sudoku puzzle.

Definition 1 (Sudoku Solution)**.** *The solution of a Sudoku puzzle requires that every row, column, and box contain all the numbers in the set* $[1, 2, \ldots, 9]$ *and that every cell be occupied by one and only one number.*

This definition implies that no row, column, or box will have any number repeated. An example of a Sudoku puzzle is shown in Figure 2. The more difficult puzzles can only be solved efficiently by writing down in each empty cell the *possible* numbers that can occupy the cell. This list of possible numbers for each cell is called the *markup of the cell*. The markup of the example puzzle in Figure 2 is shown in Figure 3.

We now define preemptive sets, which is the primary tool for solving Sudoku puzzles up to the point where either (1) a solution is found or (2) continuation requires randomly choosing one of two or more numbers from the markup of an empty cell.

Definition 2 (Preemptive Sets)**.** *A preemptive set is composed of numbers from the set* $[1, 2, \ldots, 9]$ *and is a set of size* m, $2 \le m \le 9$, *whose numbers are potential occupants of* m *cells exclusively, where exclusively means that no other numbers in the set* $[1, 2, \ldots, 9]$ *other than the members of the preemptive set are potential occupants of those* m *cells.*

A preemptive set is denoted by

$$\{[n_1, n_2, \ldots, n_m], [c(i_1, j_1), c(i_2, j_2), \ldots, c(i_m, j_m)]\},$$

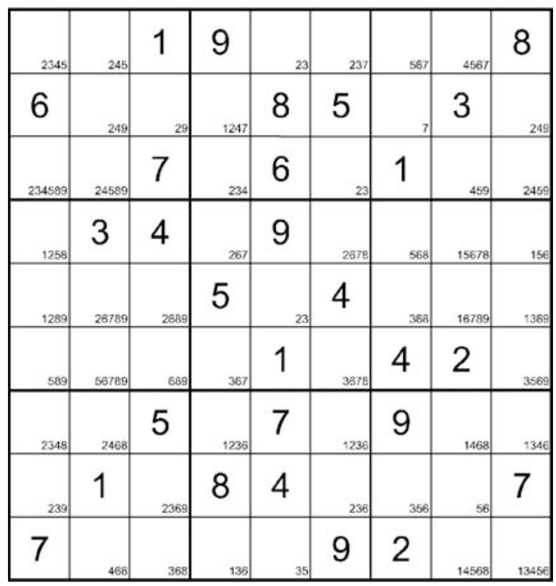

Figure 3. Markup of the example puzzle in Figure 2.

where $[n_1, n_2, \ldots, n_m]$, $1 \le n_i \le 9$ *for* $i = 1, 2, \ldots, m$, *denotes the set of numbers in the preemptive set and* $[c(i_1, j_1), c(i_2, j_2), \ldots, c(i_m, j_m)]$ *denotes the set of* m *cells in which the set* $[n_1, n_2, \ldots, n_m]$, *and subsets thereof, exclusively occur.*

Definition 3 (Range of a Preemptive Set)**.** *The range of a preemptive set is a* row, column, *or* box *in which all of the cells of the preemptive set are located. When* $m = 2$ *or* 3, *the range can be one of the sets* $[row, box]$ *or* $[column, box]$.

A description of a preemptive set is that it is a set of m distinct numbers from the set $[1, 2, \ldots, 9]$ and a set of m cells exclusively occupied by the

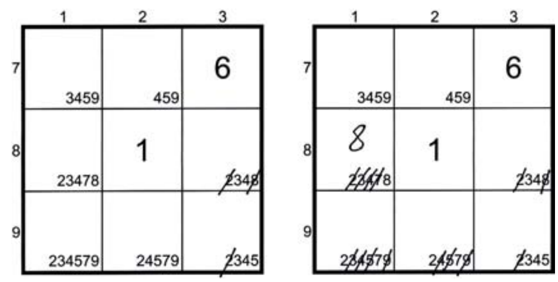

Figure 4. A box with a preemptive set of size 4.

m numbers, or subsets of them, with the property that the distribution of the m numbers across the m cells is not known at the time the preemptive set is discovered. The distribution of the m numbers into the m cells will be revealed as the solution of the puzzle progresses.

An example of a preemptive set is displayed in Figure 4, where the large-sized numbers are the numbers given in the puzzle and the list of small-sized numbers in the lower right-hand corner of each cell is the markup. In Figure 4 we observe the preemptive set

$$\{[3,4,5,9],[c(7,1),c(7,2),c(8,3),c(9,3)]\},$$

which is of size 4, a preemptive quadruple, in the box on the left.

Since 8 is a singleton in the markup of this box, we can cross out the 2, 3, 4, and 7 in $c(8,1)$ and enter 8 in $c(8,1)$. Now, the crossing out of numbers in the preemptive quadruple in cells other than those in the quadruple in the left-hand box in Figure 4 results in the preemptive pair $\{[2,7],[c(9,1),c(9,2)]\}$ in the right-hand box.

Theorem 1 (Occupancy Theorem). *Let X be a preemptive set in a Sudoku puzzle markup. Then every number in X that appears in the markup of cells not in X over the range of X cannot be a part of the puzzle solution.*

Proof. If any number in X is chosen as the entry for a cell not in X, then the number of numbers to be distributed over the m cells in X will be reduced to $m-1$, which means that one of the m cells in X will be unoccupied, which violates the Sudoku solution definition. Hence, to continue a partial solution, all numbers in X must be eliminated wherever they occur in cells not in X over the range of X.

The literature on Sudoku frequently uses the notion of a hidden pair or hidden triple and so forth; see, for example, Mepham (2005, p. 9). Hidden pairs, triples, and so forth are simply preemptive sets waiting to be discovered, as we will now explain. An example of a hidden pair, the pair $[3,5]$, is presented in Figure 5 in the left-hand box. This hidden pair—called a hidden pair because of the presence of the 6 in one case and the 2 and 8 in the other—will turn out to be the preemptive set $\{[3,5],[c(2,7),c(3,8)]\}$, as follows: On a first examination of the left-hand box in Figure 5, we should immediately notice the presence of the preemptive set

$$\{[1,2,6],[c(1,7),c(1,9),c(2,9)]\},$$

which, after crossing out the 1's, 2's, and 6's in cells not in the preemptive set, produces the box in the middle of Figure 5, where we observe the singleton 8 in $c(1,8)$, which means that 8 is the only entry possible in $c(1,8)$ and the 8 in the markup of $c(3,8)$ must, therefore, be crossed out. The result in the right-hand box is the revelation of the preemptive pair

$$\{[3,5],[c(2,7),c(3,8)]\}.$$

We now state a theorem the reader can easily prove that establishes preemptive sets as the basic tool for solving Sudoku puzzles.

Theorem 2 (Preemptive Sets). *There is always a preemptive set that can be invoked to unhide a hidden set, which then changes the hidden set into a preemptive set except in the case of a singleton.*

Hidden tuples are, however, quite useful, because they are often easier to spot than the accompanying preemptive set, especially hidden singletons and pairs.

The three boxes in Figure 5 are an excellent example of the process of breaking down the preemptive set in a box, row, or column that is present by default as soon as the markup of a puzzle is completed. That is to say, the box on the left in Figure 5, for example, is *covered* by the preemptive set

$$\{[1,2,3,5,6,8],[c(1,7),c(1,8),\\ c(1,9),c(2,7),c(2,9),c(3,8)]\}$$

by default. Now, the right-most box in Figure 5 is covered by the preemptive set

$$\{[1,2,3,5,6],[c(1,7),c(1,9),c(2,7),\\ c(2,9),c(3,8)]\},$$

but this preemptive set is the union of the two smaller preemptive sets

$$\{[1,2,6],[c(1,7),c(1,9),c(2,9)]\}\\ \text{and } \{[3,5],[c(2,7),c(3,8)]\}.$$

The fact that both of these preemptive sets occur in the same box means that their intersection must be empty. That is to say, $[1,2,6]\cap[3,5]=\emptyset$ and $[c(1,7),c(1,9),c(2,9)]\cap[c(2,7),c(3,8)]=\emptyset$. The remarks here also apply to rows and columns, of course.

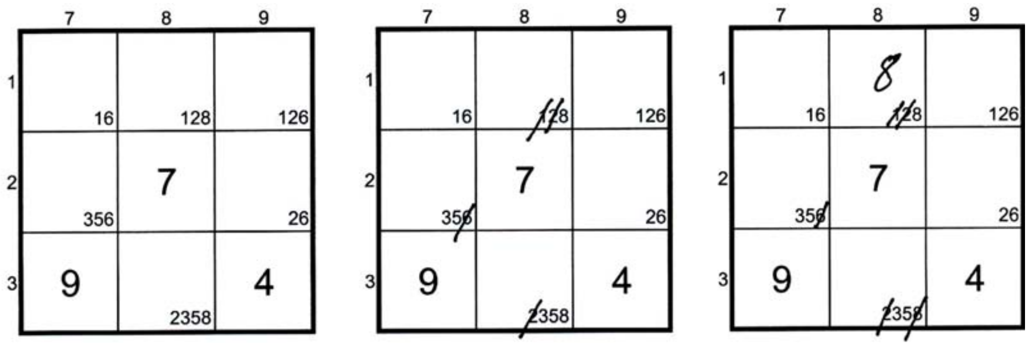

Figure 5. Changing a hidden pair into a preemptive pair.

An Algorithm for Solving Sudoku Puzzles

In this section we develop an algorithm that solves Sudoku puzzles.[2] The first puzzle we use as an example only requires the use of preemptive sets to arrive at a solution.

The second puzzle uses preemptive sets to reach the point where continuation requires random choice. After the random choice of a number for entry into a cell is made, the algorithm returns to the use of preemptive sets until a solution is reached or another random choice has to be made, and so forth.

	3	9	5					
			8				7	
				1		9		4
1			4					3
		7				8	6	
		6	7		8	2		
	1			9				5
					1			8

Figure 6. Will Shortz's puzzle 301.

One should always begin the solution of a Sudoku puzzle by looking for cells within boxes to enter numbers within that box that are missing. The easiest approach is to begin with the highest frequency number(s) given in the puzzle. The method proceeds by finding a box that is missing this high-frequency number and determining whether there is one and only one cell into which this number can be entered. If such a cell exists, the entry of that number into the cell is *forced*. When no more of this number can be forced, continue with the number with the next highest frequency number and so forth until no more numbers can be forced into cells.

Our example puzzle for finding a solution using only preemptive sets is Will Shortz's 301 (Shortz, 2006, puzzle 301), which we shall refer to as Shortz 301. The puzzle is shown in Figure 6. Shortz 301 is classified as *Beware! Very Challenging.*

---

[2] *The solution of a Sudoku puzzle is not necessarily unique, which is apparently widely known among serious Sudoku players (see the section "On the Question of the Uniqueness of Sudoku Puzzle Solutions"). Some experts, such as Sheldon (2006, p. xx), argue that all published Sudoku puzzles should have unique solutions. However, this point of view is not universal. As a matter of fact, Thomas Snyder of the USA, who won the 2007 World Sudoku Championship in Prague, Czech Republic, solved a puzzle that has exactly two solutions to win the championship. The puzzle Snyder solved is*

	5			2			3	
2					1	7		8
4		7	6					
					5			
5	2						4	7
			7					
					3	5		4
3		6	5					1
	9			7			6	

(see `http://wpc.puzzles.com/sudoku/jigsawFinals.htm`*). We found two solutions for this puzzle, and they differ only in one of the* 2×2 *squares that straddles boxes 5 and 6, as follows:*

$$\begin{array}{ccc} & {\scriptstyle 6} & {\scriptstyle 7} \\ {\scriptstyle 5} & 6 & 8 \\ {\scriptstyle 6} & 8 & 6 \end{array} \quad \text{and} \quad \begin{array}{ccc} & {\scriptstyle 6} & {\scriptstyle 7} \\ {\scriptstyle 5} & 8 & 6 \\ {\scriptstyle 6} & 6 & 8 \end{array}.$$

The puzzle is, of course, hard, but application of the algorithm given in this paper resulted in a solution in less than one hour. The London Times *reported that Snyder took about five minutes to find a solution, including checking that his solution was valid.*

Herzberg and Murty (2007, Figures 3–5, pp. 711–12) display a puzzle with the same properties as the one Snyder solved.

	1	2	3	4	5	6	7	8	9
1	24678	3	9	5	2467	2467	16	128	126
2	2456	2456	1	8	2346	9	356	7	26
3	25678	25678	258	236	1	2367	9	2358	4
4	1	25689	258	4	25678	2567	57	259	3
5	2345689	245689	23458	12369	235678	23567	1457	12459	1279
6	23459	2459	7	1239	235	235	8	6	129
7	3459	459	6	7	345	8	2	1349	19
8	23478	1	2348	236	9	2346	3467	34	5
9	234579	24579	2345	236	23456	1	3467	349	8

Figure 7. The markup of Shortz 301.

The Solution of Shortz 301

There are two forced numbers in Shortz 301: 1 in $c\,(2,3)$ and 9 in $c\,(2,6)$. After entering these numbers into the puzzle, we mark up the puzzle. The markup of a Sudoku puzzle can be done either manually or by a computer program. The sensible approach is to use a computer program for this tedious task, which is prone to error when done manually.[3] The markup of Shortz 301, along with the inclusion of the two forced numbers, is shown in Figure 7.

The reader is strongly advised to make a copy of Figure 7 and make the entries and cross-outs described below on the copy.

On a first scan through the puzzle we note the preemptive triple

$$(*) \qquad \{[1,2,6], [c\,(1,7), c\,(1,9), (2,9)]\}$$

in box 3, which the reader will recognize from Figure 5. Hence, cross out all 1's, 2's, and 6's in cells not in $(*)$ in box 3, and as a result enter 8 in $c\,(1,8)$ in Figure 8.

The purpose of subscripting the 8 in $c\,(1,8)$ in Figure 8 is to denote the sequential place of entry in the progress toward a solution of the puzzle. When a number is entered into a cell, all other instances of that number in the markup of the cells of the appropriate row, column, and box must be crossed out, so we cross out the 8 in $c\,(1,1)$ in Figure 7, as shown in Figure 8.

In the markup of column 9 in Figure 7 there is the hidden singleton 7 in $c\,(5,9)$, so cross out the 1, 2, and 9 in $c\,(5,9)$ in Figure 8 and enter 7_2, and then cross out the 7's in $c\,(4,7)$, $c\,(5,5)$, $c\,(5,6)$, and $c\,(5,7)$, as shown in Figure 8.[4]

The entries 5_3 up to 5_5 in Figure 8 are self-explanatory. Next, we note the preemptive triple $\{[2,3,6], [c\,(3,4), c\,(8,4), c\,(9,4)]\}$ in column 4 of Figure 7. Therefore, we cross out all 2's, 3's, and 6's in the other cells of column 4, as shown in Figure 8. This crossing out in column 4 results in the preemptive pair $[1,9]$ in column 4 and box 5. There are no 1's or 9's to be crossed out in the other cells of box 5, so the presence of this pair does not result in pushing the solution forward immediately.

There is a preemptive pair $[2,8]$ in column 3 in cells $c\,(3,3)$ and $c\,(4,3)$ of Figure 8, which gives rise to the preemptive quadruple

$$\{[3,4,5,9], [c\,(7,1), c\,(7,2), c\,(8,3), c\,(9,3)]\}$$

in box 7, which the reader will recognize from Figure 4; the preemptive pair

$$\{[3,4], [c\,(8,3), c\,(8,8)]\}$$

in row 8; the preemptive pair

$$\{[2,7], [c\,(9,1), c\,(9,2)]\}$$

in row 9; and the preemptive pair

$$\{[2,6], [c\,(8,4), c\,(8,6)]\}$$

in row 8 and box 8.

Now, as a result of discovering these five preemptive sets, there is a singleton 7 in row 8 in the markup of $c\,(8,7)$ in Figure 8, so enter 7_6 in $c\,(8,7)$.

There is also a singleton 8 in the markup of $c\,(8,1)$ and a singleton 3 in the markup of $c\,(9,4)$, so enter 8_7 and 3_8 in $c\,(8,1)$ and $c\,(9,4)$ respectively. Lastly, at this point, the preemptive pair $\{[4,5], [c\,(9,3), c\,(9,5)]\}$ implies 6_9 and 9_{10}.

The remaining entries required to arrive at a solution are easy to follow, beginning with 1_{11} in $c\,(7,9)$ in Figure 8 up to the last one: 3_{56} in $c\,(5,1)$. Figure 9 displays the solution of Shortz 301 in "clean form".

A Sudoku Puzzle Whose Solution Requires Random Choice

The technique of random choice will be required for continuation when the following condition is met:

Condition 1 (Random Choice). When no preemptive set in any row, column, or box can be broken into smaller preemptive sets, then an empty cell

---

[3] *J. F. Crook (2007).* Visual Basic program for marking up Sudoku puzzles. *The program runs as a macro under Microsoft's Excel spreadsheet program. Excel was chosen because the Excel data structure is perfect for holding Sudoku puzzles.*

[4] *The discovery of the preemptive quadruple*

$$\{[1,2,6,9], [c\,(1,9), c\,(2,9), c\,(6,9), c\,(7,9)]\}$$

in column 9 of Figure 7 would also have revealed the singleton 7 in column 9.

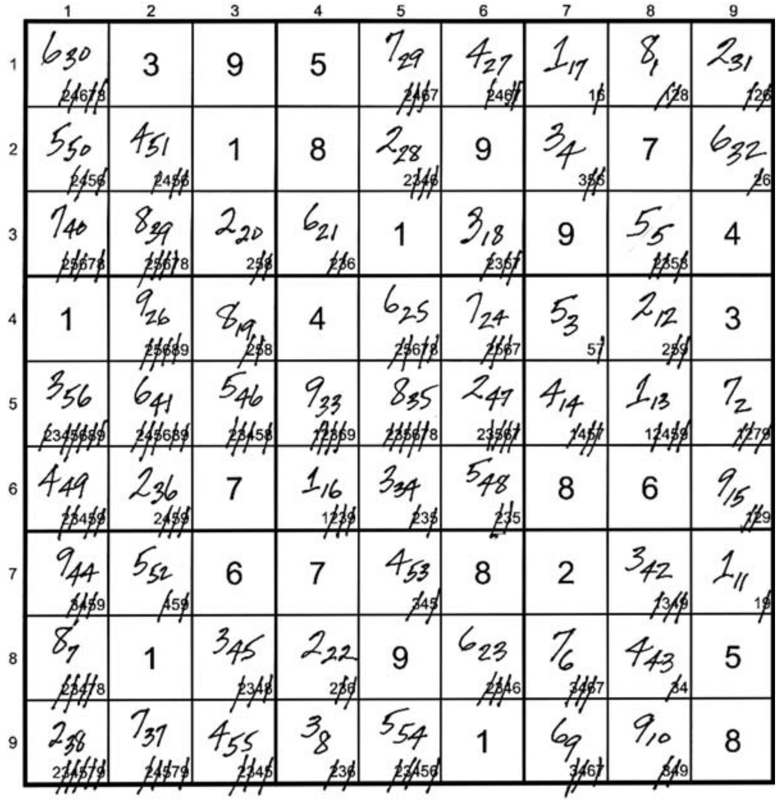

Figure 8. Shortz 301 solution.

6	3	9	5	7	4	1	8	2
5	4	1	8	2	9	3	7	6
7	8	2	6	1	3	9	5	4
1	9	8	4	6	7	5	2	3
3	6	5	9	8	2	4	1	7
4	2	7	1	3	5	8	6	9
9	5	6	7	4	8	2	3	1
8	1	3	2	9	6	7	4	5
2	7	4	3	5	1	6	9	8

Figure 9. Shortz 301 solution, again.

must be chosen and a number randomly chosen from that cell's markup in order to continue solving the puzzle.

The existence of preemptive sets means that the solution is still partial, because there are no preemptive sets in a solution.

The cell chosen to begin the continuation becomes the vertex of a search path that we *generate on the fly*, and the number chosen by random choice from the markup of that vertex is the label of the vertex.[5] The easiest way to keep track of the various paths in a "puzzle" tree is to use pencils of different colors. The path coloring scheme we use here is red for the first path, green for any path that begins at the end of the red path, and then blue and so forth.

Before continuing we define a Sudoku violation:

Definition 4 (Sudoku Violation)**.** *A violation in Sudoku occurs when the same number occurs two or more times in the same row, column, or box.*

The point of using a colored pencil to record a path is that if a *violation* occurs while generating the path, backtracking is easy when one can identify, by their color, which entries on the board have to be erased. An erasure of a path would also include, of course, erasing all cross-outs of numbers with that path's color.

---

[5] *There are, in this context, two kinds of trees. The kind we are talking about here are built from scratch as required and never consist of more than a path through the tree. Hence, we say that they are generated on the fly.*

The other kind of tree exists in "storage" as complete trees that can be accessed for various uses. For example, a binary search tree of n names and associated data is a permanent data object on which various operations such as searching can be performed as required.

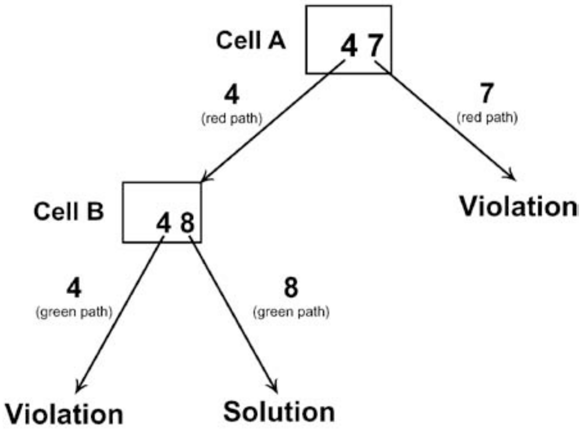

Figure 10. An example search tree with three paths.

	9		7			8	6	
	3	1			5		2	
8		6						
		7		5				6
			3		7			
5				1		7		
						1		9
	2		6			3	5	
	5	4			8		7	

Figure 11. The Mepham diabolical Sudoku puzzle.

When one has completely erased a path that led to a violation, one can begin a new path with the erased color. Now, assume that two green paths lead to violations where only two green paths are possible. Then, you must erase not only the second green path but also the red path that is its parent, because the first green path has already led to a violation.

After erasing the red path, delete the number that led to the violation from the red path's choice set, which is initially the markup of the vertex, and then randomly choose a number from the reduced choice set as the new label of the vertex.

In the absence of an epiphany concerning the likely path to a solution, always choose a preemptive pair if one exists, because one of the two numbers will be the correct choice for its cell and will force the choice in the other cell of the preemptive pair. If no preemptive pair exists, then choose the cell with the smallest number of numbers in its markup.

2	9	5	7	34	134	8	6	134
47	3	1	8	6	5	49	2	47
8	47	6	1249	2349	12349	459	1349	13457
1349	148	7	249	5	249	249	13489	6
1469	146	29	3	8	7	2459	149	145
5	48	2389	249	1	6	7	3489	348
367	678	38	5	2347	234	1	48	9
179	2	89	6	479	149	3	5	48
139	5	4	19	39	8	6	7	2

Figure 12. Markup of Mepham's diabolical puzzle.

In Figure 10 we have diagrammed an example search tree that has three paths, only one of which leads to a solution. The three possible paths through this tree are

$$\left(\underrightarrow{4}, \underrightarrow{4}\right), \left(\underrightarrow{4}, \underrightarrow{8}\right), \left(\underrightarrow{7}\right).$$

The path $\left(\underrightarrow{4}, \underrightarrow{8}\right)$—4 for cell A and 8 for cell B—leads to a solution, but the probability of choosing this path is only $\frac{1}{4}$, and therefore the odds against choosing this path are 3 to 1.

In the next section we display and solve a Sudoku puzzle whose solution requires the use of random choice.

The Mepham Diabolical Sudoku Puzzle

The puzzle appearing in Figure 11 was published by Mepham (2005, p. 14), who characterizes this puzzle as *diabolical* because it requires generating search paths on the fly. We shall refer to this puzzle as *Mepham's D*.

After entering the forced numbers in Mepham's D and marking it up, it is transformed into the puzzle shown in Figure 12.

The preemptive pair

$$\{[4,7], [c\,(2,1), c\,(2,9)]\}$$

in row 2 of Figure 12 isolates the singleton 9 in $c\,(2,7)$. Hence, we enter 9_1, which is shown in large boldface type in Figure 13. The preemptive triple

$$\{[2,4,9], [c\,(4,4), c\,(4,6), c\,(4,7)]\}$$

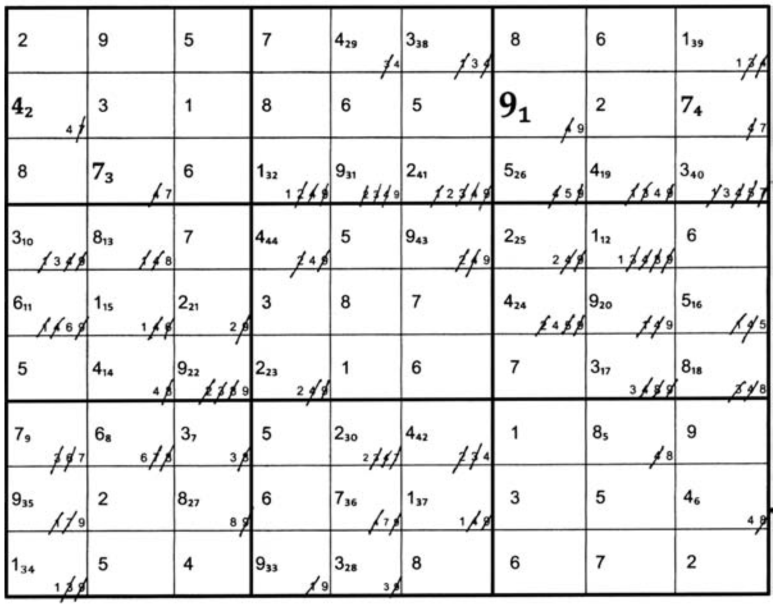

Figure 13. Solution of Mepham's diabolical Sudoku puzzle.

in row 4 in Figure 12 results in the preemptive triple $\{[1,3,8],[c(4,1),c(4,2),c(4,8)]\}$ in row 4 after appropriate crossing out.

Since there are no more undiscovered preemptive sets at this point, we choose the cell $c(2,1)$, which is a member of the preemptive pair $\{[4,7],[(2,1),(3,2)]\}$ in box 1, as the vertex of the search path that we will generate on the fly. We then randomly choose one of the two numbers $[4,7]$ in the markup of $c(2,1)$ by flipping a coin. The choice was 4. The red path is denoted by the boldface type in Figure 13 and is the sequence 4_2, 7_3, 7_4.

The red path in Figure 13 plays out quickly and does not, through crossouts, generate any new preemptive subsets, so we begin a green path by choosing randomly from the set $[4,8]$ in $c(7,8)$. The choice was 8. We denote the green path in Figure 13 by non-boldface type. The green path in Figure 13, which is the sequence 8_5, 4_6, …, 4_{44}, leads to the solution of Mepham's D displayed in Figure 13.

The search tree for Mepham's D is the one we gave as an example in Figure 10. In Figure 14 we display the solution of Mepham's D in "clean form".

Summary: Statement of the Algorithm

The steps in the algorithm for solving Sudoku puzzles are:

(1) Find all forced numbers in the puzzle.
(2) Mark up the puzzle.
(3) Search iteratively for preemptive sets in all rows, columns, and boxes—taking appropriate crossout action as each new preemptive set is discovered—until
(4) either
  (a) a solution is found or
  (b) a random choice must be made for continuation.
(5) If 4(a), then end; if 4(b), then go to step 3.

On the Question of the Uniqueness of Sudoku Puzzle Solutions

In the context of mathematics a Sudoku puzzle can be recast as a vertex coloring problem in graph theory. Indeed, just replace the set $[1,2,\ldots 9]$ in a Sudoku puzzle with a set of nine different *colors* and call the Sudoku board a *graph* and call the cells of the Sudoku board *vertices*.

The proper language to use here is to speak of coloring a graph in such a way that each of the nine colors appears in every row, column, and box of the puzzle. Such a coloring is a solution of the puzzle and is known in graph theory as a *proper coloring*. The minimum number of colors required for a proper coloring of G is called the chromatic number of G, and for Sudoku graphs the chromatic number is, of course, 9.

Now, the chromatic polynomial of a graph G is a function of the number of colors used to color G, which we denote by λ. The function computes the number of ways to color G with λ colors.

Herzberg and Murty (2007, p. 709) prove the following important theorem:

Theorem 3 (Completion Chromatic Polynomial). *Let G be a finite graph with v vertices. Let C be a*

2	9	5	7	4	3	8	6	1
4	3	1	8	6	5	9	2	7
8	7	6	1	9	2	5	4	3
3	8	7	4	5	9	2	1	6
6	1	2	3	8	7	4	9	5
5	4	9	2	1	6	7	3	8
7	6	3	5	2	4	1	8	9
9	2	8	6	7	1	3	5	4
1	5	4	9	3	8	6	7	2

Figure 14. Solution of Mepham's diabolical Sudoku puzzle, again.

partial coloring of t vertices of G using d_0 colors. Let $p_{G,C}(\lambda)$ be the number of ways of completing the coloring using λ colors to obtain a proper coloring of G. Then $p_{G,C}(\lambda)$ is a monic polynomial (in λ) with integer coefficients of degree $v - t$ for $\lambda \geq d_0$.

The value of $p_{G,C}(9)$ must be 1 for the coloring (solution) of the graph (puzzle) to be unique.[6] The computation of the integer coefficients of $p_{G,C}(9)$ is easily done by running an implementation of the inductive proof of Theorem 3 (see Herzberg and Murty, p. 710). The inductive proof is essentially the *deletion-contraction* algorithm, which is thoroughly discussed by Brualdi (Brualdi, 2004, p. 529).

References

[1] Richard A. Brualdi (2004), *Introductory Combinatorics*, 4^{th} ed., Pearson Prentice Hall, Inc., Upper Saddle River, NJ.

[2] Agnes M. Herzberg and M. Ram Murty (2007), Sudoku squares and chromatic polynomials, *Notices Amer. Math. Soc.* **54**(6), 708–717.

[3] Michael Mepham (2005), *Solving Sudoku*, Crosswords Ltd., Frome, England. This article is available online at the Web address http://www.sudoku.org.uk/PDF/Solving_Sudoku.pdf.

[4] Tom Sheldon (2006), *Sudoku Master Class*, Plume (Penguin), New York.

[5] Will Shortz (ed.) (2006), *The Little Black Book of Sudoku*, St. Martin's Griffin, New York, NY.

[6]*Herzberg and Murty (2007, p. 712) prove that a necessary condition for a unique coloring (solution) is that C, the puzzle, must use at least eight colors (numbers), that is, $d_0 \geq 8$.*