CATALAN NUMBERS AND APPLICATIONS

Aybeyan Selimi, Muzafer SARAČEVIĆ **page 99-114**

ABSTRACT

Catalan numbers have a significant place and major importance in combinatorics and computer science. They form a sequence of natural numbers that occur in studying astonishingly many combinatorial problems. They appear in the triangulation problem of polygon and polyhedron, binary trees, multiplication ordering, lattice path problem, etc. Today, application of the Catalan numbers we see in engineering in the field of computational geometry, geographic information systems, geodesy, cryptography, and medicine. In the problems of computational geometry, they are generally used in geometric modeling. In cryptography are used in the forming of keys for secure transfer of information. In this study, we consider Catalan numbers, their properties, generating function and related problems with them.

Keywords. Catalan numbers, triangulation, lattice path, multiplication ordering and binary tree.

Aybeyan Selimi Ph.D,

Faculty of Informatics, International Vision University, Gostivar - North Macedonia;
e-mail:
aybeyan@vizyon.edu.mk

Prof. Muzafer SARAČEVIĆ PhD
Associate Professor, Departman of Computer Science, International University of Novi Pazar, Serbia,
email:muzafers@uninp.edu.rs

UDK: 511.176

Date of received:
20.01.2019
Date of acceptance:
21.02.2019

Declaration of interest:
The authors reported no conflict of interest related to this article.

1. Introduction

In this paper we give some of applications of Catalan numbers. Catalan numbers appear as a sequence in many combinatorial problems such as polygon triangulation, balanced parenthesis, lattice path, Dyck's paths, the problem of binary trees, multiplication ordering and in the many other problems. These interesting numbers were first encountered by Leonhard Euler (1703-1783) and Johann Andreas von Segner (1704-1777), by studying the problem of triangulation of the convex polygon. The problem was first resolved by Euler in 1760, who by using combinatorial tools obtain the number of triangulations in the convex polygon. Recursive relations of this numbers are introduced from Segner, while many of the properties and identities of these numbers by studying well-formed sequences of parentheses in 1838 are discovered by side of Eugene Charles Catalan (1814-1894). In his honor, these numbers today are called Catalan numbers.

Definition. *Catalan numbers C_n are integer sequence defined by*

$$C_n = \frac{(2n)!}{(n+1)!\,n!} = \frac{1}{n+1}\binom{2n}{n}, \quad n \geq 0 \tag{1.1}$$

Catalan numbers are implemented in the Mathematica Sofware Package as

CatalanNumber[n].

The first 10 values of Catalan number given in the table below are find with code:

CatalanNumber /@ Range[0, 10]

$$C_n = \{1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796\}$$

There is an alternative way of defining C_n:

$$\binom{2n}{n} - \binom{2n}{n-1} = \frac{(2n)!}{(n!)^2} - \frac{(2n)!}{(n-1)!\,(n+1)!} \qquad (1.2)$$
$$= \frac{(2n)!}{n!\,(n+1)!} = C_n$$

These numbers were completely independent discovered from mongolian mathematician Ming An-Tu (1692-1763) in 1730. Ming's work was published in Chinese and was not known in the West long time. He used this sequence of numbers to express $sin(2x)$ and $sin(4x)$ in terms of $sin(x)$, i.e

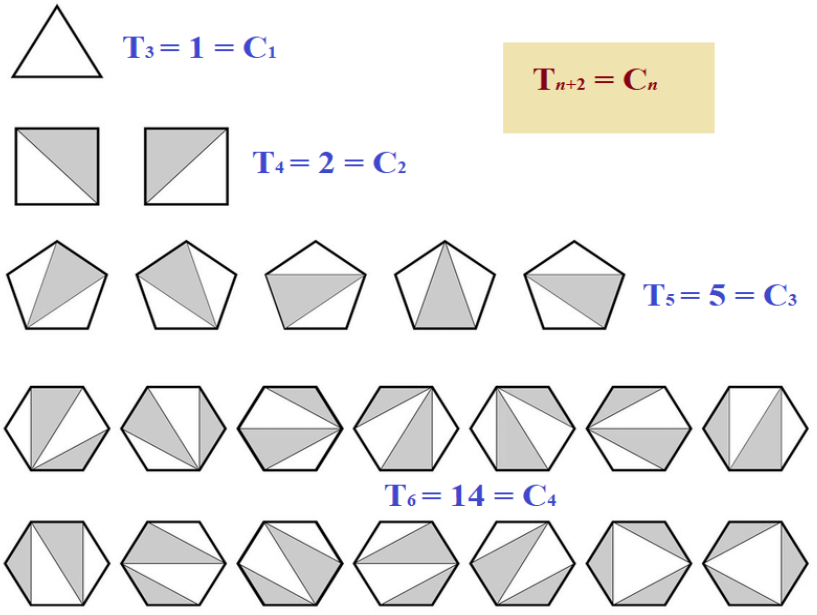
$$sin2x = 2sinx - \sum_{k=1}^{n} \frac{C_{k-1}}{4^{k-1}} (sinx)^{2k+1}$$

2. Related Problems with Catalan Numbers

2.1 Triangulation of convex n – gon

We will denote the convex polygon with *n* sides with P_n. It is described by a sequence of vertices $v_1, v_2, \dots, v_n$. The inner diagonal connecting the vertices v_i and v_j is marked with δ_{ij} . Also, and the polygon sides are considered as diagonals, so also $\delta_{i,i+1}$ represents the side $v_i v_{i+1}$. Set of triangles of the maximal way of decomposition of a convex polygon P_n to the $n-2$ triangles is denoted with T_n . For triangulation is necessary to draw $n-3$ non-intersect diagonals (O'Rourke, 1997). Since the triangle is already triangulated, there is only one way of triangulation and therefore $T_3 = 1$. For a square ($n = 4$) we can draw one diagonal. This can be done in two ways (because the square has two diagonals) so it is $T_4 = 2$.

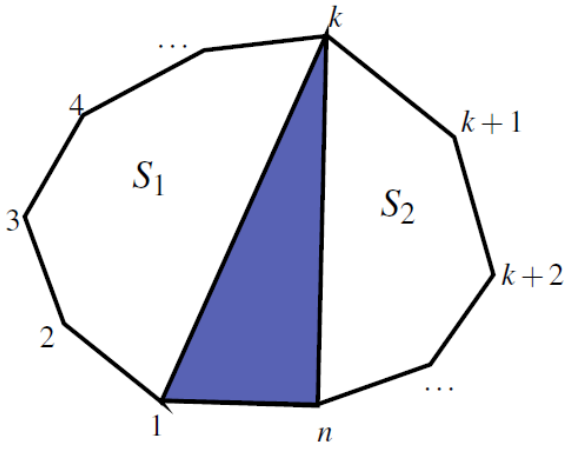
Figure 2.1 Triangulations of an *n*-gon, where $3 \leq n \leq 6$



For the pentagon $(n = 5)$ solution is less obvious, there are 5 ways of triangulation. Let us now look at the general solution for the number T_n of triangulation of $n-$gon. Note that each $n-$gon side is a part of the triangle in a triangulation. For counting of different triangulations of a convex polygon, we will use recursion and the following way. Randomly select and fix one particular side and call him the base. We count the triangulations in which each of the triangles has the base $(1,\ n)$ as a side. In every triangulation the edge $(1,\ n)$ is a side of exactly one triangle. In the polygons $S_1$and S_2 there are k and $(n - k + 1)$ vertices, respectively. Since the polygons $S_1$and S_2 can be triangulated independently, the number of triangulations in S_1 is T_k , and in the S_2 is T_{n-k+1} . We define $T_2 = 1$.

Since the triangulation choices of the polygons in decomposition are independent of each other, the combinatorial principle of the product is valid, and for arbitrary points k is equal to the $T_k T_{n-k+1}$, see Figure 2.2.

Figure 2.2 Polygonal Dissection (De Loera, 2010: 3)



The choice of this point can also be done in several (independent) ways, so we still have to go through all the possible values of k. The final form of recursive formula for triangulation T_n is:

$$T_n = \sum_{k=2}^{n-1} T_k T_{n-k+1} = T_2 T_{n-1} + T_3 T_{n-2} + \cdots T_{n-1} T_2 \,, \qquad n \geq 3$$

Therefore, the value of the Catalan number C_{n-2} determines the number of triangulations corresponding to the polygon P_n (Saračević, 2013: 5). Based on the formula (1.1) for obtaining the values of Catalan numbers we can define value T_n, for $n \geq 3$:

$$T_n = \frac{1}{n-1} \cdot \binom{2n-4}{n-2} = \frac{(2n-4)!}{(n-1)!\,(n-2)!}$$

The value of T_{n+2} also can be expressed equivalent form to (2.9):

$$T_{n+2} = \frac{2}{n+1} \cdot \binom{2n-1}{n} = \frac{1}{n+1}\binom{2n}{n} = C_n$$

Since $C_n = T_{n+2}$, for $n \geq 1$ and $C_0 = 1$ we obtain the *Segner's recurrence relation* for C_n:

$$C_n = \sum_{k=0}^{n-1} C_k C_{n-k-1}$$

2.2 Catalan Numbers and Multiplication Ordering/ Balanced Parenthesis

Suppose we have $n+1$ numbers to multiply together, meaning that there are n multiplications to perform. By changing only the multiplication order, without changing the order of the numbers themselves this numbers can be multiplied in many orders (Stanley, 2015). Multiplication is a binary operation and therefore in the multiplication of the $n+1$ numbers first we have to multiply the two, then multiply their multiplicity by the next element and so on. Multiplication is ended when all numbers are multiplied. To show the multiplication ordering explicitly, we will use parenthesis. For $n = 1$ we have 2 numbers to multiply and 1 multiplication $(a \cdot b)$, $n = 2$ we have 3 numbers to multiply and 2 different multiplication ordering $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. For $n = 3$ we have 4 numbers to multiply and 5 different multiplication ordering $(((a \cdot b) \cdot c) \cdot d) = ((a \cdot b) \cdot (c \cdot d)) = ((a \cdot (b \cdot c)) \cdot d) = (a \cdot ((b \cdot c) \cdot d)) = (a \cdot (b \cdot (c \cdot d)))$, and so on.

The multiplication ordering is equivalent to the problem of balanced parenthesis variables. To show that the problem of the balanced parenthesis satisfies the Catalan number sequence at the first we look the problem of balanced parenthesis for *n* numbers $a_1, a_2, \dots, a_n$. Let the number of multiplications of these *n* numbers is marked with M_n. Then from discussion above $M_0 = 1 = M_1$.

The first i pairs can be correctly grouped in M_i ways and the remaining $n - 1 - i = n - i - 1$ pairs in M_{n-i-1} ways. Using the multiplication principle, these two events can take place together in $M_i\, M_{n-i-1}$ different ways. From addition principle we have

$$M_n = M_1 M_{n-1} + M_2 M_{n-2} + \cdots \dots M_{n-1} M_1$$

From this equation, we see that M_n satisfies the Segner's recurrence formula as a C_n, thus $M_n = C_n$.

2.3 Triangulation and Binary Trees

A rooted tree is called an *m*-ary tree if every internal vertex has no more than *m* children. The tree is called a full *m*-ary tree if every internal vertex has exactly *m* children. An *m*-ary tree with $m = 2$ is called a binary tree (Rosen, 2012: 747). From the definition of the binary tree for n ≥ 1 follows that the number of B_n binary trees with n vertices is equal to the number of arranged pairs (B, B') of binary trees with a total of $n - 1$ vertices, where the tree B has $n - k + 1$ vertices, and B' has k. Again from the rule sum and the product follows

$$B_n = B_1 B_{n-1} + B_2 B_{n-2} + \cdots \ldots + B_{n-1} B_1$$

A binary tree with the degree of root one and each internal vertex three is called the planted trivalent binary tree (Koshy, 2009: 235). There is a bijection between the set of binary trees and the set of planted binary trees. By attaching a new root at the existing root of a binary tree, we get a planted binary tree, by deleting the root of a planted trivalent binary tree, we get an ordinary binary tree. It means every planted binary tree with n vertices correspondent to the binary tree with $n - 1$ vertices. There is a bijection between the set of triangulations of an n −gon and the set of balanced parentheses with n pairs, where each containing C_n elements. From the equation between the number of planted trivalent binary trees with *n* vertices and the number of binary trees with *n* - 1 vertex, we can establish a one-to-one correspondence between polygonal triangulations and planted trivalent binary trees. There is a bijection between the set of triangulations of a convex n −gon and the set of planted trivalent binary trees with $n - 1$ leaves (Saracevic, 2019).

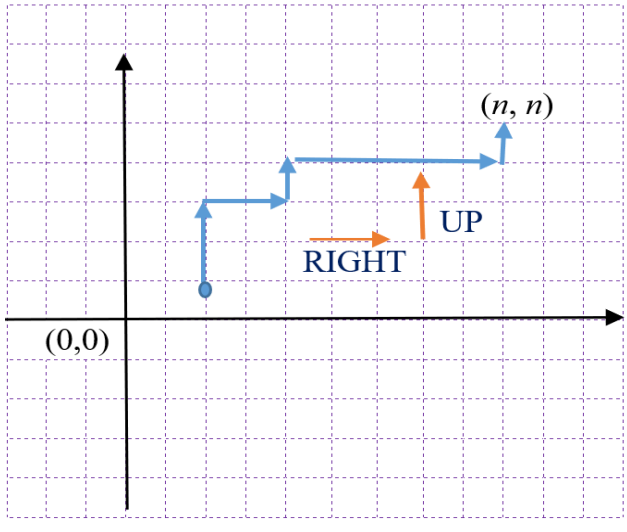
2.4 Catalan Numbers and Lattice Paths

The integer lines parallel to coordinate axes in the Cartesian coordinate system form the lattice of integer lines or discrete lattice. The set $L = \{(m,n): m,n = 0,1,2,\dots\}$ are the points of the lattice and the lines joining these points are called the edges of the lattice. Let's fix two points (m_1,n_1) and (m_2,n_2) in this lattice, such that $m_2 \geq m_1$ and $n_2 \geq n_1$. A **increasing/lattice** path from (m_1,n_1) to (m_2,n_2) is a subset $\{e_1,e_2,\dots,e_k\}$ of L such that:

- either $e_1 = (m_1,n_1 + 1)$ or $e_1 = (m_1 + 1,n_1)$;
- either $e_k = (m_2,n_2 - 1)$ or $e_k = (m_2 - 1,n_2)$; and
- if we represent the tuple $e_i = (a_i,b_i)$, for $1 \leq i \leq k$, then for $2 \leq j \leq k$,
  - either $a_j = a_j - 1$ and $b_j = b_j - 1 + 1$
  - or $b_j = b_j - 1$ and $a_j = a_j - 1 + 1$.

The movement on the lattice is either to the **right** or **up**, see Figure 2.3.

Figure 2.3 A lattice with a path



Let's look the $n \times n$ size discrete lattice. How many different paths can be drawn in this discrete lattice? That is, the number of possible lattice paths from $(0,0)$ to the lattice point (n,n) on the discrete lattice such that from any lattice point (x,y), we can walk one block right $(\boldsymbol{R})$ or one block up $(\boldsymbol{U})$, where $x,y \geq 0$. Every path can be represented by a word made up of exactly n $\boldsymbol{R}'s$ and n $\boldsymbol{U}'s$. So, the total number of paths through the discrete lattice to the point (n,n) is $\binom{2n}{n}$. The shortest paths in the lattice are the paths that do not go above the diagonal $(y = x)$. The number of shortest lattice paths or valid paths from $(0,0)$ to (n,n) on the $n \times n$ grid is given by

$$\binom{2n}{n} - \binom{2n}{n-1} = \frac{(2n)!}{n!\,n!} - \frac{(2n)!}{(n+1)!\,(n-1)!} = \frac{(2n)!}{(n+1)n!} = C_n$$

2.5 Catalan Numbers and Dyck Paths

The problem of Dyck paths consists of drawing all possible mountain ranges consisting of *n* upstrokes and *n* downstrokes, assuming they are always legal. If $n = 0$, it can be drawn only one path, for $n = 1$, there is also only one path. For $n = 2$ it is possible to draw two and for $n = 3$ five different paths. By replacing each upstroke with an open parenthesis and a downstroke with a closed parenthesis is not hard to see that the problem of the Dyck mountain paths corresponds to the problem of balanced parenthesis.

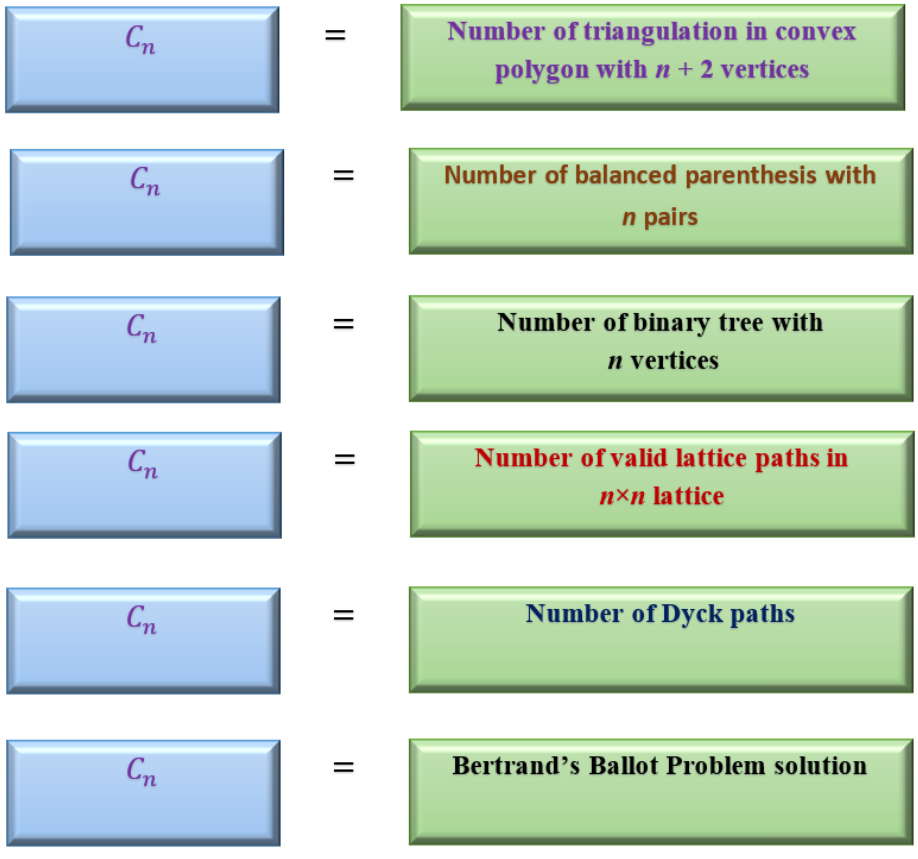
Figure 2.4 Number of Dyck paths

$n = 0$:	*	1 way
$n = 1$:	/\	1 way
$n = 2$:	/\ /\/\, / \	2 ways
$n = 3$:	/\ /\ /\ /\/\ / \ /\/\/\, /\/ \, / \/\, / \, / \	5 ways

2.6 Bertrand's Ballot Problem

Consider an election with two candidates ***A*** and ***B*** where *n* voters vote one at a time. The candidate ***A*** receives *a* votes and candidate ***B*** receives *b* votes, where $a \geq kb$ for $k \in \mathbb{Z}^+$. For example, $\boldsymbol{w} = \boldsymbol{ABAABBBA}$ means Voter 1 voted for Candidate A, Voter 2 voted for Candidate B, and so on. Bertrand's ballot problem is to determine the probability (assuming that all voting histories are equally likely) that Candidate A is always ahead of Candidate B in the election (Roman, 2015: 9). The number of ballots in which Candidate A has *k* times more votes then Candidate B is $\frac{a-kb}{a+b}$. This problem is reminded of the other problems above mentioned, and all they are in the correspondence with the Catalan numbers, see Figure 2.5.

Figure 2.5 Catalan numbers correspondences



3. Applications

Computational geometry is a discipline of computer science that deals with the research of algorithms from the aspect of geometry (Selimi, 2019a). As a branch of computer science is dedicated to the study of algorithms that can be expressed in terms of geometry. Some of these studies the purely geometric problems, while others are obtained as the corollary of examining computational geometric algorithms. Algorithms of computational geometry today are applied in numerical computation, geometric modeling, computer vision, computer graphics, geodesy, dynamic computing, isothetic computational geometry, and parallel computing.

In (Stanley, 2011) are listed 95 problems that are solved with Catalan numbers. Catalan numbers as a correspondence of triangulation problem are applied in various areas of computational geometry, cryptography, medicine, and banking. Triangulation is a procedure which is applied in the representation of three-dimensional objects. As a procedure provides the glazing mechanism which is used for approximation of the three-dimensional objects with polygons. The techniques of triangulation also are applied in obtaining of triangular networks in the surface decomposition which are presented with the three-dimensional image. These triangular networks share the image in several non-overlapping regions and have similar characteristics (Selimi, 2018).

Catalan numbers in combination with dynamic programming technique in the paper (Selimi, 2019) are used for obtaining and storing of minimum weighted triangulation of polygons. In this paper optimal triangulation is based on matrix chain product and memorization method. The values optimal triangulations are stored in a table which is divided with its diagonal into two parts.

Throughout the history of mankind, there was a need for a secure exchange of information. The problem of secure communication has already been dealt with by Egyptians and Indians more than 3,000 years ago, and since then the basic idea has not changed - to transfer a message from one place to another that is safer, to make an algorithm that would allow the original message to be hiding so that it is completely incomprehensible to persons who would have unauthorized access to its property. The first methods used were not complex mathematical algorithms, but began using alternative languages that were known only to a small number of people.

The development of more complex methods of secure communication began only with the development of a letter, which allowed for any information to be displayed by a certain number of characters that would, after using a particular key, form a re-initial message. Eventually, the idea of displaying letters with other symbols appeared. Examples that are still in use today are Morse code, Braille letter, and ASCII code. The task of cryptography is to enable two persons (sender and receiver) to maintain the secrecy of messages, even communicating with an unstable communication channel (computer network, telephone line), which is available to third parties. The discipline that allows this is cryptography. Cryptography has been used for centuries to ensure the secrecy of predominantly military and diplomatic communications. Cryptography is the science of preparing secret codes and breaking them. Protocol and algorithms are used to ensure communication security. Today, encryption methods are used in many different ways to ensure secure transfer of information. Cryptography is a very dynamic, current, and widespread discipline, a contribution to the application of combinatorics and computational geometry (Saračević, 2018: 30).

The importance of Catalan numbers in the cryptography, are seen primarily in the development of algorithms for generating binary sequences that are necessary for generating keys. The sequence of these numbers are applied a simple polygon triangulation algorithm in the process of generating hidden cryptographic keys from one segment of a 3D image is represented.

In the paper (Saračević, 2017a), is shown the other application of Catalan numbers in the encryption process. Catalan keys are implemented on the ballot problem, stack permutation, and balanced parentheses. The ballot problem is a problem which determines the combinations there are to put the 2n votes in such way that in each adding a new vote, the number of votes that have been won by candidate A is greater or equal to the votes of candidate B. In this paper are covered some mathematical concepts of cryptography and is represented a contribution of Catalan numbers in this field. Also, this paper contains information about bit balance property in the binary notation of the Catalan numbers. There are given a few suggestions and examples of Catalan number application in encrypting files. The emphasis is placed on the application of stack permutation in text encryption, where also Catalan numbers are used in the encoding.

One other application of Catalan numbers are seen in the paper (Saračević, 2017b). In this research, the authors represent an application of the decomposition method in the dynamic generation of the valid binary Catalan-keys. In the paper is presented one example of Catalan numbers decomposition and their application in steganography. The sequences are used in the form of combinations of Catalan numbers sequences – decomposition of Catalan numbers. In developed technique, sequence represents one set of constants and the other sequence is represented by variables which are conditioned by the given variables, that is, they are

calculated based on them. The formula for Catalan numbers serve as a generator for the first numbers sequence (constants) and the decomposition method serves as a generator for the second sequence (variables). These two sequences are interchangeably connected. The decomposition method generates variables based on constant and avoids the application of Catalan numbers with other numbers like Fibonacci or Lucas numbers.

A new application of Catalan numbers, primarily as a generator of pseudo-random numbers in combination with several combinatorial problems, with the purpose of text encryption and decryption, is given in (Saračević, 2018c). This paper shows the application of lattice path in text encryption.

There are provided examples and are given experimental results of text encryption speed for some combinatorial encryption methods ballot problem, stack permutations and balanced parentheses in comparison with Lattice Path. Also is applied the NIST statistical test battery in assessing of the quality of the Catalan keys.

A novel approach to steganography based on the properties of Catalan numbers, Dyck words and machine learning is presented in the (Saračević, 2019e). In this paper, authors show the application of Catalan numbers in a data hiding method and Dyck words. According to the analysis of this research is concluded that computing resources limit the entire Dyck words set generation process, hence that brute force steganalysis is impossible.

4. CONCLUSION

In this paper, we give some of the application of Catalan numbers. Today, we can see their application as a correspondent of the triangulation method in the biometric identification process. There are many polygon triangulation methods in field of computational geometry which are based in Catalan numbers and are used in face recognition techniques. Steganography and visual cryptography give the possibility of application the Catalan numbers in quantum cryptography. Quantum cryptography and DNA will present the basis for the protection of confidential documents in the near future. According to that, we can expect the scientific works with application of Catalan numbers in medicine, banking, e-commerce, networks, and information security. Further works direction for Catalan numbers application lies in their combination with cyclic sieving phenomenon in the building of triangular networks in convex and concave polygons and polyhedron.

5. REFERENCES

1. De Loera A. J., Rambau J. and Santos F., Triangulations Structures for Algorithms, Springer-Verlag Berlin Heidelberg, Volume 25, 2010
2. Koshy, T. Catalan Numbers with Applications, Oxford University Press, New York, 2009.
3. O'Rourke J., Computational Geometry in C, Second Edition, Cambridge University Press, 1997
4. Richard P. Stanley, Catalan Numbers, Cambridge University Press 2015.
5. Richard P. Stanley, Enumerative Combinatorics, Volume 1, 2nd edition, 2011.
6. Roman S., An Introduction to Catalan Numbers, Springer Cham Heidelberg New York Dordrecht London, 2015
7. Rosen Kenneth H., Discrete Mathematics and Its Applications, Seventh Edition, McGraw-Hill Companies, USA, 2012.

8. Saračević M., Selimi A., and Selimovic F., Generation of cryptographic keys with algorithm of polygon triangulation and Catalan numbers, Computer Science, 19(3), 2018
9. Saračević, M. Application of Catalan numbers and some combinatorial problems in cryptography (Bachelor's thesis), Faculty of Informatics and Computing, Singidunum University in Belgrade, 2017.
10. Saračević, M. Methods for solving the polygon triangulation problem and their implementation (PhD thesis), Faculty of Science and Mathematics, University of Niš, 2013.
11. Saračevic, M., Koričanin, E., Biševac, E. (2017a). Encryption based on Ballot, Stack permutations and Balanced Parentheses using Catalan-keys. Journal of Information Technology and Applications, 7 (2), 69–77.
12. Saracević, M., Hadzić, M., Koričanin, E. (2017b). Generating Catalan-keys based on dynamic programming and their application in steganography. International Journal of Industrial Engineering and Management, 8 (4), 219-227.
13. Saračević, M., Adamović, S., Biševac, E. (2018c). Applications of Catalan numbers and Lattice Path combinatorial problem in cryptography. Acta Polytechnica Hungarica: Journal of Applied Sciences, 15 (7), 91-110.
14. Saračević, M., Adamović, S., Miškovic, V., Maček, N., Šarac, M. (2019e). A novel approach to steganography based on the properties of Catalan numbers and Dyck words, Future Generation Computer Systems - Elsevier, 100, 186 – 197.
15. Saracevic M., Selimi A., Convex polygon triangulation based on planted trivalent binary tree and ballot problem, Turkish Journal of Electrical Engineering and Computer Sciences, 2019; Vol. 27, pp. 346-361.
16. Selimi A. and Saračević M., Computational Geometry Applications, Southeast Europe Journal of Soft Computing., Vol.7, No.2, 2018.
17. Selimi, A. Algorithm of computational geometry and their application in linear optimization and dynamic programming (PhD thesis), Department of Computer Science, University of Novi Pazar, 2019a.
18. Selimi A., Krrabaj S., Saračević M. and Pepić S., Memoization method for storing of minimum-weight triangulation of a convex polygon, Computer Science 20(2) 2019: 195–211
19. Wilf Herbert S., Generatingfunctionology, Third Edition, A K Peters, Ltd, 2006.