

Journal of Financial and Actuarial Mathematics and Management

A joint edition of the Institute of Information Sciences and Institute of Social Sciences of Perun Academy
Publishing Company: PSP Publishing
ISSN: 2367-5810
Issue 10, Year X (2022)
Founding Editor and Editor-in-Chief: Pavel Stoynov

Editorial Board: Hiroyuki Ono (Japan), Yasukazu Yoshizawa (Japan), Naoyuki Ishimura (Japan), Boris Melnikov (Russia), Elena Melnikova (Russia), Goran Karanovic (Croatia), James Osondu (Nigeria)

Contents

Y. Algani. An entire study about the development
of newer methods for transportation problem optimization……..……...3
P. Stoynov. Bulgarian travel operators market from
insurance industry perspective ..…….....9
D. Gotseva, R. Trifonov, P. Stoynov. Building a Model for Network Intrusion Detection with
Artificial Neural Networks……..…….12

An entire study about the development of newer methods for transportation problem optimization

Yousef Methkal Abd Algani
Faculty of Education, Department of Teaching
Mathematics, The Arab Academic College for Education in Israel-Haifa.
yousefabdalgani@gmail.com
ORCID: 0000-0003-2801-5880

Abstract.
The transportation problem is a widely use optimization technique to analyze the problems concerned with transporting goods.The transportation sector has become eminent based on the economic, social factors and also it influences the relationships existing among various countries. With the movement of goods and people, the transportation sector promotes the travel and tourism sector. The transportation problem is one among the original endeavors of a linear programming model. It is an optimization problem having linear constraints and linear objectives.The main goal of the transportation problem is minimization of transportation cost, minimization of distance with respect to time and so on. There are several optimization techniques developed in recent years, to solve the transportation problem.This explorative paper explores about the transportation problems and the recent advancements in the optimization methods developed for decoding the transportation problems. The proposed exploration provides insights about the methods used for finding initial basic feasible solution to a transportation problem and the newly developed methods based on these techniques. There are many alternative methods for providing solutions to the transportation problems. For any transportation problem, the usual methods, which are employed to find the initial basic feasible solution is the Vogel's Approximation Method (VAM), the Least Cost Method (LCM) and the North-West Corner method. This research helps in researching further about the developing cost-effective methods that assists in minimizing the transportation costs.

Keywords: Transportation cost, transportation problem, feasible solution, optimal solution etc.

Introduction

The transportation sector forms an eminent part of the human beings. The movement of things is essential to be done and they are achieved with the help of transportation systems. These transportation systems possess high efficiency, speed and development over time with a drastic enhancement (impoff, 2020). The eminence of transportation is highlighted in how the governments, businesses and individuals rely on this sector for accessing the resources. An optimal functionality cannot be contributed by a society, when it lacks transportation facilities. A sustainable development and an enhanced total productivity is contributed by the transportation sector. Economies, job, resources and people require indefectible movement for retaining the further advancement of the whole world. The transportation problem is one among the original endeavors of a linear programming model. It is an optimization problem having linear constraints and linear objectives. Lots of special structures are exhibited by the transportation problem. For instance, each of the variables occurs in accurately with two non-zero coefficients. The variable encompassing this coefficient possesses either negative or positive value. There are many alternative methods for providing solutions to the transportation problems. For any transportation problem, the usual methods, which are employed to find the initial basic feasible solution is the Vogel's Approximation Method (VAM), the Least Cost Method (LCM) and the North-West Corner method. For a transportation problem, the optimality is checked with the MODI method. The newer methods for transportation problems' optimization is designed with the alterations of the above mentioned methods.

Theoretical background of the study

The theoretical concepts and terms of the proposed research are discussed here.

Operations Research

Operations research is defined as the application of quantitative methods for helping analysts and decision-makers for enhancing, analyzing and designing the operation or performance of the systems. The operations research field encapsulates analytical tools from various fields that could be embedded in an efficient way for assisting the decision-makers in resolving various problems in an advantageous or practical way. A system's performance could be optimized with the tools of operations research. Probability and statistics, optimization and simulation are eminent features of an operations research. Identification of a problem,

construction of a model having variables in the real-world; usage of this model for deriving resolutions for a problem and testing and analyzing the solutions of the implemented model are eminent steps of operations research. Some endeavors of operations research are; risk management, packet routing optimization, inventory management, time management and scheduling, agricultural and urban planning, engineering and network optimization, supply chain management and in enterprise resource planning.

Linear programming problem

A linear programming problem encompass finding of a minimized or maximized subject involving a linear function to some constraints in linear inequalities or equalities form. This programming model is employed for gaining the greater optimum solution for a constraints carried transportation problem. Some of the types of a linear programming model are; linear programming model with an open solver, linear programming with graphical method, linear programming by R method, linear programming by simplex method. The components of a linear programming model are; objective functions, decision variables, data and constraints. Non-negativity, finiteness, linearity are some features of linear programming. Some of the linear programming applications are; transportation optimization, energy industry, efficient manufacturing and engineering.

Transportation problem

A transportation problem is one among the forms of a linear programming problem, which involves the transportation of services and goods to various destinations from various sources. This is in accordance with the demand and supply of the suitable destinations and sources with the lowest minimum cost of transportation. The main objective of a transportation problem is the minimization of the transportation cost.

Mathematical representation of transportation problems

A transportation problem is concerned with the movement of products from m number of supply sources to n number of distribution units or destinations. Minimization of the total transportation cost is the main objective. O is the supply source and D is the destination. Hence goods are transported from O_1, O_2, O_3……O_m to D_1, D_2, D_3…….D_n. The supply units are indicated by a_i and the demand units are indicated by b_j and c_{ij} is the cost of transportation per unit from O to D. here i= 1,2, 3, 4,…., m and j =1, 2, 3, 4,….,n. For fulfilling the demand and supply needs with the low transportation cost, the associated linear model is expressed as follows;

$$\text{Minimum } z = \sum_{i=1}^{m} \sum_{j=1}^{n} xijcij$$

Which is in accordance with the constraints, $\sum_{i=1}^{m} xij \geq bij and \sum_{j=1}^{n} xij \leq aij$ and $x_{ij} \geq 0$ for i and j values. Thus, these are the standard mathematical formulations of a transportation problem.

Initial basic feasible solution

A group of non-negative allocations, where there is a satisfaction of columns and row restrictions is referred to as a feasible solution. A feasible solution for origin m and destination n is regarded as a basic feasible solution where the number of optimistic allocations is m+n=1.

Optimum solution

The optimum solution is a type of feasible solution, which involves the reaching of the objective function to minimum or maximum value; for instance, the highest least cost or the profit. When there are no feasible solutions having functional values, then it is called a global optimum solution.

Balanced transportation problem and unbalanced transportation problem

When the total supply units' number equals with the total demand units' number, then it is called a balanced transportation problem. When the supply units' number and the demand units' number are unequal, then it is an unbalanced transportation problem.

Literature Review

Wagener (1965) developed a new model for providing solutions to the transportation problems. There a new procedure was introduced for the transportation model computation, where every column in the cost matrix is utilized and thus the method was suitable for the computers having tape storage. However, some limitations and errors were identified in his proposed model, which lead to further research on the field.

Ekanayake et.al (2020) developed a modified ant colony optimization technique for getting transportation problem solution. A modified ant colony optimization technique was proposed by them with a meta-heuristic algorithm for getting the initial basic feasible solution for transportation problems. The model possessed finite iterations and was simple in execution. The accuracy of the algorithm was made with numerical examples and it offered satisfactory results. With their new algorithm, the time consumption of the transportation problems is reduced.

Amaliaha, Fatichaha and Suryanib (2020) constructed a newer heuristic method for determining the initial basic feasible solution for a transportation problem. This new method was indicated as Bilqis Chastine Erma (BCE) method. 35 numerical archetypes were utilized for evaluating their new method's performance. The experimental results of the study made a conclusion that this newer method provided high efficiency of 88% when compared with the other transportation methods.

Shaikh, Shah and Memon (2018) constructed an enhanced algorithm for getting initial basic feasible solution for a transportation problem. The new method was tested for accuracy against the North-West Corner method and the Least Cost methods, utilizing five numerical archetypes. It was concluded that much reliable results were got with the new method, when compared to the other methods.

There are other new methods developed and designed with some alterations made to the North West corner method, least cost method and Vogel approximation method.

Methodology

The traditional optimization techniques were helpful in determining the unconstrained minima or maxima of differentiable and continuous functions or the optimum solution. The above are analytical techniques and they utilize different calculus for optimum solution location. Some of the numerical optimization techniques are, combinational optimization, constraint satisfaction, dynamic programming, stochastic programming, non-linear programming, quadratic programming, integer programming, linear programming and dynamic programming. Some of the advanced optimization methods are, hill climbing, simulated annealing, genetic algorithm, ant colony optimization etc (nptel, 2020). Among the other eminent optimization techniques, the north-west corner rule, Vogel approximation method and least cost method are important techniques of optimization that helps in finding the initial basic feasible solution for a transportation problem.

North West Corner Method

In the North-west corner method, focus must be made on the North-west comes of the transportation table to find out the initial basic feasible solution. There are three important steps in a North-west corner rule which are as follows.

- Determine the cell with the North-west corner of the transportation table. Allocation must be possibly made to the chosen cell with the respective adjustment of the related quantities of demand and supply by extracting the amount allocated.
- Then the column or row that has zero demand for supply must be crossed out. When both a column or a row posses a o, then that column or row must be randomly crossed out .The following shows an empty transportation table having supplies in the right side and at the bottom the demands as shown in the below table.

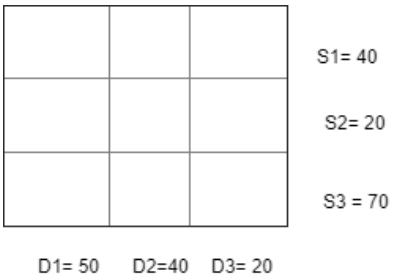


Figure 1: An empty transportation table with demand and supplies

Here, we initially choose the cell on the North-West Corner and its associated value is a minimum of S1 and D1. When D1 is greater than S1, then the first row must be crossed out. This step must be repeated until we get a table with the basic feasible solution.

New methods developed out of North West Corner Method

Metlo et.al (2016) made a modification in the North-West Corner method for making optimization of the transportation problems. This newer method is modified forms of the North West Corner Method to get the initial basic feasible solution along with culminate entries required for operating the transportation that is utilized for obtaining the optimal solution. In linear programming, the transportation problem is an eminent model class and is associated with the transportation of commodities to multiple destinations with the purpose of reducing the transportation steps or time. The aim of their modification was to search for the cheapest and best way, the suppliers are utilizing for satisfying the demand at the eminent points. There would be an adjustment from the distributor to the points of demand during a change in the shipping products cost. Then a comparison was made between the North West Corner method and the modified North West Corner method. The comparison results were equal and the newly modified method was greater in power when compared with the conventional North West method.

Sharma and Bhadane (2016) developed an alternative model of the North-west corner method to solve the transportation problem. Their new models is developed by the help of a statistical tool, which is referred to as co efficient of range. Numerical archetypes were explained in their paper for justifying their results. These numerical archetypes illustrated that the entire transportation cost, which is got by the utilization of the coefficient of range (COR) provides optimal results when compared with the North-west corner method.

In addition, the transportation cost which is got with the new COR method was similar to the least cost method. A non-degenerate solution was obtained with the North-west corner method. It showed that the initial basic feasible solution got with the COR method is very less. When compared with the iterations number.

Least Cost Method (LCM)

Least cost method (LCM) is one among the methods, which is used for getting the initial basic feasible solution in a transportation problem. Here, the cell having the minimum transportation cost is allocated first. These cells of low cost are selected over the cells of high cost with a view for getting the minimum transportation cost or least transportation cost. More solutions for transportation with optimality are generated by the Least Cost method, when compared against the North-West Corner method, as it takes into account the cost of shipping during the allocation process. Only the supply needs and availability are considered in the North-West Corner method, which is independent of the shipping cost. The demand and supply must be equal and when there is more supply, there is an addition of a dummy source in the transportation table and the demand value equals the demand and supply difference value and zero is the cost. When the demand is higher than the supply, there is an addition of a dummy origin or destination to the transportation table and the supply value is equal to the difference between demand and supply and zero is the cost value. The following shows the steps of optimization of the transportation problems using the least cost method (LCM).

- Initially we have to check whether a given transportation problem is balanced or not. When the supply is not equal to demand or in an unbalanced condition, a dummy market or source is added to the transportation table.
- Choose the lowest transportation cost out of the total transportation matrix and make an allocation of the minimum demand or supply.
- Then the column or row that has a fulfilled demand or supply must be removed and then a new matrix must be prepared.
- The above procedure must be repeated until the end of all allocations.

New methods developed out of LCM Method

Mehmood Shah et.al (2010) made a modification of the least cost method for resource allocation in the grid. The criteria for sharing resources are the grid. Any service quality is influenced by proper grid management. In their research, they developed a new method for resource allocation in the grid. That method was a modified least cost method, which prepares tasks for computing optimum nodes. That was an enhanced technique of the least cost method. They evaluated that method for the scenario of resource allocation in the grid. In addition, for grid scheduling, a linear programming model was proposed, which encapsulated a comparison of the proposed method with that of the available methods. Satisfactory results were obtained with their proposed technique.

Kavitha and Srinivasan (2018) proposed another new method in place of the least cost method for the optimization of the transportation problem. That method is otherwise known as the Lowest Even Cost method. The lowest even cost method was developed for finding the initial basic feasible solution for a given problem in transportation. Their method was found to be a complex method for calculation, when compared with the other methods. A better initial basic feasible solution was gained with the method, but the accuracy of the method was found to be unsure.

Vogel's Approximation Method (VAM)

Vogel's Approximation Method is one among the technique used for determining the first basic feasible solution for a given transportation problem. The steps for the optimization of a transportation problem using Vogel Approximation Method are given below.

- The technique initiates with the determination of the two minimum transportation cost cells in each column and row in the array of the transportation problem. The minimum of the transportation costs is subtracted from each other, which generates a Vogel number for each column and row.
- Choose the highest Vogel number and then assume in the respective minimum cost cell, where the assumption must be highest amount, which is transferred to the respective destination. Post this assignment, the recompilation of the Vogel numbers is made, contemplated on the existing columns and rows of the transportation array. The entire process is repeated till the entire shipments are done.

Computer based software is essential for the extra work of computation utilizing the Vogel approximation Method in offering transportation problems' solutions.

New methods developed out of VAM method

Pratihar et. al (2020) made a modification of the Vogel's approximation method for solving the transportation problems in an uncertain environment. In their research, the interval type 2 fuzzy set was embedded in the fuzzy set transportation problem for highlighting the supply, transportation cost and demand. The fuzzy sets and their types embedded in the transportation problems are discussed in this research. The Vogel's approximation method was modified to provide solution to the transportation problem of fuzzy set. Numerical examples were discussed for describing the system accuracy. Their research insists future analysis for application of their proposed technique in a practical transportation problem on a large scale.

Shimshak et. al (2016) developed an alternative form of Vogel's approximation method with the application of heuristics were utilized with the Vogel Approximation Method (VAM) in their proposed technique. Computations were made with test problems, which highlighted the accuracy of the heuristics. The method was found to consume less effort and time needed for obtaining an initial basic feasible solution utilizing Vogel approximation method. This result was found by both manual and computational operation.

Conclusion

Thus, this paper of exploration initially explores about the theoretical concepts associated with the transportation problems of operations research, then a disquisition review about the literature existing for the proposed research, the methods used to solve the transportation problems and the review on the recently developed advanced methods for optimizing the problems in transportation. Transportation problems were earlier resolved utilizing simplex method. But it was a time-consuming process and produced many errors. Artificial variables are ignored and main variables are only considered in a transportation problem. Thus, the above-mentioned linear programming models were developed for resolving both transportation and simplex problems. Transportation cost is an eminent parameter for optimization of problems in transportation. The new transportation methods are developed based on this objective.

References

1. Trustrum, Kathleen (1971) "The Transportation Problem." Linear Programming, edited by Kathleen Trustrum, Springer Netherlands, 1971, pp. 31–45. Springer Link, doi:10.1007/978-94-010-9462-7_3.

2. Wagener, Ulrich A. "A New Method of Solving the Transportation Problem." Journal of the Operational Research Society, vol. 16, no. 4, Dec. 1965, pp. 453–69. Springer Link, doi:10.1057/jors.1965.80.

3. "A New Heuristic Method of Finding the Initial Basic Feasible Solution to Solve the Transportation Problem." Journal of King Saud University - Computer and Information Sciences, July 2020. www.sciencedirect.com, doi:10.1016/j.jksuci.2020.07.007.

4. Ekanayake, E. M. U. S. B., et al. "A Modified Ant Colony Optimization Algorithm for Solving a Transportation Problem." Journal of Advances in Mathematics and Computer Science, Aug. 2020, pp. 83–101. journaljamcs.com, doi:10.9734/jamcs/2020/v35i530284.

5. BilqisAmaliaha, ChastineFatichaha, Erma Suryanib. A New Heuristic Method of Finding the Initial Basic Feasible Solution to Solve the Transportation Problem. July 2020, https://www.sciencedirect.com/science/article/pii/S131915782030421

6. "Essay on Transportation Problem| Operations Research | Linear Programming." Engineering Notes India, 21 Mar. 2017, https://www.engineeringenotes.com/project-management-2/operations-research/essay-on-transportation-problem-operations-research-linear-programming/15392.

7. "Importance of Transportation and Role of The Economic Development." IMPOFF, 4 Mar. 2020, https://impoff.com/importance-of-transportation/.

8. What Is Transportation Problem - Quantitative Techniques: Theory and Problems [Book]. https://www.oreilly.com/library/view/quantitative-techniques-theory/9789332512085/xhtml/ch5sec1.xhtml. Accessed 26 Apr. 2021.

9. "Transportation Problem | Set 1 (Introduction)." Geeks for Geeks, 25 Nov. 2019, https://www.geeksforgeeks.org/transportation-problem-set-1-introduction/.

10. "Linear Programming | Applications Of Linear Programming." Analytics Vidhya, 28 Feb. 2017, https://www.analyticsvidhya.com/blog/2017/02/lintroductory-guide-on-linear-programming-explained-in-simple-english/.

11. Metlo, M. A., et al. "Modified North-West Corner Method for the Optimal Solution for Transportation Problems." Sindh University Research Journal - SURJ (Science Series), vol. 48, no. 4, Dec. 2016. sujo-old.usindh.edu.pk, https://sujo-old.usindh.edu.pk/index.php/SURJ/article/view/2775.

12. Neetu M. Sharma and Ashok P. Bhadane. A New Method for Solving Fuzzy Transportation Problems. no. 12, Mar. 2016.

13. Shah, Syed Nasir Mehmood, et al. "Modified Least Cost Method for Grid Resource Allocation." 2010 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 2010, pp. 218–25. IEEE Xplore, doi:10.1109/CyberC.2010.47.

14. Kavitha, M., and N. Srinivasan. "Lowest Even Cost Method (OR) An Alter Method to Least Cost Method for the Transportation Problem." International Journal of Engineering and Management Research, vol. 8, no. 3, June 2018, pp. 109–11.

15. Pratihar, Jayanta, et al. “Modified Vogel’s Approximation Method for Transportation Problem under Uncertain Environment.” Complex & Intelligent Systems, vol. 7, no. 1, Feb. 2021, pp. 29–40. Springer Link, doi:10.1007/s40747-020-00153-4.

16. Shimshak, Daniel, et al. “A Modification Of Vogel’S Approximation Method Through The Use Of Heuristics.” INFOR: Information Systems and Operational Research, vol. 19, no. 3, Aug. 1981, pp. 259–63. Taylor and Francis+NEJM, doi:10.1080/03155986.1981.11731827.

Bulgarian travel operators market from insurance industry perspective

Pavel Todorov Stoynov
Technical University – Sofia,
Perun R&D Academy, Sofia, Bulgaria
pavel_todorov_stoynov@yahoo.com

Abstract.
During COVID pandemy, travel operators in Bulgaria met considerable difficulties and many of them received aid from Bulgarian Government. As a sequence, the price of the mandatory insurance for travel operators called "Responsibility of travel operators" raised and many insurers even denaide to sell this type of insurance policy.
Keywords: travel operators, insurance, policy "Responsibility of travel operator".

Introduction

During COVID pandemy, travel operators in Bulgaria met considerable difficulties and many of them received aid from Bulgarian Government. As a sequence, the price of the mandatory insurance for travel operators called "Responsibility of travel operators" raised and many insurers even denaide to sell this type of insurance policy.

Travel operators and isurers

Ministry of Tourism of Republic of Bulgaria maintains complete list of travel operators acting on the territory of Bulgaria as of 2022 (Ministry of Tourism of Republic of Bulgaria, 2022a). The list includes some important attributes like: type of the company (both travel agent and travel operator or only travel operator), registration ID of the travel operator, name of the company, company EIC number, date of expiration of the insurance policy, insurer, limit of the insurance policy.

The travel operator are insured by 11 insurers in 2022 as can be seen in Table 1. Actually, two insurers - LEV INS AD and ZD EUROINS AD – make 95% of all insurance policies "Insurance of travel operator".

Table 1. Insurers and number of Travel operators insured by them in 2022 г.

Insurer	Number of insured travel operators
GENERALLI AD	4
ZAD ALLIANCE BULGARIA AD	1
ZAD ARMEEEC AD	5
ZAD ASSET INSURANSE AD	5
ZAD OZK ZASTRAHOVANE AD	19
ZK AKSIOM AD	2
ZD BULINS AD	7
ZD EUROINS AD	558
ZEAD DAL BOG Live and health EAD	4
ZK LEVINS AD	700
ZA UNIQA AD	17

The insurance policies are for different limits, as can be seen in Table 2. The majority of insurances are for the lowest limit (25 000 BGN) which are 81% of all insurances.

Table 2. Limits of insurance and number of insurances of travel operators with this limit in 2022 г.

Limit, BGN	Number of insured travel operators
1,100,000.00	1
1,400,000.00	7

10,000.00	1
125,000.00	32
175,000.00	11
25,000.00	1071
250,000.00	17
30,000.00	2
350,000.00	8
450,000.00	5
50,000.00	83
500,000.00	16
65,000.00	1
700,000.00	3
75,000.00	63
850,000.00	1

There exists a relation between the insurance limit of the travel operator and the revenues of the travel operator declared for the previous year. This relation is defined in the **„Regulation of conditions for signing mandatory insurance covering the responsibility of the travel operator" (**Ministry of Tourism of Republic of Bulgaria, 2022b).

This also means that the risk of default is related to insurance limits as they are related to the income and hence to the risk exposure.

Only 7 travel operators have the maximal limit of 1 400 000 BGN for their insurance policies in 2022, i. e. have turnover more than 3 000 000 BGN: ALBENA TOUR EAD, ALEXANDRIA TRAVEL EOOD, BOHEMIA OOD, EMERALD TRAVEL BULGARIA OOD, ONEKS TOUR EOOD, SUNNY TOURS OOD and TEZ TOUR BULGARIA OOD.

The insurance limits, based on net income of the previous year, are presented in Table 3 below.

Table 3. Insurance limits defined by Ministry of tourism of Bulgaria

Insurance limit
25,000
30,000
50,000
65,000
75,000
125,000
175,000
250,000
350,000
450,000
500,000
700,000
850,000
1,100,000
1,400,000

Actually, large claims exceeding the the insurance limits are possible only for the large travel operators and hence a second level of insurance should be introduced for them. This is the intention of Ministry of Tourism. A Guarantee fund is planned to be created in Bulgaria following the experience of ither EU countries.

Conclusion

Large travel operators can meet considerable liquidity difficulties and even can default when catasrofic events like COVID pandemy occure. Because of this, a second level of insurance like Guarantee fund would be good practice and actually it is currently introduced in EU countries.

References

Ministry of Tourism of Republic of Bulgaria (2022a) List of travel operators acting on the territory of Bulgaria as of 2022. Work paper.

Ministry of Tourism of Republic of Bulgaria (2022b) Regulation of conditions for signing mandatory insurance covering the responsibility of the travel operator

Building a Model for Network Intrusion Detection with Artificial Neural Networks

Daniela Gotseva
dgoceva@tu-sofia.bg
Roumen Trifonov
r_trifonov@tu-sofia.bg
Pavel Stoynov
Technicla University – Sofia,
Faculty of Computer Science and Technology
Sofia, Bulgaria

Abstract.
In this article, application of Artificial Neural Networks (ANN) for building Intrusion Detection Systems (IDS) is presented. Data models are necessary and important when building IDS. In the article, the accent is put on feed-forward ANN and anomaly-based IDS where data models are needed to train and evaluate the ANN. Network-based IDS are considered and a statistical approach for building the ANN for intrusion detection is presented.

Key words: Artificial Neural Networks (ANN), IT security, Intrusion Detection Systems (IDS), Data Models, Host-based IDS, R language

Introduction

Feed-forward Artificial Neural Networks (FF ANN) are Artificial Neural networks (ANN) with no feedback loops. In general we may say that a network is feed-forward if it is possible to attach successive numbers to the inputs and to all of the hidden and output units such that each unit only receives connections from inputs or units having smaller numbers (Bishop, 1995).

Intrusion Detection Systems (IDS) are IT security systems for detecting hostile activity. When these systems also prevent the hostile activity, they are called Intrusion Detection and Prevention Systems (IDPS). They are host-based or network based (Gotseva, Trifonov and Stoynov, 2019).

The issue of data selection and preparation when working with neural networks is of high importance. It is crutial for building well adequate models and this motivates the researchers in the area to look for reliable data and for consistent ways for data preparation. The lack of public data sets is a challenges for anomaly-based intrusion detection (Sommer and Paxson, 2010; Haider et al., 2017) and the research community is working on this issue. A good review of the data sets for network intrusion detection is given by (Ring et al., 2019).

In this article, data models for network-based Intrusion Detection Systems (IDS) based on Feed-forward Artificial Neural Networks (FF ANN) are presented.

Data

Some data sets for intrusion detection are considered in Stoynov (2019a, 2019b).

Network-based data appear in packet-based or flow-based format. Flow-based data contain only meta information about network connections, packet-based data also contain payload (Ring et al., 2019).

It is possible to convert packet-based data to flow-based data (but not vice versa) with tools like nfdump 2 or YAF 3 (Haddadi and Zincir-Heywood, 2014).

The data set which we use to build the model is generated by Al-Kasassbeh, Al-Naymat, Al-Hawari (2018). The authors create a realistic dataset with 4998 records for anomaly detection. This dataset was actually based on SNMP variables collected from network devices in a real test network. The variables are grouped in 5 categories: Interface, IP, ICMP, TCP, UDP. The data set is labeled with 8 traffic categories: bruteForce, httpFlood, icmp-echo, normal, slowloris, slowpost, tcp-syn, udp-flood.

The statistics of the output variable – traffic category - is presented in Table 1.

All input variables are: ifInOctets11, ifOutOctets11, ifoutDiscards11, ifInUcastPkts11, ifInNUcastPkts11, ifInDiscards11, ifOutUcastPkts11, ifOutNUcastPkts11, tcpOutRsts, tcpInSegs, tcpOutSegs, tcpPassiveOpens, tcpRetransSegs, tcpCurrEstab, tcpEstabResets, tcpActiveOpens, udpInDatagrams, udpOutDatagrams, udpInErrors, udpNoPorts, ipInReceives, ipInDelivers, ipOutRequests, ipOutDiscards, ipInDiscards, ipForwDatagrams, ipOutNoRoutes, ipInAddrErrors, icmpInMsgs, icmpInDestUnreachs, icmpOutMsgs, icmpOutDestUnreachs, icmpInEchos, icmpOutEchoReps, class.

Table 1. Absolute frequency of Execution Path

Traffic category	Frequency
bruteForce	200
httpFlood	573
icmp-echo	632
normal	600
slowloris	780
slowpost	480
tcp-syn	960
udp-flood	773

To select the input variables which best explane the type of network traffic, construct Table 2, 3 and 4 where for every class of traffic are presented the average numbers of: (tcpInSegs) and output (tcpOutSegs) TCP segments, the input (udpInDatagrams) and output (udpOutDatagrams) UDP datagrams, input IP packets delivered (ipInDelivers), input IP packets received (ipInReceives), input IP packets discarded (ipInDiscards), output IP packets requested (ipOutRequests), output IP packets discarded (ipOutDiscards). The corresponding data are presented in the following tables.

Table 2. Average number of input and output TCP segments by Traffic category.

Traffic category	Average number of input TCP segments (tcpInSegs)	Average number of output TCP segments (tcpOutSegs)
bruteForce	103	101
httpFlood	333	277
icmp-echo	1 159	970
normal	873	694
slowloris	838	677
slowpost	766	629
tcp-syn	580	479
udp-flood	1 261	1 037
All categories	805	660

The data in the Table 2 show that considerable deviation from the normal behavior in terms of TCP segments is observed with bruteForce and httpFlood (negative deviation) and with icmp-echo and udpFlood (positive deviation).

Table 3. Average number of input and output UDP diagrams by Traffic category.

Traffic category	Average number of input UDP datagrams (udpInDatagrams)	Average number of output UDP datagrams (udpOutDatagrams)
bruteForce	180 992	119 937
httpFlood	120 934	88 695
icmp-echo	339 705	267 745
normal	199 647	148 378
slowloris	118 249	75 941
slowpost	206 948	152 275
tcp-syn	245 217	185 956
udp-flood	262 299	194 303
All categories	214 028	158 882

The data in the Table 3 show that considerable deviation from the normal behavior in terms of UDP datagrams is observed with httpFlood and slowloris (negative deviation) and with icmp-echo and udp-flood (positive deviation).

The data in the Table 4 show that considerable deviation from the normal behavior in terms of input IP packets delivered is observed with httpFlood and slowloris (negative deviation) and with icmp-echo and udp-flood (positive deviation).

In terms of input IP packets received, deviation from the normal behavior is shown by bruteForce, icmp-echo, slowpost, tcp-syn and udp-flood (in positive direction).

Table 4. Average number of input and output IP packets (delivered, received, requested, discarded) by Traffic category.

Traffic category	Average number of input IP packets delivered - ipInDelivers	Average number of input IP packets received - ipInReceives	Average number of input IP packets discarded - ipInDiscards	Average number of output IP packets requested - ipOutRequests	Average number of output IP packets discarded - ipOutDiscards
bruteForce	186 708	273 848 108	19	120 052	852
httpFlood	123 427	66 331 387	24	88 996	58
icmp-echo	346 072	166 195 750	22	268 651	505
normal	203 607	46 836 956	30	149 155	383
slowloris	122 119	39 674 948	30	76 671	109
slowpost	212 981	160 545 656	39	152 967	79
tcp-syn	250 301	149 355 523	44	186 479	128
udp-flood	268 131	148 117 047	31	195 340	279
All	218 884	118 407 273	32	159 570	243

Table 5. Ranges defined for input TCP segments (variable tcpInSegs).

Average number of input TCP segments (tcpInSegs)	Code
0-450	1
450-900	2
900-1350	3
>1350	4

From these variables we select as input variables for the ANN the variables: tcpInSegs, udpInDatagrams, ipInReceives, ipOutRequests. To use variables we transform them in ranges.

Table 6. Average number of input TCP segments and their ranges by Traffic category.

Traffic category	Average number of input TCP segments (tcpInSegs)	Range
bruteForce	103	1
httpFlood	333	1
icmp-echo	1 159	3
normal	873	2
slowloris	838	2
slowpost	766	2
tcp-syn	580	2
udp-flood	1 261	3
All categories	805	2

Table 5 shows the transform of the input variable tcpInSegs into ranges.

In Table 6, the defined ranges from Table 5 are mapped to the average number of the input TCP segments for every traffic category. The normal range is 2. Negative deviation (range 1) have bruteForce and httpFlood categories. Positive deviation (range 3) have icmp-echo and udp-flood traffic categories.

Table 7 shows the transform of the input variable udpInDatagrams into ranges.

Table 7. Ranges defined for input UDP datagrams (variable udpInDatagrams).

Average number of input UDP datagrams (udpInDatagrams)	Code
0-125 000	1
125 000-250 000	2
250 000-375 000	3
>375 000	4

In Table 8, the defined ranges from Table 7 are mapped to the average number of input UDP datagrams for every traffic category. The normal range is 2. Negative deviation (range 1) have httpFlood and slowloris categories. Positive deviation (range 3) have icmp-echo and udp-flood traffic categories.

Table 8. Average number of input UDP datagrams and their ranges by Traffic category.

Traffic category	Average number of input UDP datagrams (udpInDatagrams)	Range
bruteForce	180 992	2
httpFlood	120 934	1
icmp-echo	339 705	3
normal	199 647	2
slowloris	118 249	1
slowpost	206 948	2
tcp-syn	245 217	2
udp-flood	262 299	3
All categories	214 028	2

Table 9 shows the transform of the input variable ipInReceives into ranges.

Table 9. Ranges defined for input IP packets received (variable ipInReceives).

Average number of input IP packets received -ipInReceives	Code
0-100 000 000	1
100 000 000-200 000 000	2
200 000 000-300 000 000	3
>300 000 000	4

In Table 10, the defined ranges from Table 9 are mapped to the average number of input IP packets received for every traffic category. The normal range is 1. Moderate positive deviations (range 2) have icmp-echo, slowpost, tcp-syn and udp-flood traffic categories. Large positive deviation (range 3) has bruteForce.

Table 10. Average number of input IP packets received and their ranges by Traffic category.

Traffic category	Average number of input IP packets received - ipInReceives	Range
bruteForce	273 848 108	3
httpFlood	66 331 387	1
icmp-echo	166 195 750	2
normal	46 836 956	1
slowloris	39 674 948	1
slowpost	160 545 656	2
tcp-syn	149 355 523	2
udp-flood	148 117 047	2
All categories	118 407 273	2

Table 11. Ranges defined for output IP packets requested (variable ipOutRequests).

Average number of output IP packets requested -ipOutRequests	Code
0-100 000	1
100 000-200 000	2
200 000-300 000	3
>300 000	4

Table 12. Output IP packets requested and their ranges by Traffic category.

Traffic category	Average number of output IP packets requested -ipOutRequests	Range
bruteForce	120 052	2
httpFlood	88 996	1
icmp-echo	268 651	3
normal	149 155	2
slowloris	76 671	1
slowpost	152 967	2
tcp-syn	186 479	2
udp-flood	195 340	2
All categories	159 570	2

Table 11 shows the transform of the output IP packets requested into ranges.

In Table 12, the defined ranges from Table 11 are mapped to the average number of output IP packets requested for every traffic category. The normal range is 2. Negative deviations (range 1) have httpFlood and slowloris traffic categories. Positive deviations (range 3) has icmp-echo traffic category.

As output variables we select traffic category. The encoding of this variable is given in Table 13.

Table 13. Encoding of the output variable Traffic category.

Traffic category	Code
bruteForce	1
httpFlood	2
icmp-echo	3
normal	4
slowloris	5
slowpost	6
tcp-syn	7
udp-flood	8

The final data model which we construct is a Feed-forward ANN. It has 16 binary input variables presented in Table 14 and 8 output variables y_1 to y_8 which are correspondingly the different traffic categories presented in Table 13.

Table 14. Input variables.

Variable	Observed variable	Range
x_1, x_2, x_3, x_4	input TCP segments	1,2,3,4,
x_5, x_6, x_7, x_8	input UDP datagrams	1,2,3,4
$x_9, x_{10}, x_{11}, x_{12}$	input IP packets received	1,2,3,4
$x_{13}, x_{14}, x_{15}, x_{16}$	output IP packets requested	1,2,3,4

Every input and output variable takes values 0 or 1. An input variable takes the value 1 when the corresponding observed variable is in the corresponding range. An output variables takes value 1 when the corresponding traffic category is observed.

The Feed-forward ANN has a one hidden layer with 4 hidden neurons. The activation function of input variables and output variables are linear. The activation function of the hidden neurons is logistic.

The next steps are: to train the model, to apply the model for intrusion prediction and to evaluate the model.

Conclusion

When creating data models for intrusion detection with neural network it is important to select the most suitable input variables – the ones which best describe the type of the network traffic. In the article we present a statistical approach for defining suitable input and output variables and to specify completely the neural network to use in intrusion detection.

References

Al-Kasassbeh, Al-Naymat, Al-Hawari (2018) Using machine learning methodsfor detecting network anomalies within SNMP-MIB. Intenational Journal of Wireless and Mobile Computing, vol. 15, No. 1.

Bishop, M. (1995) Neural Networks for Pattern Recognittion. Oxford University Press. ISBN 0 19 853864 2 (Pbk)

Gotseva, D., R. Trifonov and P. Stoynov (2019) Neural Networks for Intrusion Detection.

Haddadi, F., A. Zincir-Heywood (2014) Benchmarking the Effect of Flow Exporters and Protocol Filters on Botnet Traffic Classification, IEEE Systems Journal 10 (4) (2016) 1390–1401. doi:10.1109/JSYST. 2014.2364743.

Ring, M., S. Wunderlich, D. Scheuring, D. Landes and A. Hotho (2019) A Survey of Network-based Intrusion Detection Data Sets. arXiv preprint: 1903:02460v2[cs.CR] 6 Jul 2019.

Sommer, R., V. Paxson (2010), Outside the Closed World: On Using Machine Learning For Network Intrusion Detection, in: IEEE Symposium on Security and Privacy, IEEE, 2010, pp. 305–316. doi:10.1109/SP.2010.25.

Stoynov, P. (2019a) Databases of anomalous traffic for estimation of Intrusion detection. Financial and Actuarial Mathematics-FAM-2019, PERUN-SPRINT, Sofia, Bulgaria, 2019, 3-8. ISSN 1314-460X.

Stoynov, P. (2019b) Evaluating Network Intrusion Detection Systems. - "Financial and Actuarial Mathematics-FAM-2019", PERUN-SPRINT., Sofia, Bulgaria, 2019, 9-16. ISSN 1314-460X.