



Available online at www.sciencedirect.com



Procedia Computer Science 121 (2017) 152-159

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN -International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist 2017, 8-10 November 2017, Barcelona, Spain

# Modeling Neutrosophic Data by Self-Organizing Feature Map: MANETs Data Case Study

Haitham ELwahsh<sup>a</sup>\* , Mona Gamal<sup>b</sup>, A. A. Salama<sup>c</sup>, I.M. El-Henawy<sup>d</sup>

<sup>a</sup> Computer Science Department, Faculty of Computers and Information, Kafrelsheikh University, Kafrelsheikh 33516, Egypt
 <sup>b</sup>Information System Department, Faculty of Computers and Information, Kafrelsheikh University, Kafrelsheikh 33516, Egypt
 <sup>c</sup>Department of Mathematics and Computer Science, Faculty of Sciences, Port Said University, Port Said 522, Egypt.
 <sup>d</sup>Computer Science Department, Faculty of Computers and Information, Zagazig University, Zagazig, Egypt

# Abstract

Network security is a major research area for both scientists and business. Intrusion Detection System (IDS) is one of the most challenging problems in Mobile Ad Hoc Networks (MANETs). The main reason resides behind the changing and uncertain nature of MANETs networks. Hence, a compensate evolving in the IDS would be converting the whole system to rely on uncertainty and indeterminacy concepts. These concepts are the main issues in the fuzzy system and consequently in neutrosophic system. In neutrosophic system, each attack is determined by MEMEBERSHIP, INDETERMINACY and NONMEMEBERSHIP degrees. The main obstacle is that most data available are regular values which are not suitable for neutrosophic calculation. This paper is concerned by the preprocessing phase of the neutrosophic knowledge discovery system. Converting the regular data to neutrosophic values is a problem of generating the MEMEBERSHIP, NONMEMEBERSHIP and INDETERMINACY functions for each variable in the system. Self-Organized Feature Maps (SOFM) are unsupervised artificial neural networks that were used to build fuzzy MEMEBERSHIP function, hence they could be utilized to define the neutrosophic variable as well. SOFMs capabilities to cluster inputs using self-adoption techniques have been utilized in generating neutrosophic functions for the subsets of the variables. The SOFM are used to define the MEMEBERSHIP, NONMEMEBERSHIP and INDETERMINACY functions of the KDD network attacks data available in the UCI machine learning repository for further processing in knowledge discovery. Experimental results show the features and their corresponding functions.

# © 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies.

\* Corresponding author. Tel.: +2-01114488334; fax: +0-000-000-0000 . E-mail address:Haitham.elwahsh@gmail.com

1877-0509 © 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies. 10.1016/j.procs.2017.11.021 Keywords: MANETs; Neutrosophic system; Self-Organized Feature Map; Intrusion Detection System (IDS).

#### **1. INTRODUCTION**

Mobile Ad hoc Network (MANET) is a system of wireless mobile nodes that dynamically self-organized in arbitrary and temporary network topologies without communication infrastructure. This network may change quickly and unforeseeable. The unique characteristics of MANET give an adversary the opportunity to launch numerous attacks against ad-hoc networks [1].

Security in Mobile Ad Hoc Network is the most important concern for the basic functionality of network [15]. The Security issues have been met when availability of network services, confidentiality and integrity of the data can be achieved by assuring that. MANET often suffer from security attacks because of its features like open medium, changing its topology dynamically, lack of central monitoring and management, cooperative algorithms and no clear defense mechanism. These factors have changed the battle field situation for the MANET against the security threats [16, 18]. Classifying these threats is a challenging procedure that suffers from uncertainty and indeterminacy problems. Fuzzy [19] and neutrosophic [3] disciplines provide a logical base for dealing with uncertainty problems.

The concept of fuzzy sets was introduced by LotFI A. Zadeh in 1965 [20]. Since then the fuzzy sets and fuzzy logic have been applied in many real life problems in uncertain and ambiguous environments. The traditional fuzzy sets are characterized by the MEMEBERSHIP value or the grade of MEMEBERSHIP value. Sometimes it may be very difficult to assign the MEMEBERSHIP value for fuzzy sets. Consequently the concept of interval valued fuzzy sets was proposed [17] to capture the uncertainty grade of MEMEBERSHIP value. In almost realistic life problems in expert system, credence system, information coalition and so on, we must consider the truth-MEMEBERSHIP as well as the falsification-MEMEBERSHIP for convenient description of an object in unreliable, obscure environment. Neither the fuzzy sets nor the interval valued fuzzy sets is appropriate for such a situation. Intuitionistic fuzzy sets introduced by Atanassov [21] are appropriate for such a situation. The intuitionistic fuzzy sets can only handle the incomplete information considering both the truth-MEMEBERSHIP (simply MEMEBERSHIP) and falsity-MEMEBERSHIP (NON-MEMEBERSHIP) values. It does not handle the indefinite and discrepant information which exists in credence system.

Smarandache [22] introduced the concept of neutrosophic set which is a mathematical tool for handling problems involving imprecise, indeterminacy and inconsistent data. Neutrosophic system is a generalization for fuzzy system. In neutrosophic set, indeterminacy is quantified explicitly. MoreOver, truth-MEMEBERSHIP, indeterminacy MEMEBERSHIP and falsification-MEMEBERSHIP are independent. In [7] Salama introduced the concept of neutrosophic crisp set Theory, to represent any event by a triple crisp structure. Moreover, the work of Salama et al. [5-9] formed a starting point to construct new branches of neutrosophic mathematics and computer sci. Hence, Neutrosophic set theory turned out to be a generalization of both the classical and fuzzy counterparts. This assumption is very important in a lot of situations such as information fusion when we try to combine the data from different sensors. Neutrosophic set is a potential general conventional framework which propagate the concept of the classic set, fuzzy set [20], interval valued fuzzy set intuitionistic fuzzy set [3], etc. A neutrosophic set A defined on universe U.  $x = x(T, I, F) \in A$  with T, I and F being the real standard or nonstandard subsets of  $]0^{-}, 1^{+}[.T]$  is the degree of truth-MEMEBERSHIP function in the set A, I is the INDETERMINACY function in the set A and F is the falsity-MEMEBERSHIP function in the set A.Building a neutrosophic IDS is a feasible solution in dealing with ambiguity circumstances. The neutrosophic IDS is composed of two main sub modules: the preprocessing phase and the network attacks classification phase. The preprocessing phase concentrates on preparing the network data in a format suitable for the classification module. This paper is concerned in reformatting the regular data in the KDD data set [23] into neutrosophic format  $(\mathbf{x}, \boldsymbol{\mu}_A(\mathbf{x}), \boldsymbol{\sigma}_A(\mathbf{x}), \boldsymbol{\nu}_A(\mathbf{x}))$  where x is the value of attribute data,  $\boldsymbol{\mu}_A(\mathbf{x})$  is the MEMEBERSHIP value,  $\sigma_A(\mathbf{x})$  is the INDETERMINACY value and  $v_A(\mathbf{x})$  is the NON-MEMEBERSHIP MEMEBERSHIP value of the x in the data space. Although, manual procedures of human interfering could help in reformatting the neutrosophic

network data, the huge amount of data would be an obstacle for this kind of help. Semi manual techniques could be utilized here. Human expert could help in preparing a subset of common values in the network data. Then, a machine learning technique could be utilized to learn from this subset and complete the process of reformatting. The rest of this paper is organized as follows: section 2 presents the theories and overviews. Section 3 proposes the SOFM to learn the neutrosophic functions  $\mu_A(\mathbf{x})$ ,  $\sigma_A(\mathbf{x})$ ,  $\mathbf{v}_A(\mathbf{x})$ . Section 4 and 5 show the experimental results and conclusion respectively.

## 2. THEORIES AND OVERVIEW

#### 2.1 Neutrosophic System

Neutrosophy has set the establishment for a whole family of new mathematical theories propagating both their classical and fuzzy counterparts [2, 3, 4] such as a neutrosophic set theory. We mention some relevant basic preliminaries, and in particular, the work of Smarandache in [5, 6, 7] and Salama et al. [8,9]. Smarandache introduced the neutrosophic components **T**, **I**, **F** which represent the MEMEBERSHIP, indeterminacy, and NON-MEMEBERSHIP values respectively, where  $]0^-, 1^+[$  is nonstandard unit interval. Salama et al. introduced the following: Let X is a non-empty fixed set. A neutrosophic set is an object having the form  $A = (x, \mu_A(x), \sigma_A(x), \nu_A(x))$  where  $\mu_A(x), \sigma_A(x)$ , and  $\nu_A(x)$  which represent the degree of MEMEBERSHIP function (namely), the degree of indeterminacy (namely), and the degree of NON-MEMEBERSHIP (namely  $\mu_A(x)$  the degree of indeterminacy namely ( $\sigma_A(x)$ ) and the degree of non-member ship (namely  $\nu_A(x)$ ) respectively of each element  $x \in X$  to the set A where

$$\mathbf{0}^{-} \leq \left( \boldsymbol{\mu}_{A}(\mathbf{x}), \boldsymbol{\sigma}_{A}(\mathbf{x}), \boldsymbol{\nu}_{A}(\mathbf{x}) \right) \leq \mathbf{1}^{+}$$
 and

$$0^{-} \leq (\mu_A(\mathbf{x}) + \sigma_A(\mathbf{x}) + \nu_A(\mathbf{x})) \leq 3$$

Smarandache introduced the following: Let T, I,F be real standard or nonstandard subsets of ]  $0^-$ ,  $1^+$ [, with Sup\_T=t\_sup, inf\_T=t\_inf; Sup\_I=i\_sup, inf\_I=i\_inf; Sup\_F=f\_sup, inf\_F=f\_inf; n-sup=t\_sup+i\_sup+f\_sup n-inf=t\_inf+i\_inf+f\_inf; T, I, F are called neutrosophic components.

(1)(2)

## 2.2 Kohonen's self-organizing Feature Map (SOFM)

Kohonen Self-Organizing Maps (or just Self-Organizing Maps, or SOMs for short), are a type of neural network [10, 12]. They were developed in 1982 by Tuevo Kohonen, a professor emeritus of the Academy of Finland. Self-Organizing Maps are aptly named. "Self-Organizing" is because no supervision is required. SOFMs learn on their own through unsupervised competitive learning. "Maps" is because they try to chart their weights to stratify to the given input data [13]. The nodes in a SOM network attempt to become like the inputs presented to them. Assuming an input vector of  $\mathbf{z_n} = [\mathbf{x_n y_n}]$ T, the SOFM will go through two main phases. The first is the training phase which accept the input data and apply the calculations in equations 2,3 and 4 to find the winning neuron and update the network connections. The second is the retrieving phase which retrieves the weight connections associated with the winning neuron for each new instance of input data.

$$\mathbf{q}(\mathbf{x}_{\mathbf{n}}) = \min_{\substack{\substack{i \\ j \neq j \\ i \neq k}}} \|\mathbf{x}_{\mathbf{n}} - \mathbf{w}_{j}\|; \tag{3}$$

$$\eta_{qj}[t] = \begin{cases} \mu_L(t) \in N_q \\ 0 j \notin N_q \end{cases}$$
(4)

$$w_{j}[t+1] = w_{j}[t] + \eta_{qj}[t] (x_{n}[t] - w_{j}[t])$$
(5)

With, in this case, the weight vector  $\mathbf{w}_{j} = [\mathbf{w}_{j1}\mathbf{w}_{j2} \dots \mathbf{w}_{jd}\mathbf{x}_{j(d+c)}]^{T} = [\mathbf{w}_{jd}\mathbf{w}_{jc}], j = 1, \dots, J$ . After the learning phase, the SOFM can be considered as a MEMEBERSHIP generation network just like its counterpart, the feed forward multilayer neural network trained with a supervised learning algorithm [14]. However, in the retrieving phase, it is not as straightforward as in the case of the feed forward multilayer neural network and some modification, described next, is required. In the retrieving phase, the input feature vector is only $\mathbf{x}_n$ . Therefore, the input feature vector will find the best matching neuron q by considering only the weight sub vector  $\mathbf{w}_{jd} = [\mathbf{w}_{j1} \dots \mathbf{w}_{jd}]^{T}$  related to input features, which is Eq. 2. After finding the winning neuron q, the output of SOFM is the weight sub

vector  $\mathbf{w}_{qc} = \begin{bmatrix} \mathbf{w}_{q(d+1)} & \dots & \dots & \mathbf{w}_{q(d+c)} \end{bmatrix}^T$  associated with the labeling information. Also, it is the neutrosophic MEMEBERSHIP generated by SOFM.

#### 3. DESIGNING SOFM for Modeling Neutrosophic Variables

The SOFM used to identify the MEMEBERSHIP, NONMEMEBERSHIP and INDETERMINACY functions for the network features. The neurons in the SOM emulate the inputs applied to them to achieve the learning process. The topological relationships between input data are conserved when mapped to a SOFM network. This is a very important capability when inspect complex data. Fuzzy system [11] depends on the fuzzy variable to build the fuzzy rules and equations. SOFM was utilized in generating fuzzy MEMEBERSHIP function [21]. Because neutrosophic system depends partially on MEMEBERSHIP function, SOFM could be used to define neutrosophic MEMEBERSHIP function for the variables. Furthermore, SOM could be utilized to define both NONMEMEBERSHIP and INDETERMINACY functions to complete the neutrosophic variable modeling. The generation of neutrosophic MEMEBERSHIP function is generated according to the clusters in the first step. However, it is possible to integrate the two-step procedure and generate the neutrosophic MEMEBERSHIP function directly during the learning phase. The main idea is to augment the input feature vector with the class labeling information.

The focus of this research is on how SOFM could be used to handle neutrosophic information. Therefore, the information being associated is all neutrosophic variables. For the neutrosophic MEMEBERSHIP function, a key step in the proposed technique is to combine the input feature vector  $\mathbf{x}_n = [\mathbf{x}_{n1}\mathbf{x}_{n2} \dots \dots \mathbf{x}_{nd}]^T$  with the vector  $\mathbf{y}_n = [y_{n1}y_{n2} \dots \dots y_{nc}]^T$  Coding the subset labeling information of truth degrees. The dimensions of  $\mathbf{x}_n$  and  $\mathbf{y}_n$  are respectively, the number of input features d and the number of subset labels c. That is, a new vector  $\mathbf{z}_n$  of dimension c + d is constructed according to  $\mathbf{z}_n = [\mathbf{x}_n \mathbf{y}_n]^T = [\mathbf{x}_n \ 0]^T + [0 \ \mathbf{y}_n]^T$ . In the learning phase, the newly constructed  $\mathbf{z}_n$  will be the Input feature vector to SOFM. The weight updating as indicated by Eq.5. The procedure is the same for the neutrosophic NONMEMEBERSHIP function. Instead of the degree of truth, the  $\mathbf{z}_n$  vector will hold the degree of falsity for different labels of the variable values.

According to the neutrosophic set definition, the MEMEBERSHIP, NONMEMEBERSHIP and INDETERMINACY functions are independent but with one condition which is provided in Eq. 2. The summation of the three values for a neutrosophic label should not exceed 3. Hence, the indeterminacy function could be defined by knowing the MEMEBERSHIP and NONMEMEBERSHIP function as follows.

$$0^{-} - (\mu_{A}(x) + \sigma_{A}(x)) \le \nu_{A}(x) \le 3^{+} - (\mu_{A}(x) + \sigma_{A}(x))$$
(6)

Then the resulted INDETERMINACY function should be normalized to fit in the [0, 1] interval according to the second condition in Eq. 1. The algorithm of the proposed technique is illustrated as follows.



# 4. Experimental Results

The KDD-99 dataset's [23] connections are represented by 41 features; the features in Columns 2, 3, and 4 are the protocol type, the service type, and the flag. The value of the protocol type may be TCP, UDP, or ICMP; the service type could be one of the 65 different network services such as HTTP and PRIVATE; and the flag has 9 possible values such as SF or S0. After reducing KDD-99 features from each record, pre-processing will be done by reverse each feature from text or symbolic into numerical form. So for each text or symbolic an Integer code is assigned. As follow: Table 1. Numeric Values of KDD Features

PROTOCOL TYPE	FEATURE VALUE	SERVICE	FEATUREVALUE	FLAG	FEATUREVALUE
ТСР	1	HTTP	1	SF	1
UDP	2	PRIVATE	2	S0	2
ICMP	3	FTP_DATA	3	REJ	3
		SMTP	4	RSTR	4
		POP_2	65	S2	9

# 4.1. SOFM for Modeling neutrosophic variable

The proposed technique is concerned by the pre-processing phase of the neutrosophic knowledge discovery system. Self-Organized Feature Maps (SOFM) are unsupervised artificial neural networks that were used to build fuzzy MEMEBERSHIP function, hence they could be utilized to define the neutrosophic variable as well. SOFMs capabilities to cluster inputs using self-adoption techniques have been utilized in generating neutrosophic functions for the subsets of the variables. The SOFM are used to define the MEMEBERSHIP, NONMEMEBERSHIP and INDETERMINACY functions of the KDD network attacks data available in the UCI machine learning repository for further processing in knowledge discovery. Our experimental Results Shows the features and their corresponding functions.

# SOM parameters

•	Dimensions :
	Number of input dimensions: 3 inputs [ value, Normal, Up-normal]
	Outputs :Neuron: 225 neuron

• Error rate threshold: 0.15

Simulation

System features:

- Processor: Intel(R) Core (TM) i3-3227U CPU @ 1.90GHZ 1.90 GHZ
- Memory (RAM): 4.00 GB (3.87 GB usable)
- System type: 64-bit operating system, x64- based processor
  - Windows edition: windows 8.1

SOFM program simulation is implemented by C#. Each feature from the KDD data will have three different files for MEMEBERSHIP, NONMEMEBERSHIP assumption values. These files are provided to the SOFM program to build the entire functions. The generated output file contains the values of elected neuron. The INDETERMINACY function is calculated from the MEMEBERSHIP and NONMEMEBERSHIP data according to Eq. 6. The INDETERMINACY function is further normalized to fit within the interval  $]0^-$ ,  $1^+[...$  The results graphs shown below figure 1, which indicate the relation of MEMEBERSHIP, NON-MEMEBERSHIP and INDETERMINACY FUNCTIONS. The graphical representation shows of the duration, dst\_bytes, hot, count and Srv\_serror\_rate variables from the KDD data set.



Fig.1 the relation of membership, non-membership and neutrosophic membership

The first figure for each variable represent the MEMBERSHIP function while the second figure represent the NON-MEMBERSHIP function. It is clear that the two figures are the complement of each other. The third figure is the INDTERMINACY function that fills the unknown gap left by the MEMBERSHIP and NON-MEMBERSHIP functions

## 5. Conclusion and future work

The MANET network suffers from uncertainty and ambiguity due to its nature of self-organizing in arbitrary and temporary network topologies without communication infrastructure. Hence, developing a system to handle this uncertain situation is highly required. Neutrosophic system, like fuzzy system, deals with uncertainty basing on truth and falsity MEMBERSHIP degrees. In addition, the concept of INDETERMINACY is defined for all variables in neutrosophic systems. This research is a preprocessing phase in an intelligent system for detecting threats in MANET networks. The main issue in the preprocessing phase is concerting the regular data found in the KDD data from UCI machine learning repository into neutrosophic data. The resulted data should be the value of the variable combined with a triple of MEMBERSHIP, NON-MEMBERSHIP and INDTERMINACY degrees. The SOFM is utilized to create the MEMBERSHIP, NON-MEMBERSHIP functions starting from small subset of the data recorded from a human expert. The INDTERMINACY function is calculated via the definitions of the neutrosophic set. The experiments show the graphical results of the functions for the duration, dst\_bytes, hot, count and Srv\_serror\_rate variables from the KDD data set. The next phase is to build a classification pattern for the neutrosophic variables to detect threats in the MANET network. This phase can be implemented by an artificial intelligent algorithm.

## References

- 1.Raja, M.L. and Baboo, C.D.S.S. (2014), "An Overview of MANET: Applications, Attacks and Challenge".
- K. Atanassov, neutrosophicsets, in V.Sgurev, ed., Vii ITKRS Session, Sofia(June 1983 central Sci. and Techn. Library, Bulg. Academy of Sciences 1984.
- 3. K. Atanassov, neutrosophicsets, Fuzzy Sets and Systems 2087-96. 1986.
- 4. L.A. Zadeh, Fuzzy Sets, Inform and Control 8, 338-353.1965
- 5. A. A. Salama, Mohamed Eisa, Hewayda ElGhawalby and A. E. Fawzy: Neutrosophic Features for Image Retrieval, Neutrosophic Sets and Systems. Vol. 13, 2016, pp 56-61.
- A. A. Salama, Mohamed Eisa, S. A. ELhafeez and M. M. Lotfy, Review of Recommender Systems Algorithms Utilized in Social Networks based e-Learning Systems & Neutrosophic System, Neutrosophic Sets and Systems, 2015, Vol.(8)pp35-44.
- 7. A. A. Salama, F.Smarandache," Neutrosophic Crisp Set Theory "Educational Publisher, Columbus, 2015.
- A. A. Salama, Mohamed Abdelfattah, H. A. El-Ghareeb, A. M. Manie, Design and Implementation of Neutrosophic Data Operations Using Object Oriented Programming, International Journal of Computer Application Issue 4, Volume 5 (Sep - Oct 2014)pp.163-175.
- A. A. Salama, Florentin Smarandache. Neutrosophic crisp probability theory & decision making process, Critical Review. Volume XII, 2016, pp 34-48, 2016.
- 10.T. Kohonen, Self-Organizing Maps. 3rd Ed., Springer-Verlag, ISBN 3-540-67921-9, 2000
- Horikawa, S. 1997. Fuzzy classification system using self-organizing feature map.Oki Tech. Rev. 63(159), [Online]. Available from:
   http://www.oki.com/en/otr/html/nf/otr-159-05.html>.
- 12. Kohonen, T., 1990. The self-organizing map. Proc. IEEE 78 (9), 1464-1480.
- 13. Mitra, S., Pal, S.K., 1994. Self-organizing neural network as a fuzzy classifier. IEEE Trans. Systems Man Cybernet. 24 (3), 385-399.
- 14. Takagi, H., Hayashi, I., 1991. NN-driven fuzzy reasoning.Internat.J.Approx.Reason. 5, 191-212.
- S.Sarika, A.Pravin, A.Vijayakumar, K.Selvamanid "Security Issues in Mobile Ad Hoc Networks" Procedia Computer Science, Volume 92, Pages 1-594 (2016.)
- Kalpana Sharma, M K Ghose "Wireless Sensor Networks: An Overview on its Security Threats" IJCA Special Issue on "Mobile Ad-hoc Networks" MANETs, 2010
- 17. Chih-Chung Yang, and N.K. Bose, *Generating fuzzy membership function with self-organizing feature mapPattern Recognition Letters*, Volume 27, Issue 5, 1 April 2006, Pages 356–365
- Haitham Elwahsh, Mohamed Hashem, Mohamed Amin, "Secure Service Discovery Protocols for Ad Hoc Networks", Springer (LNCS) in Computer Science, Advances in Computer Science and Information Technology Communications in Computer and Information Science, 2011, Volume 131, Part 1, 147-157, DOI: 10.1007/978-3-642-17857-3\_15
- Bilgic T. and Tiirksen LB. (1999). Measurement of membership functions: Theoretical and empirical work, Fundamentals of Fuzzy Sets (Dubois D. and Prade H. eds.), Kluwer Acad. Publ., 1999.
- 20. L. A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338-353.
- 21. I. Turksen, Interval valued fuzzy sets based on normal forms, Fuzzy Sets and Systems 20 (1986) 191-210.
- 22. F. Smarandache, Neutrosophic set, a generalisation of the intuitionistic fuzzy sets, Inter. J. Pure Appl. Math. 24 (2005) 287 (297.
- 23. KDD Cup 1999 Data, https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html