I J C T A, 10(9), 2017, pp. 923-929 ISSN: 0974-5572

Implementation of Classification Techniques to Detect Congestion in Wireless Sensor Networks

Jayashri B. Madalgi* **and S. Anupama Kumar****

ABSTRACT

wireless Sensor Network -WSN is network of hundreds or thousands of sensors. Congestion occurs in wireless sensor networks when an event occurs. Congestion leads to performance degradation of a system. The data mining techniques help to detect congestion and then it can be mitigated by adjusting transmission rate. In this paper we discuss the implementation of data mining classification technique to detect the different congestion levels low, medium or high in WSN. We have implemented Non-Nested Generalized Exemplars (NNGE) and Best-First Decision Tree (BFTREE)classification algorithms to detect the congestion over the network. For the given data set, it is found that NNGE is more accurate than BFTREE in detecting the congestion as accuracy of NNGE is 100% and accuracy of BFTREE is 92.8572 %.

Keywords: WSN, Data mining, Congestion control, classification, NNGE, BFTREE.

1. INTRODUCTION

Sensor networks has many applications in most of the domains like habitat monitoring, object tracking or identification, monitoring of environment , military, disaster management etc,[9]. With the growth of networks and increased link speeds *congestion* in networks is a significant. Congestion occurs when the packet traffic load in the network is greater than the capacity of the network. It is a situation in which too many packets are present in a part of the subnet, thus leading to performance degradation.[2]

1.1. Congestion in Sensor Networks

The traffic in WSN can be divided into 2 streams. The downstream is stream from the sink or base station to the sensors and the upstream is from sensors to the sink or base station [12]. The downstream traffic is of one-to-many where as upstream is many-to-one nature.[2]

1.2. Nature of congestion

In WSN applications congestion can be differentiated into 2 categories: the congestion near the sink and congestion which arises near the sources. [14]

1.2.1. Congestion near the event-sources

Occurrence of an event results in traffic burst generated from sensor nodes near the event area, leading to collisions and loss in packets at the source nodes.

* Gogte Institute of technology, VTU Research center, *Email: jayashri@git.edu*

** RV College of Engineering, Bengluru, India, *Email: Kumaranu050s6@gmail.com*

1.2.2. Congestion near the sink

The traffic generated at multiple source nodes proceeds or travels in multi hop fashion towards the base station, The data is generated when an event is detected at the same a time by the source nodes .It results in more packet flow in the region near base station due to funnel-like communication path structure. This increase in packet flow as traffic results in congestion.

2. BACK GROUND WORK

Machine learning techniques [14] are applied to congestion control over wired and wireless networks. The authors of [4], [5] and [7] have analyzed the status of wired and wireless network using the parameters like routing algorithm used ,traffic load, queue length, bandwidth, route available and predict possibility of congestion and provide alternate route for congestion control [5]. In our paper[27],We have applied j48,RIPPER[23] and PART [22] classifiers using different WSN data set available. Many techniques exists which are specifically invented for the wireless sensor networks. These protocols are deployed by the layers of the WSN OSI stack. [2]

2.1. Techniques used by Data Link Layer

In wireless sensor networks mainly there are two congestion types-channel collision and buffer congestion. This layer overcomes channel collision by using the following mechanisms CSMA(Carrier Sense Multiple Access), FDMA(Frequency Division Multiple Access), and TDMA(Time Division Multiple Access).

2.2. Techniques used by Network Layer Techniques

BOBRED- Beacon Order Based Random Early Detection-(AQM) i.e active queue management techniques are effective in a limited network with few sensors and intermediate devices (routers).[9]

2.3. Transport Layer Techniques

The Protocol-Datagram Congestion Control Protocol(DCCP) is developed by the IETF -the standard was accepted in year 2006. [12].

2.4. Techniques with Cross Layer Nature

This technique combines the mechanisms of different layers of the network operating system. Hop-by-hop flow control used by transport layer, traffic with limited rate at source and prioritized data link layer[2].The network protocols mechanisms provide congestion control by concentrating on buffer length and channel capacity. These protocols do not provide packet recovery mechanism.

3. IMPLEMENTATION

The implementation has been carried out in two parts: Simulating a WSN and applying classification techniques to predict the congestion.

3.1. Simulation of WSN

The WSN simulation of 7 nodes has been arranged in funnel like structure as shown below. The 2 FTP applications have started with at 0.5 and 1.0s. The FTP1 is at source 2 and destination 6 here as FTP 2 is at source 3 and destination 1. The simlation is traced for different time 3, 5, 7, 10, 13, 16, 19, 20, 21, 30, 40, 50 and with two different buffel length values with 10 and 20.The network parameters used for simulation are

Table 1
Network configuration parameters

Parameter	*Values*
Network Interface	WirelessPhy
MAC Type	802.11
Channel	Wireless Channel
Propagation	Two Ray Ground
Antenna	Omni Antenna
Queue	DropTail / PriQueue
Initial energy	Energy in joules
Reception Power	Receiving power in watts
Transmission Power	Transmitting power in watts

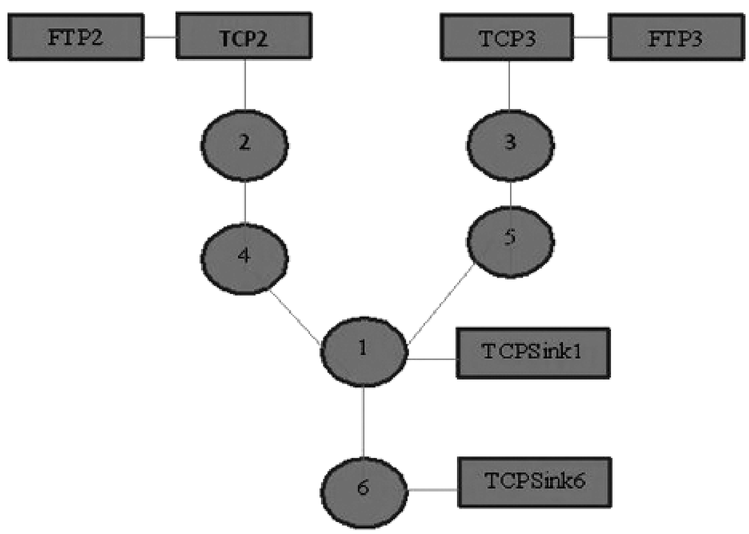
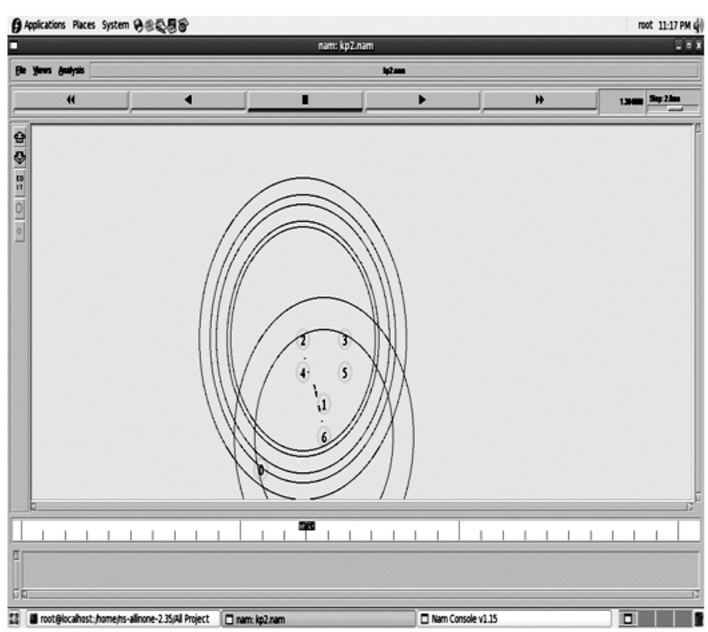


Figure 1: Network funnel like topology and application architecture



The network topology and application architecture is as shown below:

The simulation scenario snapshot is shown below: The packet transmission from node 2 to node 6 is visible in snap shot.

3.2. Application of Classification Techniques

The database used is from the network traffic data in wireless sensor network simulation. The AWK scripts are written to get the parameters-queue length, packet loss, throughput, packets delivery ratio and delay. The delay of a packet is the delay between the time packet was sent and the received time. The packet loss is the no. of packets dropped as buffer is full .The throughput is number of packets sent by the sender and the number of packets received by receiver .The packet delivery ratio(pdr) is no of bytes successfully received. These parameter values are dependent on the state of congestion in network and thus it makes it possible the occurrence of congestion to be predicted by using these parameters for a number of packets. The data set consist of ARFF (attribute relation file format) data file with attributes queuelength, delay, pdr, throughput, time, packetloss and classification attribute class {nocon, low, med, high}levels of congestion are low,medium or high. The delay in the computer networks is nothing but the time taken by the network to transmit a single packet from one node (source) to the other node (destination). This time interval is the delay for that packet. The network throughput is the rate of successful packet delivery over a communication channel. Throughput is usually measured in bits per second (bit/s or bps), and sometimes in data packets per second (p/s or pps) or data packets per time slot. The packet loss packet dropped due to no buffer available at node**.**The packet delivery ratio is the ratio proportion between the total numbers of packet sent and the total number of packets received. It is taken as the fraction of total packet sent by total packets received.

3.3. NNGE (Non-Nested Generalized Exemplars) classifier

Instance based classifiers rely on using directly the examples from the training set as concept models without constructing abstractions. Generalised exemplars uses examples that share the same class are grouped together, and so represent large rules more fully. This reduces the role of the distance function to determining the class when no rule covers the new example. It also reduces the number of classification errors that result from inaccuracies of the distance function, and increases the influence of large rules.

3.3.1. Rules generated

class high IF: 40.0<=stime<=50.0 ^ 20.0<=queueLength<=25.0 ^ 13725.6<=delay<=16319.1 ^ 670.91<=throughput<=674.4 ^ 0.0<=packetloss<=2.0 ^ 1.0<=pdr<=1.013 (3)class med IF : 21.0<=stime<=40.0 ^ 20.0<=queueLength<=25.0 ^ 8478.78<=delay<=12341.9 ^ 660.85<=throughput<=671.36 ^ 0.0<=packetloss<=10.0 ^ 1.01179<=pdr<=1.013 (3)class low IF : 19.0<=stime<=21.0 ^ queueLength=25.0 ^ delay=9137.69 ^ 655.53<=throughput<=657.99 ^ packetloss=0.0 ^ 1.001<=pdr<=1.025 (2)

class low IF: 3.0<=stime<=20.0 ^ 10.0<=queueLength<=20.0 ^ 888.535<=delay<=8513.02 ^ 486.7<=throughput<=658.88 ^ 1.0<=packetloss<=16.0 ^ 1.018<=pdr<=1.11 (6)

class high IF: 10.0<=stime<=50.0 ^ queueLength=10.0 ^ 888.535<=delay<=14520.3 ^ 627.77<=throughput<=675.54 ^ 36.0<=packetloss<=150.0 ^ 1.022<=pdr<=1.04 (9)

class med IF: stime=7.0 ^ queueLength=10.0 ^ delay=888.535 ^ throughput=603.5 ^ packetloss=23.0 ^ pdr=1.06 (1)

classnocon IF: 3.0<=stime<=10.0 ^ queueLength=20.0 ^ 824.966<=delay<=889.113 ^ 484.88<=throughput<=628.46 ^ packetloss=0.0 ^ 1.007<=pdr<=1.2 (4)

The confusion matrix indicates the actual class and predicted class by the applied classifier for all the instances in data set. For eg. There are 4 instances with no congestion in data set and classifier also has predicted as no congestion.

Table 1
Confusion matrix of NNGE algorithm

nocon	*low*	*med*	*high*	*class*
4	0	0	0	nocon
0	8	0	0	low
0	0	4	0	med
0	0	0	12	high

3.4. Best-First Decision Tree (BFTREE)

In best-first top-down induction of decision trees, the best split is added in each step, fore.g. the split that reduces the Gini index maximally in contrast to depth-first tree traversal.

packetloss< 29.5

| throughput< 659.865

| | delay < 856.7505: nocon(3.0/0.0)

| | delay >= 856.7505: low(8.0/2.0)

| throughput>= 659.865

| | delay < 13033.75: med(3.0/0.0)

| | delay >= 13033.75: high(3.0/0.0)

packetloss>= 29.5: high(9.0/0.0)

Size of the Tree: 9

Number of Leaf Nodes: 5

Table 2
Confusion matrix of BFTREE algorithm

nocon	*low*	*med*	*high*	*class*
3	1	0	0	nocon
0	8	0	0	low
0	0	1	3	med
0	0	0	12	high

After analyzing the results it can be stated that queue length and packet loss are 2 important attributes and the values of other attribute less than threshold indicates no congestion or different levels of congestion in network.

CONCLUSION

The different data mining classification algorithms can be applied to detect congestion in wireless sensor networks. The simulation is observed for 28 times with different time and buffer length and network data is used to build the data set.

Table 3
Comparison of 2 algorithms

Name of the Algorithm	*Accuracy in Detecting Congestion (%)*	*Time taken(seconds)*
NNGE	100.00	0.02
BFTREE	92.8572	0.02

The time taken to build model is almost same in both algorithm. However Based on confusion matrix and it can be clearly seen that accuracy of NNGE is100% and accuracy of BFTREE is 92.8572.We suggest that NNGE classification technique is more suitable to detect congestion in WSN and can be used more effectively.

REFERENCES

[1] Ekaterina Dawhkova and Andrei Gurtov "Survey on congestion control mechanisms for Wireless Sensor networks "center for wireless communication,finland,2013

[2] C. Wang.B. Li,KSohrarby and M. Daneshmand " UPstream Congestion control on wireless sensor networks through cross layer optimization" IEE journal on selected areas in communications, VOL 25, No. 4, May 2007.

[3] L. Breiman, J. Friedman, R. Olsen, and C. Stone."*Classification and Regression Trees*." Wadsworth International (California)

[4] Pierre Geurts, Ibtissam El Khayat, Guy Leduc," A Machine Learning Approach to Improve Congestion Control over Wireless Computer Networks"

[5] SasanAdibi ,"Data Mining - A Captured Wired Traffic Approach", International Journal of Advanced Science and Technology Vol. 21, No. 2, Aug, 2010

[6] Soukaena Hassan Hasheem"An Association Rules Analysis to enhance solving theCongestion Problem", Journal of Emerging Trends in Computing and Information Sciences Volume 2 No.7, JULY 2011

[7] Manoj Devare and Ajay Kumar , "Clustering and Classification (Time Series analysis) Based Congestion Control algorithm: Data Mining Approach", IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.9, September 2007 pp 241-246

[8] Sudip Misra, Isaac Woungang, Subhas Chandra Misra ,"Guide to Wireless Sensor Networks"

[9] Ameer Ahmed Abbasi, Mohamed Younis "A survey on clustering algorithms for wireless sensor networks"june2007**.**" Springer publications

[10] Shivangi Borasia and Vijay Raisinghani, "A Review of Congestion Control Mechanisms for Wireless Sensor Networks "

[11**]** Md. Abdur Rahman, Abdulmotaleb El Saddik and Wail Gueaieb, " Wireless Sensor Network Transport Layer:State of the Art " Springer-Verlag Berlin Heidelberg 2008

[12] Md. Mamun-Or-Rashid, Muhammad MahbubAlam, Md. AbdurRazzaque,andChoongSeon Hong "Reliable Event Detection and Congestion Avoidance in Wireless Sensor Networks"© Springer-Verlag Berlin Heidelberg 2007

[13] Vikram P. Munishwar, Sameer S. Tilak, and Nael B. Abu-Ghazaleh" Congestion and Flow Control in Wireless Sensor Networks "© Springer-Verlag London Limited 2009

[14] Tan,Steinbach,Vipinkumar"Introduction to Data mining"

[15] Han J., and Kamber, M."Data Mining: Conceptsand Techniques". Morgan Kaufmann, 2006.

[16] Mohammad Abu Alsheikh, Shaowei Lin, DusitNiyato and Hwee-Pink Tan"Machine Learning in Wireless Sensor Networks:Algorithms, Strategies, and Applications

[17] Qi Li, ZongwuKe, Duanfeng Xia, Yuxia Sun "A Routing Protocol for Wireless SensorNetworks with Congestion Control"

[18] HemmatSheikhi MousaDashti Mehdi Dehghan "Congestion Detection for Video Traffic in Wireless Sensor Networks"2011 IEEE

[19] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Pe-terReutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Exlorations, Volume 11, Issue 1.

[20] Prabu Rajkumar P, Arul Treesa Mathew, Sruthi N Paul, Sujitha B Cherkottu "Congestion Control in Healthcare Wireless Sensor Networks- A Data Centric Approach"Technology Research (IJSETR) Volume 2, Issue 7, July 2013

[21] Eibe Frank, Ian H. witten "Generating Accurate rule sets without global optimization" Fifteenth International Conference on Machine Learning, 144-151, 1998.

[22] William H. Cohen "Fast Effective Rule Induction" Proceedings of the Twelth International Conference on Machine Learning,1995

[23] C4.5: by J. Ross Quinlan.. Programs for M Publishers, Inc., 1993Machine Learning Morgan Kaufman

[24] Mrs. U. Urathal alias Sri Swathiga, Dr. C. Chandrasekar " An Efficient Fuzzy based Congestion Control Technique for Wireless Sensor Networks International Journal of Computer Applications (0975 – 8887) Volume 40– No.14, February 2012

[25] Mrs. U. Urathal alias Sri Swathiga, Dr. C. Chandrasekar "Congestion Prediction And Adaptive Rate Adjustment Technique For Wireless Sensor Networks" International Journal of Research in Computer Science ISSN 2249-8265 Volume 3 Issue 4 (2013) pp. 1-7 www.ijorcs.org, A Unit of White Globe Publications doi: 10.7815/ijorcs.34.2013.067

[26] Mrs. Jayashri B. Madalgi and Dr. S. Anupama Kumar "Efficiency of Classification Techniques in Detecting Congestion in Wireless Sensor Networks "International Journal of Advanced Research in Computer and Communication Engineering,Vol. 5, Issue 3, March 2016,DOI 10.17148 / IJARCCE.2016.5332