Cooperative Learning for Distributed In-Network Traffic Classification

Cooperative Learning for Distributed In-Network Traffic Classification

S.B.Joseph[1], H.R.Loo[1], I.Ismail [1],T. Andromeda[2], M.N.Marsono[1]

[1]Department of Electronics and Computer Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia
[2]Department of Electrical Engineering, Faculty of Engineering, Universitas Diponegoro, Semarang, Indonesia

sjbassi74@gmail. com, loohuiru@gmail.com, ismahani@fke .utm.my, trias1972@gmail.com; nadzir@fke . utm.my

Abstract. Inspired by the concept of autonomic distributed/decentralized network management schemes, we consider the issue of information exchange among distributed network nodes to network performance and promote scalability for in-network monitoring. In this paper, we propose a cooperative learning algorithm for propagation and synchronization of network information among autonomic distributed network nodes for online traffic classification. The results show that network nodes with sharing capability perform better with a higher average accuracy of 89.21% (sharing data) and 88.37% (sharing clusters) compared to 88.06% for nodes without cooperative learning capability. The overall performance indicates that cooperative learning is promising for distributed in-network traffic classification.

1. Introduction

Network traffic classification is a crucial network processing task for network traffic management. Traffic measurement and classification enable network administrators to be aware of the current network state. Data stream mining algorithms [1, 2] have been explored recently for online traffic classification to overcome the shortcoming of conventional data mining algorithms. They are designed to cope with concept drift, adapt to new knowledge and react to changes promptly.

Recently, distributed network management (DNM) strategies such as [3], have been introduced as a solution to the increasing management complexity of communication networks. These strategies are autonomic and decentralized in nature, towards improving network performance, scalability and to reduce human participation. Each DNM entity participates in a distributed management process, which requires cooperation among DNM entities to monitor, analyze and make decision to achieve global network objectives. Cooperative Learning (CL) enables network nodes to discover each other, exchange information, disseminate local decision, enhance self-adaptation of nodes, improve scalability and finally enforce management decision [4].

This paper proposes an analysis for online distributed in-network traffic classification based on our CL framework in [5], using incremental k-means network traffic classification [1]. This paper analyzes the effect of information exchange among nodes on the overall classification accuracy. We tested our proposed method on two schemes: sharing of training labeled data in the form of information and sharing of clusters in the form of knowledge. Our proposed system has been applied to UPC [6] real network datasets in order to evaluate the performance of such system. This system is able to classify network traffic online with an average accuracy of 89.21% and 88.37% sharing clusters, compared to 88.06% over the non-sharing approach.

The remainder of this paper is structured as follows. Section 2 introduces related works on monitoring networks, DNM and CL. Section 3 presents online traffic classification and our proposed method. Section 4 analyzes the experimental results. Conclusion is in Section 5.

2. Related Works

The rapid growth of Internet traffic today practically demonstrates the concept of Big Data, characterized by having large data size with millions of network connections (variety, volume, veracity and velocity), which motivated the proposals of embedding network management functions into the network itself, e.g. [3] towards managing the network easier, automatically and with reduced complexity.

Several research works have also proposed different monitoring schemes for different network architecture over the years. Stadler et al. [7] has outlined principles for decentralised monitoring using spanning trees and gossiping methods. However, their techniques are targeted for network aggregates and not for online measurements. Raz et al [8] has presented an insight of network monitoring algorithms and also several monitoring algorithmics that utilize monitoring techniques. Although these techniques are proposed for distributed monitoring, they did not consider the issue of collaborative/cooperative learning among autonomic nodes.

In a distributed system where each node is semi or fully autonomic in solving a global problem, the role of interaction/cooperation among such distributed nodes can not be over emphasized. References [4, 9] proposed interaction/cooperative learning among distributed networks for improving accuracy and scalability. The issues of nodes discovery for interaction among nodes is presented in [9], network nodes discover one another in a network in other to cooperate with each other. In the same note, reference [4] proposed bootstrapping and discovery mechanisms to ensure the proper information dissemination in distributed network process. However, these methods are proposed for discovering neighbouring nodes, they did not consider the task of nodes interaction for information sharing. Finally, we consider the issue of information/knowledge sharing among distributed network nodes, through cooperative learning.

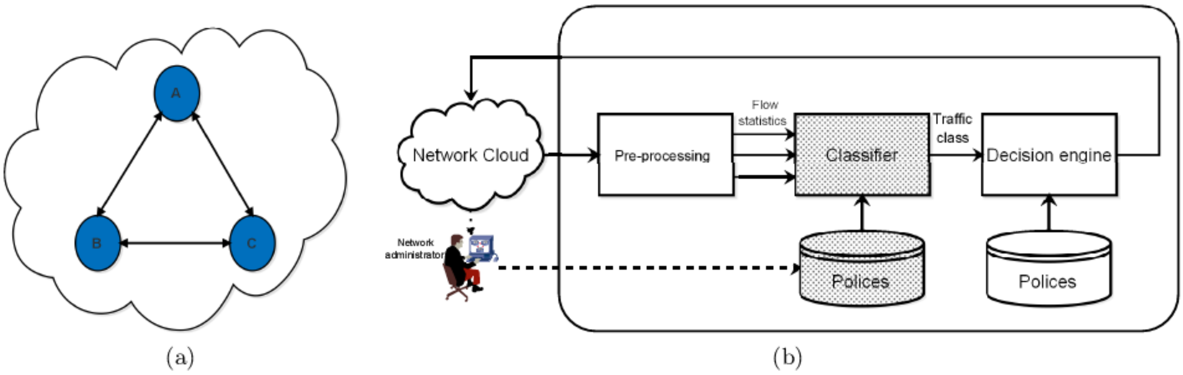
3. Online Network Traffic Classification With Cooperative Learning

Table 1: List of online features selected for online classification

ID	Name	Long Description
01	Source Port	Source port number
02	Destination	Destination port number
03	TL IP	Total bytes in IP packet
04	UL IP	Total bytes in IP packet (uplink)
05	DL IP	Total bytes in IP packet (downlink)
06	TL Eth	Total bytes in Ethernet packet
07	UL Eth	Total bytes in Ethernet packet
08	DL Eth	Total bytes in Ethernet packet
09	TL Ctr	Total control bytes in packet
10	UL Ctr	Total control bytes in packet
11	DL Ctr	Total control bytes in packet

We propose a network traffic classification technique with cooperative learning capabilities to support distributed network monitoring and to enhance the monitoring capabilities of the DNM nodes. The system framework is based on our proposed system in [5], which consists of three distinct layers, each performing a different successive task towards the achievement of global management objectives. In this paper, we only focus on classification layer and the effect of cooperative learning. Different from [5], we increase the number of nodes to show the scalability of the system framework. Figure 1 shows the framework of our proposed system. Network traffic are sensed at different network locations by autonomic nodes. These nodes share newly derived information in the form of information or knowledge among each other to improve their states.

In this paper, we propose two sharing methods for cooperative learning for network traffic classification. We consider the sender of information as host node, while the receiving node as client node.



(a) (b)

Figure 1: System block diagram (a) Cooperating nodes, (b) Individual node structure

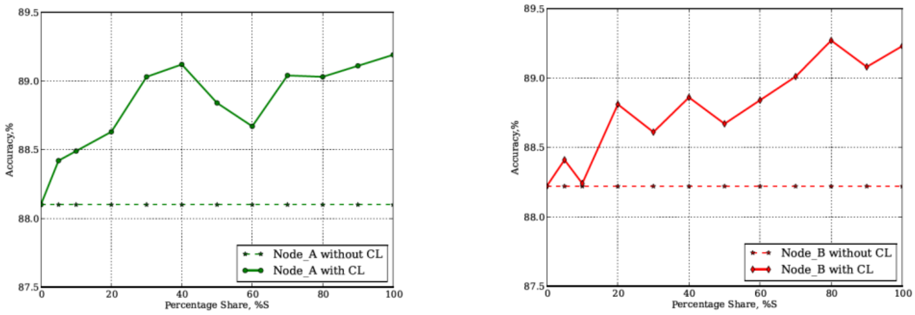
3.1. *Scheme 1: Sharing of information*

In this scheme cooperative learning is initiated by the host node upon receiving a labeled flow instance (x_iy_i) from the network administrator for retraining. The host retrains its model and at the same time shares the labeled instance with randomly selected neighboring node. Upon receiving shared labeled instance, a client node will update its state by retraining its model with the new instance. This sharing is done based on Algorithms 1.

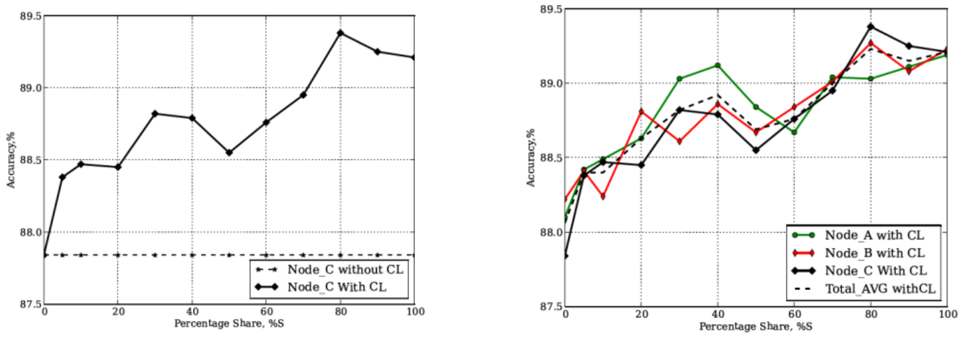
3.2. *Scheme 1: Sharing of information*

This sharing scheme is activated by a node upon creating a new cluster after retraining with received labeled flow instance, the host shares newly created cluster (c_j) with randomly selected neighboring nodes. Each of the nodes will either merges or creates new cluster. This sharing is done based on Algorithms 2.

Let x_i, represents incoming data stream; M is classification model; c_1c_2 are first and second nearest cluster from x_i; y_i and y'_i are the true and predicted labels for x_i; x_j and y'_j are labeled data from nearest neighboring node; c_j is newly created cluster. The classifications are initiated upon receiving incoming flow instances. In both schemes not all information received from network administrator will be shared, host will only share a percentage (%S) of the received information to a neighboring client. This is to minimize network resources overhead due to sharing.



(a) (b)



(c) (d)

Figure 2: Comparison of average accuracy plots for scheme 1 in (a) Node A, (b) Node B, (c) Node C, (d) All Nodes

Algorithm 1 Proposed Cooperative Learning Algorithm (Case 1)

```
while new $x_i$ do
    $y_i$ = Classify M,$x_i$
    if (x_i is labeled) then
    $y_i$ is known
    retrain(M,$x_i$,$y_i$)
    select random neighbor
    share ($x_i$,$y_i$)to selected neighbor
end if
end while
Re-Training
if ($x_j$ $y_j$) received then
    retrain (M,$x_j$,$y_j$)
end if
```

Algorithm 2 Proposed Cooperativ Learning Algorithm (Case 2)

```
while new $x_i$ do
    $y_i$ = Classify M,$x_i$
    if ($x_i$ is labeled) then
        $y_i$ is known
        retrain (M,$x_j$,$y_j$)
    end if
    if ($c_j$)is created then
        select random neighbor
        share ($c_j$)to selected neighbor
    end if
end while
Re-Training
if ($c_j$)is received then
    find nearest cluster to ($c_j$)
    if ($c_j$)is close to nearest cluster then
        merge ($c_j$)to nearest cluster
    else
        adapt($c_j$)as new cluster
    end if
end if
```

Table 2: Summary of results

Nodes	Scheme 1			Scheme 2		
	Accuracy (%) without CL	Accuracy (%) with CL	Improvement (%)	Accuracy (%) without CL	Accuracy (%) with CL	Improvement (%)
A	88.10	89.19	1.09	88.10	88.59	0.49
B	88.22	89.23	1.01	88.22	88.31	0.09
C	87.84	89.21	1.37	87.84	88.22	0.38

4. Results and Discussion

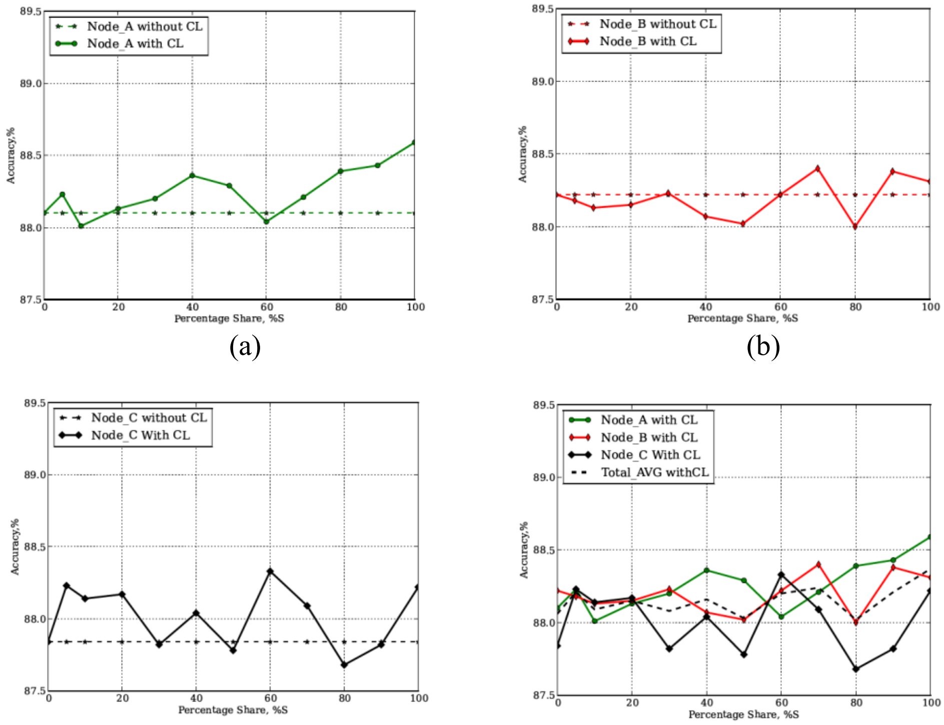
4.1. Dataset

We evaluated our proposed method using real network datasets. The UPC dataset [6] is used for the experiment; a total of 339061 flow instances was extracted. We represented each host address as an autonomic node representing different sensing segment in our experiment. We used eleven (11) attributes, and six (6) traffic classes as suggested in [1] as listed in Table 1.

4.2. Performance Evaluation

The average accuracy plots of individual nodes with and without cooperative learning for scheme 1 is presented in Figure 2. Figure 3 shows the cumulative accuracy plots for nodes with and without cooperative learning for scheme 2. Both plots show higher classification accuracy on each nodes with cooperative learning compared to nodes without cooperative learning. The comparison of average accuracy of each nodes with different percentage of sharing (%S) shows that with at least 5% of sharing we can increase the accuracy by 0.34% for scheme 1 and 0.15% for scheme 2.

Table 2 presents a summary of accuracy at each node for our proposed method. The experimental results indicate an improvement on the classification accuracy when cooperative learning is activated compared to independent accuracy without cooperative learning. An improvement in accuracy was observed for both classification models. The overall performance indicates that cooperative learning is promising for distributed network monitoring.



(a) (b)

(c) (d)

Figure 3: Comparison of average accuracy plots for scheme 2 in (a) Node A, (b) Node B, (c) Node C, (d) All Nodes

5. Conclusion

The proposed CL algorithm for online distributed network performance monitoring based on incremental k-means traffic classification algorithm has been presented. In difference to existing classification methods, the proposed method explores the capabilities of cooperative learning to improve the accuracy, reliability and scalability of classification accuracy. The performance of the proposed method was justified with real network traces. The obtained results have shown the significant increment in classification knowledge of nodes, and improvement in accuracy of classification in cooperative learning network nodes

Acknowledgment

This work is supported in part by CREST grant (UTM Vote No. 4B176) and Universiti Teknologi Malaysia Grant matching (UTM Vote No. 00M75)

References

[1] H. Loo, S. Joseph, and M. Marsono, "Online incremental learning for high bandwidth network traffic classification," Applied Computational Intelligence and Soft Computing, vol. 2016, 2016.

[2] G. Mingliang, H. Xiaohong, T. Xu, M. Yan, and W. Zhenhua, "Data stream mining based real-time highspeed traffic classification," in Broadband Network & Multimedia Technology, 2009. IC-BNMT'09. 2nd IEEE International Conference on, pp. 700–705, IEEE, 2009.

[3] D. Dudkowski, M. Brunner, G. Nunzi, C. Mingardi, C. Foley, M. P. de Leon, C. Meirosu, and S. Engberg, "Architectural principles and elements of in-network management," in Integrated Network Management, 2009. IM'09. IFIP/IEEE International Symposium on, pp. 529–536, IEEE, 2009.

[4] L. Guardalben, S. Sargento, P. Salvador, and V. Mirones, "A cooperative hide and seek discovery over in network management," in Network Operations and Management Symposium Workshops (NOMS Wksps), 2010 IEEE/IFIP, pp. 217–224, IEEE, 2010.

[5] S. Joseph, H. Loo, I. Ismail, and M. Marsono, "Cooperative learning for online in-network performance monitoring," in Communications (MICC), 2015 IEEE Malaysia International Conference on, pp. 219–224, Nov 2015.

[6] V. Carela-Español, T. Bujlow, and P. Barlet-Ros, "Is our ground-truth for traffic classification reliable?," in Passive and Active Measurement, pp. 98–108, Springer, 2014.

[7] R. Stadler, M. Dam, A. Gonzalez, and F. Wuhib, "Decentralized real-time monitoring of network-wide aggregates," in Proceedings of the 2nd Workshop on Large-Scale Distributed Systems and Middleware, p. 7, ACM, 2008.

[8] D. Raz, R. Stadler, C. Elster, and M. Dam, "In-network monitoring," in Algorithms for Next Generation Networks, pp. 287–317, Springer, 2010.

[9] K. M. Konwar, D. Kowalski, and A. A. Shvartsman, "Node discovery in networks," Journal of Parallel and Distributed Computing, vol. 69, no. 4, pp. 337–348, 2009.