10th International Scientific Conference
Technics, Informatics, and Education – TIE 2024
20-22 September 2024

Session: *Computer Sciences and Information Technology*
Professional paper
DOI: 10.46793/TIE24.136P

Detection of Broadcast Storms in Local Area Networks

Uroš Pešović[1*] [0000-0001-8722-6544], Slađana Đurašević Pešović[1] [0000-0002-3598-5781], Biljana Savić[1] [0000-0002-9560-6077] and Dušan Marković[2] [0000-0002-7270-6702]
[1] Faculty of technical sciences, University of Kragujevac, Čačak, Serbia
[2] Faculty of agriculture, University of Kragujevac, Čačak, Serbia
[*] uros.pesovic@ftn.kg.ac.rs

Abstract: *Broadcast storms represent events in the local area network (LAN) which is the result of excessive amounts of packet retransmissions by network switches which drastically reduce network performance and even overload network infrastructure which can result in that network becoming inoperational. Broadcast storms originate from broadcast and multicast packets which are targeted for all or groups of stations in LAN. Since these packets are addressed to multiple destinations, they are usually forwarded by network switches on multiple ports to reach all targeted destinations. This can become a serious problem in institutional LANs in which broadcast storms can create a temporary or permanent overload of the entire network infrastructure. This is most notably observed during the videoconferencing calls which could cause packet loss or connection interruptions. In this paper, we analyzed the influence of broadcast storms on network performance in institutional LAN and proposed a machine learning algorithm for the detection of broadcast storms based on the network traffic data collected by the Packet Capture (PCAP) Wireshark software.*

Keywords: *broadcast storm; network performance; network switch; PCAP; Wireshark*

1. INTRODUCTION

Computer networks are an important infrastructure component in every institution since computers are used in almost all segments of human activity. Computer networks connect computers to provide the following activities: communication, global access to information, resource sharing, remote access, etc. The physical organization and complexity of a computer network are usually invisible to the ordinary user, who expects a high speed of access, reliability and security of transmission and zero downtime deployment. Seamless operation of computer networks is the responsibility of the network administrator, who maintains and troubleshoots problems in computer networks.

Computer networks in institutions are organized in the form of local area networks, in which computers are connected in tree topology using networking components called switches. The primary role of network switches is to redirect packets sent from the source towards the destination. Switch analyzes the header of the received packet and based on the destination MAC address sends the packet via the appropriate port [1]. After the bootup, the switch starts to build its SAT (Source Address Table) based on the packet MAC (Medium Access Control) addresses of received packets. This table is used to associate to which port a certain network station is connected. To redirect the packet, the switch searches its SAT table to determine which port it should use for the redirection. If there is no such SAT entry, the packet is redirected to all other ports except the packet source port which ensures that the packet will reach an unknown station. When this unknown station replies to the received packet, the reply packet will come through one of the ports and the MAC address of the unknown station will be entered in the SAT table.

This concept of switching using SAT tables works well only if one station is connected to just one of the ports. Thus in local area networks the only accepted topology is tree topology, in which there is only one route between source and destination stations. Loops are not allowed because they could confuse the switching logic since one station can appear to be connected to two ports. This situation could cause that packet to be routed through an infinite loop. These redundant loops are detected by STP (Spanning Tree Protocol) which is used by switches to disable loop connections to establish tree topology.

Besides unicast traffic which is destined for specific network stations, multiple stations can be marked as packet destinations in case of multicast traffic, or all stations in the network can be marked in case of broadcast traffic. Broadcast is a special type of packet in which the destination address is set to all binary ones (or FF:FF:FF:FF:FF:FF in hexadecimal) and is intended to be received by all stations in the network. In the presence of loops in LAN topology these packets could cause a high inrush in network traffic a so-called broadcast storm [2]. These situations can temporarily or permanently interfere with regular LAN operation, which could cause a significant drop in network performance.

Broadcast storms can also occur in LAN without the loops where there is a source that generates frequent broadcast packets [3]. The source of such behavior can be multiple, improperly configured network services, malicious software and others. Troubleshooting of broadcast storms represents a challenge for every administrator, especially in large LAN networks. In this paper, we presented the algorithm that can be used to monitor network traffic and classify the presence of broadcast storms.

The most common networking technology based on the IEEE 802 standard uses Ethernet for wired networks and IEEE 802.11 for wireless networks. Each packet contains two address fields, destination MAC address and source MAC address. MAC addresses are uniquely assigned to every NIC (Network Interface Controller), and network switches use these MAC addresses to redirect packets from the source towards their destination. Switching operation is transparent to networked stations which are not aware of the presence of the switch. All network traffic that reaches certain network stations can be monitored with packet capture tools, such as Wireshark [4]. Wireshark is a software networking tool used to analyze network traffic via the network interface card. This tool uses NIC in promiscuous mode in which NIC allows all frames to pass, allowing the computer to read frames intended for other machines or network devices. Under regular operation (non-promiscuous mode) when the NIC receives a frame, it drops if it’s not addressed to the NIC's MAC address.

Packet capture is a networking tool that intercepts packets that are sent through the network. After these packets have been captured they are saved for further analysis by the network administrators. Analysis of these captured packets enables administrators to identify and solve networking problems that can influence regular daily operations. Packet capture can help in the analysis and identification of problems regarding network performance (packet loss, network congestion, etc), network security, intrusion detection, etc. In order to monitor traffic on certain links that are not redirected to packet capture NIC, administrators use managed switches to perform port mirroring. By using port mirroring all data traffic on certain ports can be mirrored to a port that is connected to packet capture NIC, where is been analyzed by packet capture software (Fig 1).

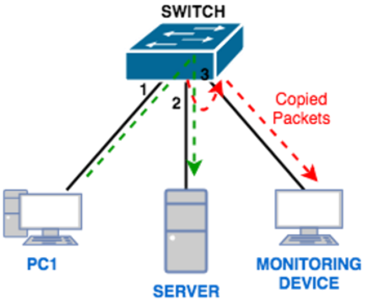


Figure 1. *Packet capture using port mirroring*

Broadcast storm needs to be separated from regular network protocols which rely on the broadcast packets. For example, DHCP (Dynamic Host Configuration Protocol) uses DHCP Discover and DHCP Request broadcast packets to find and request IP configuration from the DHCP server. Also, ARP (Address Resolution Protocol) sends a broadcast packet in order to find the MAC address of the unknown station.

Clues that could identify the source of the broadcast storm in LAN are:

- frequent broadcast packets
- broadcast packets with the same length
- broadcast packets which use the same protocol
- broadcast messages sent by a single network station
- broadcast messages sent by a group of closely addressed stations (similar IP or MAC addresses).

2. METHODOLOGY

In this paper, we proposed a clustering algorithm that enables to detect the presence of broadcast storms in captured packets and identify sources of such traffic. Since the number of packets in large networks can be exceptionally high it is difficult and impractical to classify them by hand. We employ machine learning algorithms to classify the data as regular traffic and broadcast storms.

Clustering is a key task in discovering useful patterns in large data sets that are not pre-labeled to belong to a certain class [5]. Clustering is a process in which objects, according to predefined properties, are arranged into groups, called clusters in such a way that objects in the same cluster are more similar to each other than to objects in other clusters. Thus by being able to separate objects into classes, clustering provides new knowledge about observed phenomena. Unlike supervised learning, clustering cannot be fully automated, since it doesn’t use label data to calculate exact model performance. Instead, clustering requires human judgment and domain knowledge in order to select appropriate data and iteratively adjust model parameters to achieve the desired result. The key criteria in clustering revolve around usefulness of generated clusters and did they discover new patterns in the data which haven't been known before clustering.

The first step in the clustering process is to define the expected number of clusters in the data. Sometimes if the number of expected clusters is not known in advance, the optimal number can be determined by iteratively increasing the number of clusters and finding the most satisfactory result. Figure 2 shows the process of clustering the unlabeled data shown on the left figure into six clusters marked with different colors shown in the right figure.

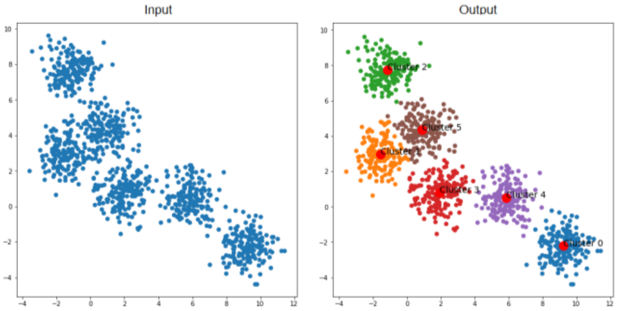


Figure 2. *Clustering of unlabeled data*

Clustering algorithms can be separated into following categories:

- Centroid-based Clustering,
- Density-based Clustering
- Connectivity-based Clustering
- Distribution-based Clustering

The choice of clustering algorithm depends mainly on the nature of the data. Data which tends to group around certain points can be cluttered with centroid-based algorithms, such as K-means. In case of data which is not clearly separable, density based clustering algorithms are used, such as DBScan. When data set has tendency to group in elongated shapes connectivity based clustering algorithms are used. The agglomerative hierarchical clustering is the most widely used type of connectivity-based clustering [6].

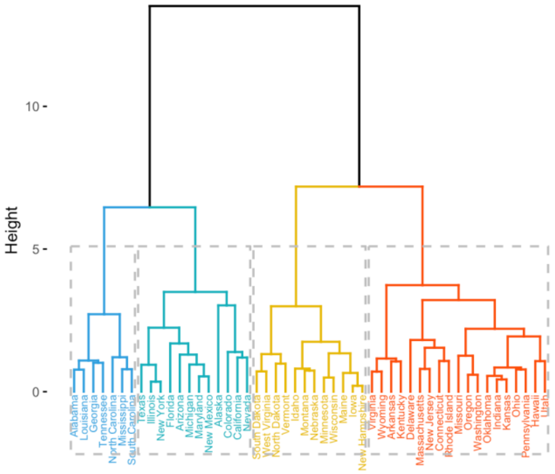
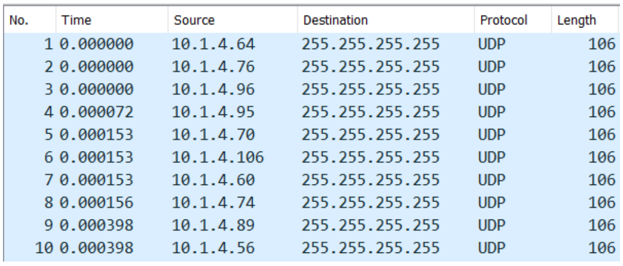


Figure 3. *Clustering dendogram*

First, the algorithm treats every data point as a separate cluster, after which it starts to merge most similar points into the cluster. The algorithm ends when all data points are connected into one big cluster, resulting in the formation of a tree-based representation, called a dendogram. Depending on how many clusters are required, the dendrogram is cut at a certain value, as shown in Fig 3, with four separate clusters.

3. RESULTS AND DISCUSSION

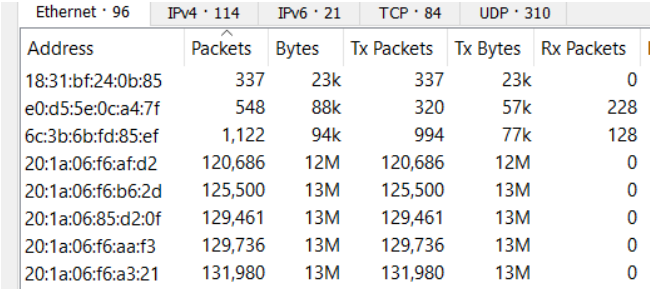
In this paper, we analyzed broadcast storms that originate from an improperly configured network service installed in one of the computing classrooms. Upon powering on, all student computers in this classroom start the service which sends broadcast packets in order to find the running server. This operation should end briefly as soon as the server responds, but when this server is not powered on host computers start to permanently send these broadcast packets. These broadcast packets created a broadcast storm in the entire faculty network reducing the available bandwidth and inducing unnecessary network traffic. Part of the broadcast traffic captured by Wireshark is presented in Fig 4.



No.	Time	Source	Destination	Protocol	Length
1	0.000000	10.1.4.64	255.255.255.255	UDP	106
2	0.000000	10.1.4.76	255.255.255.255	UDP	106
3	0.000000	10.1.4.96	255.255.255.255	UDP	106
4	0.000072	10.1.4.95	255.255.255.255	UDP	106
5	0.000153	10.1.4.70	255.255.255.255	UDP	106
6	0.000153	10.1.4.106	255.255.255.255	UDP	106
7	0.000153	10.1.4.60	255.255.255.255	UDP	106
8	0.000156	10.1.4.74	255.255.255.255	UDP	106
9	0.000398	10.1.4.89	255.255.255.255	UDP	106
10	0.000398	10.1.4.56	255.255.255.255	UDP	106

Figure 4. *Wireshark captured broadcast traffic*

By analyzing this traffic we can observe that multiple stations generate frequent UDP broadcast requests each having the same length. The entire packet capture file was recorded at two-minute interval, in which 2.6 million broadcast packets were sent, on average 20000 packets per second. This traffic generated cumulative load on entire LAN of around 16Mbps. which can cause significant drop of LAN performance, especially in networks segments which operate at 100Mbps. Ethernet statistics for the packet capture were exported in the form shown in Fig. 5.



Address	Packets	Bytes	Tx Packets	Tx Bytes	Rx Packets
18:31:bf:24:0b:85	337	23k	337	23k	0
e0:d5:5e:0c:a4:7f	548	88k	320	57k	228
6c:3b:6b:fd:85:ef	1,122	94k	994	77k	128
20:1a:06:f6:af:d2	120,686	12M	120,686	12M	0
20:1a:06:f6:b6:2d	125,500	13M	125,500	13M	0
20:1a:06:85:d2:0f	129,461	13M	129,461	13M	0
20:1a:06:f6:aa:f3	129,736	13M	129,736	13M	0
20:1a:06:f6:a3:21	131,980	13M	131,980	13M	0

Figure 5. *Statistics of broadcast generation by MAC addresses*

Ethernet statistics was exported into MATLAB where we perform agglomerative hierarchical clustering analysis. First MAC addresses which are shown in hexadecimal format were converted into numerical data. Data used for clustering algorithm was MAC address, number of packets and amount of transferred data in bytes. We expected that selected data will be well grouped for nodes which are generating broadcast storm, than for nodes which generate regular network traffic. Clustering algorithm was set to find two clusters, one which will be marked as regular traffic and other marked as broadcast storm as shown in Fig 6.

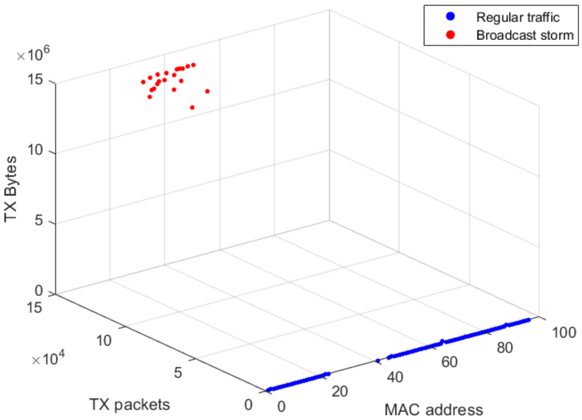


Figure 6. *Clustering broadcast storm traffic*

As a result, we were able to classify and identify 20 MAC addresses that are the source of broadcast storm as shown in Table 1.

Table 1. *Identified NIC's MAC addresses as source of broadcast storms*

No.	MAC address
1.	20:1a:06:85:d2:0f
2.	20:1a:06:f6:a2:f8
3.	20:1a:06:f6:a3:21
4.	20:1a:06:f6:a4:ee
5.	20:1a:06:f6:a7:b4
6.	20:1a:06:f6:a8:2e
7.	20:1a:06:f6:a8:c6
8.	20:1a:06:f6:a9:29
9.	20:1a:06:f6:aa:38
10.	20:1a:06:f6:aa:f3
11.	20:1a:06:f6:ad:8c
12.	20:1a:06:f6:af:d2
13.	20:1a:06:f6:b3:09
14.	20:1a:06:f6:b3:29
15.	20:1a:06:f6:b3:42
16.	20:1a:06:f6:b3:44
17.	20:1a:06:f6:b3:91
18.	20:1a:06:f6:b5:ff
19.	20:1a:06:f6:b6:2d
20.	20:89:84:30:6a:ca

4. CONCLUSION

In this paper, we analyzed the influence of broadcast storms in large institutional LANs. We proposed a clustering algorithm that can be used to classify and identify sources of broadcast storms from data obtained from packet capture software. As a result, we identified 20 MACs from which broadcast storms are transmitted and network service running on those computers will be configured to mitigate its influence on the institution's LAN. The proposed algorithm can be realized as a service implemented on the managed switch. Further work will be focused on the application of other clustering algorithms for the detection of broadcast storms.

ACKNOWLEDGEMENTS

This study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and these results are parts of the Grant No. 451-03-66/2024-03/200132 with University of Kragujevac – Faculty of Technical Sciences Čačak, and Grant 451-03-66/2024-03/ 200088 with University of Kragujevac – Faculty of Agriculture Čačak.

REFERENCES

[1] Stallings W. (2013). *Data and Computer Communications (10th. ed.)*. Prentice Hall Press, USA, ISBN:978-0-13-350648-8

[2] Tamrakar S. (2021). *Transmission Redundancy and Avoidance of Broadcast Storm Problem in Mobile Ad Hoc Networks*. LAMBERT Academic Publishing, ISBN: 978-6203042115

[3] Sarkar D., Rakesh N. and Mishra K. (2018). *Broadcast Storm Problem—A Hidden Consequence of Content Distribution in Content Delivery Networks*, Chapter in book n book: Networking Communication and Data Knowledge Engineering (pp.155-165), DOI: 10.1007/978-981-10-4585-1_13

[4] Sharpe R., Warnicke E., Lamping U., Wireshark User's Guide Version 4.5.0, https://www.wireshark.org/download/docs/Wireshark%20User%27s%20Guide.pdf

[5] Tan P.N, Steinbach M., Karpatne A, Kumar V. (2018), *Introduction to Data Mining (Second Edition)*, Praeson, ISBN: 978-0133128901

[6] Miyamoto S. (2022), *Theory of Agglomerative Hierarchical Clustering*, Springer, ISBN: 978-981-19-0419-6