applied sciences

MDPI

Article

A Hybrid Online Classifier System for Internet Traffic Based on Statistical Machine Learning Approach and Flow Port Number

Hamza Awad Hamza Ibrahim *, Omer Radhi A. L. Zuobi, Awad M. Abaker and Musab B. Alzghoul

College of Computer at Al-Gunfudah, Umm Al-Qura University, Mecca 24382, Saudi Arabia; orzubi@uqu.edu.sa (O.R.A.L.Z.); amabker@uqu.edu.sa (A.M.A.); mbzghool@uqu.edu.sa (M.B.A.)
* Correspondence: haibrahim@uqu.edu.sa

Abstract: Internet traffic classification is a beneficial technique in the direction of intrusion detection and network monitoring. After several years of searching, there are still many open problems in Internet traffic classification. The hybrid classifier combines more than one classification method to identify Internet traffic. Using only one method to classify Internet traffic poses many risks. In addition, an online classifier is very important in order to manage threats on traffic such as denial of service, flooding attack and other similar threats. Therefore, this paper provides some information to differentiate between real and live internet traffic. In addition, this paper proposes a hybrid online classifier (HOC) system. HOC is based on two common classification methods, port-base and ML-base. HOC is able to perform an online classification since it can identify live Internet traffic at the same time as it is generated. HOC was used to classify three common Internet application classes, namely web, WhatsApp and Twitter. HOC produces more than 90% accuracy, which is higher than any individual classifiers.

Keywords: Internet traffic classification; machine learning; classification methods; port-based method; hybrid classifier; online Internet traffic classification; live Internet traffic

Citation: Ibrahim, H.A.H.; Zuobi, O.R.A.L.; Abaker, A.M.; Alzghoul, M.B. A Hybrid Online Classifier System for Internet Traffic Based on Statistical Machine Learning Approach and Flow Port Number. *Appl. Sci.* **2021**, *11*, 12113. https://doi.org/10.3390/app112412113

Academic Editors: Andrea Prati, Carlos A. Iglesias, Luis Javier García Villalba and Vincent A. Cicirello

Received: 18 November 2021
Accepted: 14 December 2021
Published: 20 December 2021

1. Introduction

Internet has grown explosively and rapidly with the ability to provide numerous services for a wide range of Internet applications. The Internet includes a huge amount of applications and protocols such as http, https, file transfer, e-mail, databases, VOIP, etc.

Internet Server Provider (ISP) and network operators are usually interested in knowing the traffic carried in their networks for the purpose of optimizing network performance and security issues. Therefore, network traffic classification is an important foundation for identifying unknown Internet applications which have abnormal behaviors. In particular, network classification can detect traffic which includes threats, such as denial of service, flooding attacks and other such threats [1,2].

With the increase in Internet usage, a lot of Internet applications have been developed. However, these new applications can carry abnormal Internet traffic, which has a negative effect on network performance. Some of the Internet applications generate several types of versions with different traffic attributes. For instance, online games usually constitute a huge number of the overall games across the world each year. Network traffic classification is invaluable for identifying these large applications. Network traffic management is another issue which shows the importance of Internet traffic classification. When network managers plan to control network users through fair usage of bandwidth, they need to first know which type of applications they are dealing with. Thus, the managers cannot achieve their administrative tasks, unless they classify the network traffic. In the home network, traffic classification can help to enhance the Quality of Service (QoS) of Internet services. In general, the identification of Internet traffic helps in classifying of different types of attackers [1].

1.1. Classification Methods

The port-based classification method is based on the 16-bit port numbers on the transport layer, which consist of the information on source and destination ports. Put simply, the classifier uses these port numbers to determine the application classes. In other words, the classifier reads the port number from the Internet Assigned Numbers Authority (IANA) and then uses this number to distinguish between the Internet application types. The port-based classification method has the following advantages: (i) it is very simple, (ii) it can be used to limit worm traffic, (iii) it is very fast, (iv) it can be applied by all routers and layer three switches, and (v) it is efficient in classifying protocols carried by a fixed port number [2]. However, this method is not sufficient to classify the new Internet applications that use unknown-port numbers [3–5].

Payload-based classification or Deep Packet Inspection (DPI) is an individual packet inspection looking for unique signatures. This means that the packets will be investigated one by one to find a unique signature. This helps in establishing that the packet belongs to a particular class (application). According to researchers [6–8], payload-based classification methods overcome the port number dependency problem and achieve higher accuracy. This is generically due to the fact that the unique signature (if it exists) always tells the truth, hiding nothing. The type of signature used in the classification of traffic is based on the Internet application type and can be found in applications or transport layers. The classifier can use the signature in text (string) or hexadecimal (HEX) formats. The classifier uses these signatures to decide which packet/flow belongs to which application.

Another method is machine learning (ML)-based or flow statistical-based and it uses a collection of information to classify the network traffic. The main advantage of this approach is that it can be used at any point in the network [2]. Unlike port and payload-based methods, which are based on specific port numbers and unique signatures, the statistical-based method can identify the traffic based only on statistical features calculated from network flows. Machine learning (ML) is the most common technique used for statistical-based classification. Machine learning is one of the modern application classification techniques, which uses Artificial Intelligence to identify IP traffic. Machine learning overcomes the limitations of payload-based technique [9]. Moreover, some of ML algorithms such as Support vector machines (SVM) are suitable for detecting non-Tor traffic [10]. The ML technique is performed in several steps; firstly, selection of a dataset which contains all or some of the feature values. These features are attributes of traffic flow, such as packet length, inter arrival time, protocol, idle time and other such attributes. Second comes the application of the training stage for ML to establish classification rules; this is based on statistical computation extracted from the features. Finally, ML classification is applied to unknown packets using the training rules from the second step. Due to the rapid nature of real time applications, an important issue that must be considered when classifying Internet applications is the time of collecting the statistical values (to build the rules), which is assumed to be very short. ML consists of different algorithms categorized into two main types, supervised learning and unsupervised learning.

Another method of classification is to use a hybrid. A hybrid network classifier is defined as a classifier which uses more than one classification method. Port-based, payload-based, statistical-based and hardware-based are the common methods that are used in building hybrid classifiers. Each of the classification methods has some advantages as well as some limitations. The hybrid multistage classifier makes full use of the advantages of each of the partial methods [7]. However, the disadvantage of the hybrid classifier is the complexity involved when using more than one stage. This complexity can be evaluated through the classification time versus classification accuracy. In other words, what is the trade-off between complexity and accuracy (which is expected when more than one method is used) that can be achieved

This paper proposes a hybrid online classifier (HOC) based on two of the previous methods, machine learning and port number.

1.2. Online Classification

In online classification, the decision about which packet (or flow) belongs to which particular class is based on the traffic speed. This is the same as any hardware classifier (Packet Shaper, SANGFOR [11]) which is installed on the network path in order to classify the traffic at the network speed. Therefore, the online classifier is normally installed in line with the switch/router, to identify the total traffic that passes through this device. An online classifier is very important to manage threats in the traffic, such as denial of service, flooding attacks and other similar threats. Online classification provides early flow prediction while the flow is in progress [12]. Furthermore, most of the published articles have only focused on the classifier's accuracy, with the classifier trained based on the full flow. However, this approach cannot be implemented successfully for online classification [13]. One of the problems in online classification is the high traffic speed. The challenge will be to do all the following steps in the case of high network traffic speed, i.e., (i) capturing the traffic, (ii) dividing it into flows, and (iii) calculating the statistical features or checking the payload.

Offline classification is not helpful for online management and control, mainly due to the performance [14]. Online network traffic classification is very important for several reasons:

- Online classification is the basis for managing the real time network traffic. Therefore, in order to manage and control the Internet traffic, there is a real need for online classification;
- Online classification helps to prevent network threats and abnormal behaviors, such as denial of service, flooding attacks and other such threats;
- Developing of effective software-based online classification algorithms helps to reduce the use of hardware classifiers (such as Packet-shaper) which have very high costs.

One of the goals of this paper is to differentiate between "real Internet traffic" and "online Internet traffic". There is a big difference between these two terms: real traffic can be defined as any real Internet traffic which is captured on any network level, and at the current time this is not live traffic, while online traffic means the traffic which is currently running in the network (live traffic). Figure 1 illustrates the difference between real traffic and live traffic. Real traffic is a more extended definition than live traffic. In other words, the live online traffic is a real traffic, but the real traffic is not always an online traffic. In the same manner, there are big differences between online classification and real traffic classification. Real traffic classification is the identification of the real network traffic, which can be called offline classification. This paper defines online classification as a system which can receive and classify the Internet traffic at that traffic's running time.

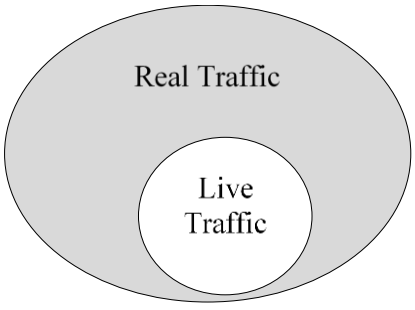

Figure 1. Real traffic and live traffic.

1.3. Hybrid-Based Classification Method

Network hybrid classifier is defined as a classifier which uses more than one classification method. Port-based, payload-based, statistical-based and hardware-based are the common methods that are used in building hybrid classifiers. Each of the classification methods has some advantages as well as some limitations. The hybrid multistage classifier

makes full use of the advantages of each of the partial methods [7]. However, the disadvantage of hybrid classifier is the complexity involved when using more than one stage. This complexity can be evaluated through the classification time versus classification accuracy. In other words, what is the trade-off between complexity and accuracy (which is expected when more than one method is used) than can be achieved?

Recent researches on network traffic classification focused on using statistical approach such as machine learning algorithm for classifying network traffic. However, the ML classification results (accuracy) becomes low over time as the application behavior changes [1].

2. Related Works

2.1. Related Internet Traffic Classification Systems

This section discusses some of the previous Internet traffic classification work. Several research papers [9,15,16] have considered the ML classifier, which was used to classify datasets in different ways, such as packet traffic features, flow traffic features, statistical packet features, etc. made a comparison between five ML algorithms (MLP, RBF, C 4.5, Bayes Net and Naïve Bayes). The authors developed a real Internet traffic dataset which included seven applications: web, e-mail, web media, P2P, FTP data, instant messaging and VoIP. In their work, they used Wireshark as the capturing tool. The results showed that, in the case of a full features dataset, Bayes Net classifier provides the best accuracy at 85.33%; when the authors applied the approach of reduced features, C4.5 provided the highest accuracy at 93.66%.

In [17] the authors proposed LASER (Longest Common Subsequence (LCS)-based Application Signature ExtRaction) technique to classify P2P traffic. The proposed method is a hybrid algorithm which investigate packet header to extract a specific signature from the payload information. Four of P2P application (Storm, Waledac, BitTorrent, eDonkey) was used to be classified by this classifier which is shows accepted accuracy results (more than 91%).

In [18], based on the analysis of P2P traffic classification technologies, a combination of packet-level classifier and flow level classifier was proposed. The first level was a deep packet inspection-based classifier which works at the packet level to identify the specific P2P traffic. The second step was a machine learning approach, which classified the remaining unknown P2P traffic at the flow level.

Shim KS et al. in [14] Proposed traffic classification method using payload size sequence signature (PSS). The proposed classifier generates unique sequence PSS for each internet application from the first N packets of the flow. The v and groups PSS vectors on the basis of the similarity between PSS vectors to identify flow patterns of the application. The proposed methods were evaluated by identifying some internet applications such as Skype, GomTV, PuTTY and other such applications.

The authors in [19] proposed an algorithm (named Skype-Hunter) to identify Skype traffic. The proposed algorithm relied on both signature-based and statistical traffic features. The experimental part of this work considered different scenarios, such as: (1) no restrictions on the transport protocols; which means using the direct connection between the Skype clients; (2) the presence of a NAT IP; this means using the IP network address translator, which is a router function that can be configured to allow the addresses of a stub-domain to be reused by any other stub-domain; (3) the presence of a firewall, which does not allow the use of UDP.

The authors of [20] used a hybrid approach to classify network traffic using the SVM and NAÏVE Bayes algorithms. The paper used a flow statistical feature to enhance feature discretization.

2.2. Online Classification Related Works

There are many published articles which include the term “online classification” in the title [7]. Unfortunately, most of them have no actual online classification but only

real traffic classification (non-live). The following paragraphs discuss some related works, which mostly includes the words "online classification" in their titles.

The study [21] proposed a hybrid traffic classification (HTC) method based on machine learning and combined with IP/ASN analysis. The packets with the same IP and port number were treated as same flow. When the packet comes in, it will check whether the flow has been classified first. After that it will find the classification result and mark the differentiated services code point (DSCP). If the flow is not classified by the first stage (encrypted flow), the proposed model is used and the second stage classification are performed. The flow classification was continuous until reaches k times. Then the majority voting method is used for calculation according to the classification results. The cooperative between IP/ASN (Autonomous System Number) and the proposed model increase the classification accuracy by 10%.

In order to maintain the accuracy of ML classifier the researchers [22] proposed a retraining mechanism. The accuracy of ML classifier is checked from time to time based on some flows which were labeled by the heuristic training dataset. The mechanism was divided into three stages. In the first stage, flows were extracted from incoming packets by the offline training dataset generator. In the second stage, the accuracy of ML classifier was evaluated against the accuracy generated by the training dataset generator. In the last stage, the online ML classifier was updated in case the current ML accuracy is below a predefined threshold.

Based on the analysis of the previous studies of hybrid classifiers, four limitations are observed: first, the proposed hybrid classifier may not consider the classification time as complexity factor. Second, most of the previous hybrid classifiers did not consider the online classification. Third, the hybrid classifiers are based on hardware components such as network processors which carry additional costs. Last, the stages of the hybrid classifier may build based only on one classification method. Table 1 summarizes some hybrid classifiers proposed by previous studies. The answer "yes" means the work considered the issue, whilst the answer "no" means the work did not consider the issue.

Table 1. Some of hybrid classifier in researcher works.

Works	Hybrid Classifier Stages	Does the Work Considers or Discusses the Complexity?	Does the Hybrid Classifier Consider Online Classification?	Does the Classifier Did Not Builds Based on Any Hardware Components?
[1]	Three steps classifier: min-max method, random forest, and Support Vector Machine (SVM)	yes	no	yes
[23]	A hybrid Radial Basis Function Network (RBFN)	no	no	yes
[19]	Two stages (signatures-based and statistical procedures)	no	no	yes
[7]	Three stages (port-based, deep packet inspection-based, and statistical-based)	no	no	yes
[17]	Hybrid classifier used header and payload information	no	no	yes
[21]	Hybrid classifier with two stages: IP/ASN analysis and proposed model	no	yes	yes
[24]	Two-stage traffic classification for P2P traffic: filtering mechanism and machine learning	no	no	yes
[25]	Three stages (port-based, static payload signatures, DPI)	no	no	yes

Table 1. *Cont.*

Works	Hybrid Classifier Stages	Does the Work Considers or Discusses the Complexity?	Does the Hybrid Classifier Consider Online Classification?	Does the Classifier Did Not Builds Based on Any Hardware Components?
[26]	Three stages (three associative classification algorithms (CBA, CMAR, and CPAR))	yes	no	yes
[27]	Three stages (defined by N stages)	no	no	yes
[28]	Classifier based on three mature technologies: Gaussian mixture model (GMM), hidden Markov model (HMM) and deep neural network (DNN)	yes	no	yes
[29]	Three stages (inspect the input packets,) sort application ID, and execute memory lookup)	no	no	No
[30]	Two stages (Signature-based and statistical-based)	no	no	yes

3. Methodology of the Proposed HOC

This paper proposes enhanced hybrid online Internet traffic classification model based on the machine learning technique and port numbers. Enhanced means it can help to solve the limitations mentioned above. This classifier differs from others, since the classification decision is based on two different parallel hybrid methods. In addition, this classifier is not based on hardware components and does not make the classification based only on port-based method. Figure 2 explains in a simple way the ML classification system. The training stage is the main input and the classification result is the output of this system. If the input is valid, this means the output should be valid. On the other hand, the use of port numbers in classification was still helpful and it can be relevant for certain types of Internet application traffic [31]. Based on some priority rules, HOC makes classification decisions for each flow.

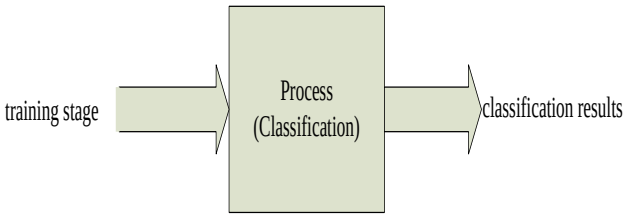

Figure 2. Classification system stages.

3.1. HOC Architecture

Figure 3 illustrates the network environment of the proposed classifier, which shows that the access point (Wi-Fi) received the Internet and distributed to the surrounding devices. This AP can be connected to different types of networks. The devices (mobile and computers) are able to access to the Internet through the AP. The volunteers that used these devices only generated the considered Internet applications. The proposed hybrid online classifier (HOC) was connected directed to the targeted AP. All the Internet traffic passing through this AP will be classified by our hybrid classifier HOC.

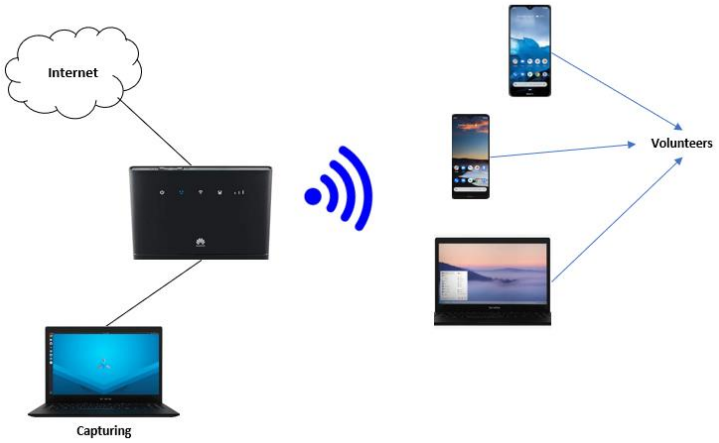

Figure 3. Network environment of proposed system.

Figure 4 illustrates the classifier stages, which start with full packets captured using a traffic mirror. Before being delivered to the two classifiers, the traffic was divided into flows based on the 5-tuples. Each flow will be classified two times by each of the classifier. The port classifier compares the captured flow port with a list of saved port numbers. If the captured flow belongs to any group of saved ports, then it will identify as its group.

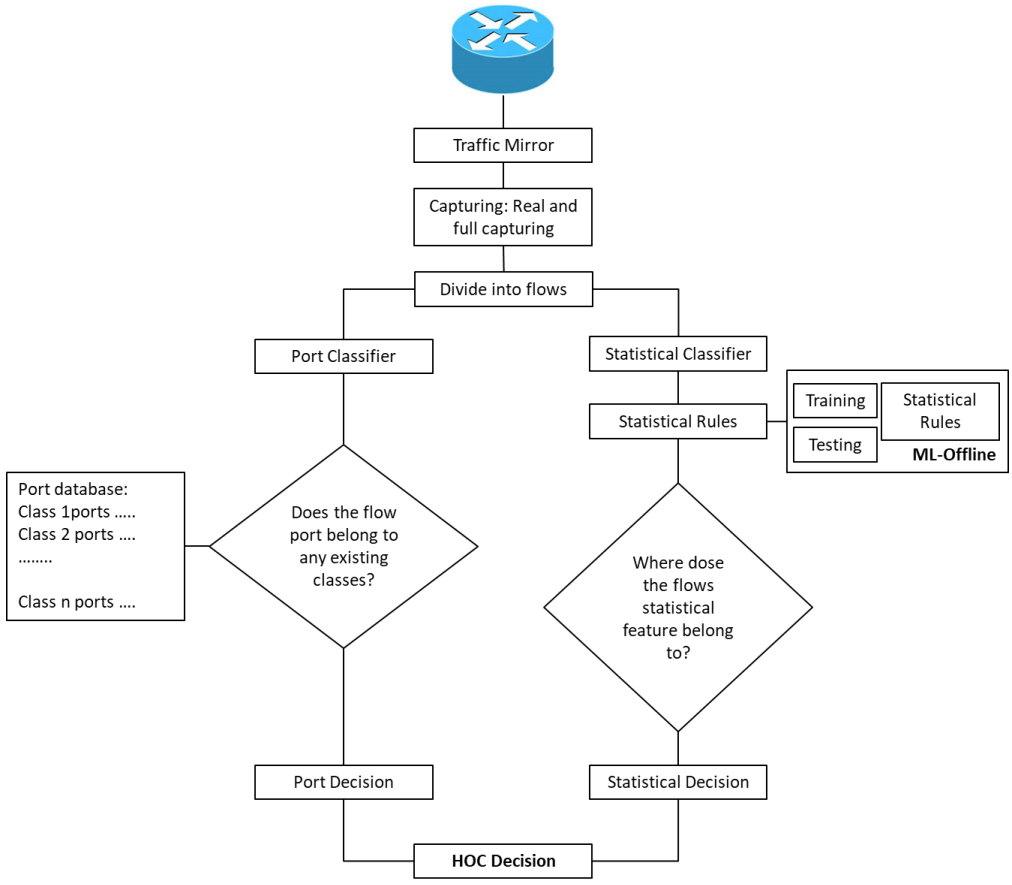

Figure 4. HOC architecture.

3.2. Port Partial Classifier

The port-based classifier is part of the HOC classification system which is entirely based on a port number. The advantage of port classification is that the identification speed is faster, but the accuracy is poor [21]. In addition, port classification methods are still relevant for certain type of Internet traffic [31] The port classifier of the HOC system

checks the flow port number. If it is similar to any number that exists in the saved database, it will be classified as equal to the class of that port number. Since the port number is numerical, the checking process is very fast and no delay is observed. Table 2 shows list of port numbers used by HOC. The top of the list illustrates some of protocol port numbers (FTP, SSH, DNS, SNMP) which collected from IANA [32]. The bottom of the list includes some port numbers which identified by extensive analysis of real captured of internet traffic in campus environment.

Table 2. Port numbers of some application classes.

Applications/Protocol (Class)	Port Numbers
FTP (data and control)	21, 21
Secure Shell (SSH) Secure Login	22
Domain Name System (DNS) service	53
Simple Network Management Protocol (SNMP)	161
www (http, https)	80, 443,
YouTube	64,021, 50,436, 50,478
WhatsApp	53,306, 53,107, 5228,5223
Twitter	56,137, 56,118, 56,117
Blackboard	52,488, 51,969, 51,894

3.3. ML Partial Classifier

The second classifier machine learning (ML) (statistical classifier) works parallel with the first classifier. Based on offline training and testing datasets, some classification rules were built. Based on these rules, the statistical classifier (algorithm) makes its online decisions to identify the captured flows. The ML classifier, plays an essential role in the HOC system. Unlike port, the statistical classifier makes a classification decision about the traffic flow in most cases. This means the port classifier will ignore the flow if there are no well-known port numbers or the traffic is encrypted, whereas the statistical classifier will automatically make the classification decision without constraint. The statistical classifier in an HOC system considers the following factors, which can help improve the ML classification quality.

The statistical algorithm imports the statistical rules from the offline training stage and uses these rules to check which class this traffic flow belongs to. In the training stage, more than 10 Weka algorithms were tested. These included: class selection, features selection, algorithms selection and building the classification rules. The statistical classifier in the HOC system is a trade-off between the rules generated by one of the three algorithms, i.e., rules.PART, Tree.J48 and RandomTree. The rules generated by the algorithm in the offline training stage are used for the classification (offline and online). After wide analysis, the rules of the PART algorithm were used (this algorithm is also used in [33]). This algorithm generates four times fewer rules than the other algorithms. In addition, the accuracy gained by the rules of this algorithm is high.

3.4. HOC Classification Decision

In the proposed hybrid online classifier (HOC), each of the two classifiers will individually classify the same traffic flow. Based on some priority rules, HOC makes classification decisions for each flow.

Table 3 shows the order of HOC priority rules. The (✓) symbol means the classifier has made a decision about the flow traffic, whilst the (×) symbol means the classifier has made no decision about this traffic flow or it classifies the flow as unknown. In the first rule, the HC decision is “unknown”, because none of the classifiers have made a decision about the current traffic flow. In the second rule, HC classifies the flow as class A when both ML and

port classifiers classify the flow in the same class (class A). In the third rule, the current flow is identified as class A by the port classifier and class B by the ML classifier. In this case, the HC decision is equal to the ML classifier (class B). In the last rule, HC classifies the flow as class A (based on the port classifier) when the ML classifiers have made no decision about this flow.

Table 3. HOC priority rules order.

HOC Priorities Order	Port Classifier	ML Classifier
1. Unknown	×	×
2. Port or Statistic classifier (class A)	✓ (class A)	✓ (class A)
3. Statistic classifier (class B)	✓ (class A)	✓ (class B)
4. Port classifier (class A)	✓ (class A)	×

4. Validation and Implementation Results

The proposed HOC was tested by identifying the traffic on three different applications: WhatsApp, Twitter and http. All the datasets of these were collected from the campus environment (Umm Al-Qura University, Computing College at AlQunfudah). WhatsApp has provided end-to-end encryption by default so messages can be seen only by the sender and recipient, and no one in between. Wireshark software [33] was used to capture and analyze this traffic. The captured file contains a large number of packets, carrying information and data about the captured application. During the capture process, the traffic needed was generated and captured manually. This means all the other Internet traffic was prevented from generating traffic. Even the windows and application updates were closed.

In ML classifier, the Weka open source was used as a machine learning tool in the training stage. The CSV file, prepared in the capturing step, was used to prepare the Weka file. Based on the previous steps, the rules of the PART algorithm (Weka algorithm) were copied and saved. These rules were prepared to be used by MATLAB, which is involved in if else statements.

Table 4 illustrates the number of packet flows which were used in the training and testing stage for each of the considered application.

Table 4. Number of flows in training and testing stages.

Application	Number of Training Flows	Number of Testing Flows
http	1500	750
WhatsApp	1500	750
Twitter	1500	750

As mentioned before, the hybrid online classifier (HOC) makes decisions based on the decisions of both the ML and port classifiers.

The rules generated by the PART algorithm were used by HOC. Figure 5 illustrates the results of the Twitter traffic classification. The figure shows the accuracy of the proposed HOC compared to ML and port classifiers. As seen, the HC and ML classifier provided high accuracy (72.5%) compared to the port classifier (0%).

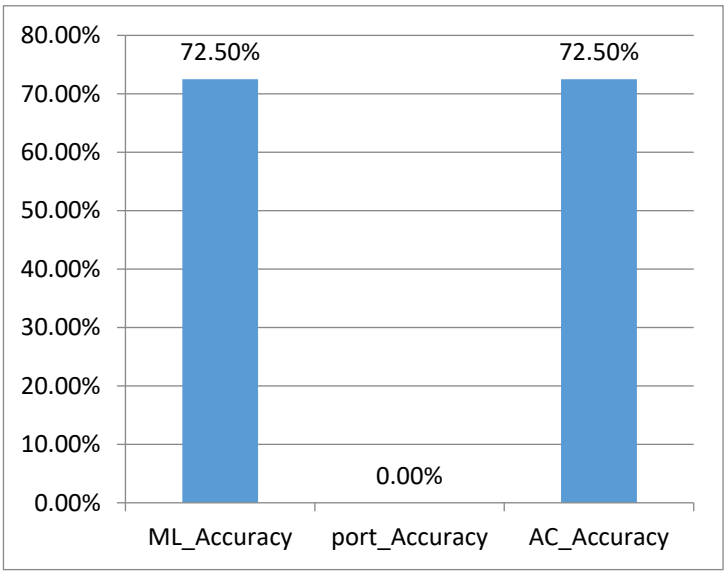

Figure 5. Twitter traffic classification results.

The testing data of http traffic was classified by the hybrid classifier. Figure 6 shows the results of http traffic classification. As shown in the figure, the hybrid classifier has a high classification accuracy (90.13%) compared to ML (84.8%) and port classifier (55.73%).

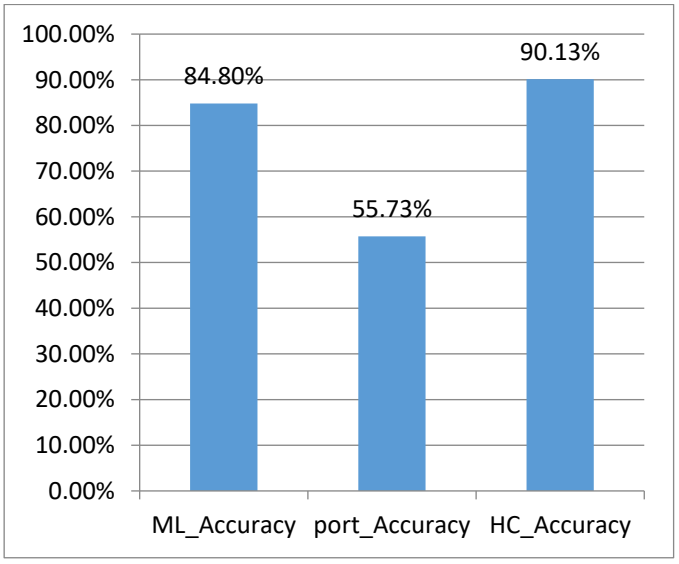

Figure 6. Http traffic classification results.

In the same way, the testing data of the WhatsApp traffic was classified by the proposed HC. Figure 7 shows the WhatsApp traffic classification results. As shown, the hybrid classifier generates a high accuracy (88.79%) when compared with the other two classifiers (85.05% and 37.9%).

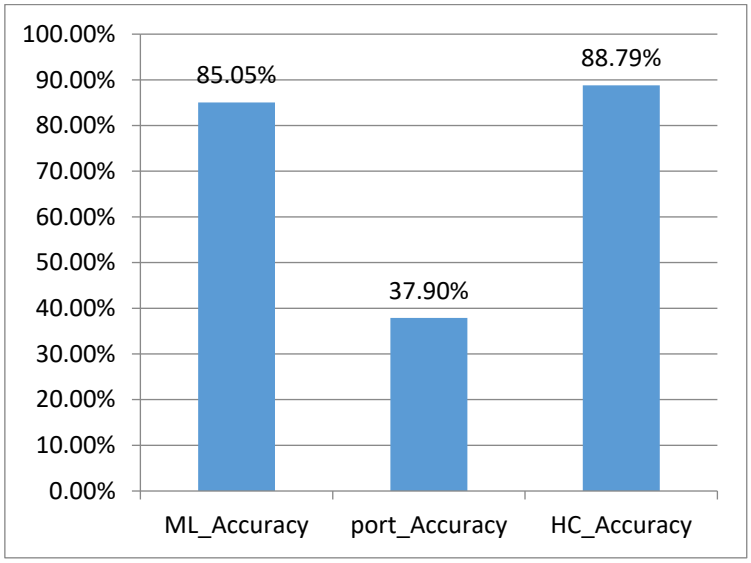

Figure 7. WhatsApp traffic classification results.

5. Conclusions

The classifier that uses only one method is constrained by the limitations of that method. Although the machine learning approach is appropriate in identifying Internet traffic and resolves the problems of classifying unknown port and encrypted traffic, this approach still has some limitations, such as features overlapping and ML dataset scenarios. The combination of more than one method is aimed at utilizing all the benefits of the individual classifiers in only one main classifier. Combining these two methods will result in a classifier that is simple and is able to identify encrypted traffic at the same time.

This paper proposed a hybrid classifier to classify Internet traffic based on two classification methods: port and statistical. The goal was to identify each packet of the Internet traffic based on their application type. The proposed hybrid classifier was tested by classifying three types of internet applications (http, WhatsApp and Twitter). Wireshark was used to capture real network traffic, which was analyzed and filtered in the manual stage. This captured file was prepared for Weka by adding the Weka header and data. More than 10 Weka algorithms were applied to train the ML classifier. The output of the training stage of statistical rules was used in the hybrid classifier. WhatsApp, http and Twitter traffic were identified using the proposed classifier. The classification showed acceptable results for each of the considered Internet application.

However, the main limitation of this work can be pointed in two issues: first, hybrid classifier complexity as well, the online classifier deal with high traffic speed. Second, the bounded network environment which provide limit live traffic (only one access point Wi-Fi data in campus network was used).

Author Contributions: Machine learning H.A.H.I., O.R.A.L.Z. and A.M.A.; methodology, H.A.H.I., O.R.A.L.Z., A.M.A. and M.B.A.; data curation, H.A.H.I., O.R.A.L.Z., A.M.A. and M.B.A.; writing—review and editing, H.A.H.I., O.R.A.L.Z. and M.B.A.; writing—original draft H.A.H.I.; writing—review and editing, H.A.H.I., O.R.A.L.Z. and M.B.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research is morally supported by College of Computer at Al-Gunfudah, Umm Al-Qura University and financially supported by the authors group.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the datasets of these research were collected from the campus environment (Umm Al-Qura University, Computing College at AlQunfudah). The datasets are available with the authors group.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mehmood, M.; Javed, T.; Nebhen, J.; Abbas, S.; Abid, R.; Bojja, G.R.; Rizwan, M. A Hybrid Approach for Network Intrusion Detection. *CMC-Comput. Mater. Contin.* **2022**, *70*, 91–107. [CrossRef]
2. Patel, B.; Somani, Z.; Ajila, S.A.; Lung, C.H. Hybrid Relabeled Model for Network Intrusion Detection. In Proceedings of the 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Halifax, NS, Canada, 30 July–3 August 2018; pp. 872–877.
3. Yamansavascilar, B.; Guvensan, M.A.; Yavuz, A.G.; Karsligil, M.E. Application identification via network traffic classification. In Proceedings of the 2017 International Conference on Computing, Networking and Communications (ICNC), Silicon Valley, CA, USA, 26–29 January 2017.
4. Kim, J.; Hwang, J.; Kim, K. High-performance internet traffic classification using a markov model and kullback-leibler divergence. *Mob. Inf. Syst.* **2016**, *2016*, 6180527. [CrossRef]
5. Ye, W.; Cho, K. P2P and P2P botnet traffic classification in two stages. *Soft Comput.* **2017**, *21*, 1315–1326. [CrossRef]
6. Lim, H.-K.; Kim, J.-B.; Hong, Y.-G.; Han, Y.-H. Payload-based traffic classification using multi-layer lstm in software defined networks. *Appl. Sci.* **2019**, *9*, 2550. [CrossRef]
7. Min, D.; Xingshu, C.; Jun, T. Online Internet traffic identification algorithm based on multistage classifier. *China Commun.* **2013**, *10*, 89–97. [CrossRef]
8. Goli, Y.D.; Ambika, R. Network traffic classification techniques-a review. In Proceedings of the 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, 21–22 December 2018; pp. 219–222.
9. Shafiq, M.; Yu, X.; Bashir, A.K.; Chaudhry, H.N.; Wang, D. A machine learning approach for feature selection traffic classification using security analysis. *J. Supercomput.* **2018**, *74*, 4867–4892. [CrossRef]
10. Hodo, E.; Bellekens, X.; Iorkyase, E.; Hamilton, A.; Tachtatzis, C.; Atkinson, R. Machine learning approach for detection of nontor traffic. In Proceedings of the 12th International Conference on Availability, Reliability and Security, Regio Callabria, Italy, 29 August 2017; pp. 1–6.
11. Sangfor 2000 [Cited 2021; Sangfor Managed Cloud Services]. Available online: https://www.sangfor.com/en (accessed on 13 December 2021).
12. Nguyen, T.T.T.; Armitage, G.; Branch, P.; Zander, S. Timely and Continuous Machine-Learning-Based Classification for In-teractive IP Traffic. *IEEE/ACM Trans. Netw.* **2012**, *20*, 1880–1894. [CrossRef]
13. Chen, Z.; Yang, B.; Chen, Y.; Abraham, A.; Grosan, C.; Peng, L. Online hybrid traffic classifier for Peer-to-Peer systems based on network processors. *Appl. Soft Comput.* **2009**, *9*, 685–694. [CrossRef]
14. Shim, K.-S.; Ham, J.-H.; Sija, B.D.; Kim, M.-S. Application traffic classification using payload size sequence signature. *Int. J. Netw. Manag.* **2017**, *27*, e1981. [CrossRef]
15. Fan, Z.; Liu, R. Investigation of machine learning based network traffic classification. In Proceedings of the 2017 International Symposium on Wireless Communication Systems (ISWCS), Bologna, Italy, 28 August 2017; pp. 1–6.
16. Anderson, B.; McGrew, D. Machine learning for encrypted malware traffic classification: Accounting for noisy labels and non-stationarity. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1723–1732.
17. Sajeev, G.P.; Nair, L.M. LASER: A novel hybrid peer to peer network traffic classification technique. In Proceedings of the 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 21–24 September 2016; pp. 1364–1370. [CrossRef]
18. Ye, W.; Cho, K. Hybrid P2P traffic classification with heuristic rules and machine learning. *Soft Comput.* **2014**, *18*, 1815–1827. [CrossRef]
19. Adami, D.; Callegari, C.; Giordano, S.; Pagano, M.; Pepe, T. Skype-Hunter: A real-time system for the detection and classification of Skype traffic. *Int. J. Commun. Syst.* **2012**, *25*, 386–403. [CrossRef]
20. Aggarwal, R.; Singh, N. A new hybrid approach for network traffic classification using SVM and Naïve Bayes algorithm. *Int. J. Comput. Sci. Mob. Comput.* **2017**, *6*, 168–174.
21. Huang, Y.-F.; Lin, C.-B.; Chung, C.-M.; Chen, C.-M. Research on QoS Classification of Network Encrypted Traffic Behavior Based on Machine Learning. *Electronics* **2021**, *10*, 1376. [CrossRef]
22. Zarei, R.; Monemi, A.; Marsono, M.N. *Retraining Mechanism for On-Line Peer-to-Peer Traffic Classification, in Intelligent Informatics*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 373–382.
23. Pradhan, A.; Behera, S.; Dash, R. Hybrid rbfn based encrypted ssh traffic classification. In Proceedings of the 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 22–23 February 2018; pp. 264–269.
24. Khan, R.U.; Kumar, R.; Alazab, M.; Zhang, X. A hybrid technique to detect botnets, based on P2P traffic similarity. In Proceedings of the 2019 Cybersecurity and Cyberforensics Conference (CCC), Melbourne, VIC, Australia, 8 May 2019; pp. 136–142.

25. Lu, W.; Ghorbani, A.A. A multiple-stage classifier for identifying unknown internet traffic. In Proceedings of the 2011 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, BC, Canada, 23 August 2011; pp. 725–729.
26. Li, L.; Kianmehr, K. Internet traffic classification based on associative classifiers. In Proceedings of the 2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), Bangkok, Thailand, 27–31 May 2012.
27. Ichino, M.; Maeda, H.; Yamashita, T.; Hoshi, K.; Komatsu, N.; Takeshita, K.; Tsujino, M.; Iwashita, M.; Yoshino, H. Internet traffic classification using score level fusion of multiple classifier. In Proceedings of the 2010 IEEE/ACIS 9th International Conference on Computer and Information Science, Yamagata, Japan, 18–20 August 2010; pp. 105–110.
28. Tan, X.; Xie, Y.; Ma, H.; Yu, S.; Hu, J. Recognizing the content types of network traffic based on a hybrid DNN-HMM model. *J. Netw. Comput. Appl.* **2019**, *142*, 51–62. [CrossRef]
29. Liu, Y.; Xu, D.; Mu, Z.; Qin, J. Efficient hybrid packet classification in traffic control system using network processors. In Proceedings of the 2009 International Conference on Advanced Computer Control, Singapore, 22–24 January 2009; pp. 57–61.
30. Sun, G.L.; Xue, Y.; Dong, Y.; Wang, D.; Li, C. An novel hybrid method for effectively classifying encrypted traffic. In Proceedings of the 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, Miami, FL, USA, 6 December 2010; pp. 1–5.
31. Aouini, Z.; Kortebi, A.; Ghamri-Doudane, Y.; Cherif, I.L. Early classification of residential networks traffic using C5. 0 machine learning algorithm. In Proceedings of the 2018 Wireless Days (WD), Dubai, United Arab Emirates, 3 April 2018; pp. 46–53.
32. Cotton, M.; Eggert, L.; Touch, J.; Westerlund, M.; Cheshire, S. Internet Assigned Numbers Authority (IANA) Procedures for the Management of the Service Name and Transport Protocol Port Number Registry. *RFC* **2011**, *6335*, 1–33.
33. Orebaugh, A.; Ramirez, G.; Beale, J. *Wireshark & Ethereal Network Protocol Analyzer Toolkit*; Elsevier: Amsterdam, The Netherlands, 2006.