# Online Traffic Classification Using Granules

Pingping Tang, Yuning Dong
*College of Telecommunications and Information Engineering*
*Nanjing University of Posts and Telecommunications*
Nanjing, China
tpping@ahnu.edu.cn,dongyn@njupt.edu.cn

Shiwen Mao
*dept. of Electrical and Computer Engineering*
*Auburn University*
Auburn, USA
smao@ieee.org

***Abstract*—Currently, it is still a great challenge to achieve online classification of massive traffic flows under dynamic network environments. Therefore, based on granular computing, an artificial intelligence computing method, which is effective to process missing, incomplete, or noisy data, a novel classification model $\mathbf{M}_{GrC}$ is proposed in this paper. In $\mathbf{M}_{GrC}$, we first define granules for the traffic flow, then explore the correlation between granules, and finally establish the structure granules to differentiate flow types. $\mathbf{M}_{GrC}$ explores the inherent relationship between packets, where the data is no longer isolated, but closely related to each other. So, it can identify the traffic more accurately when compared with the traditional classification methods, which assume the packets to be independent. The experiment results also demonstrate its superior robustness and adaptability in highly variable network environment.**

***Index Terms*—Massive traffic, on-line, flow classification, granular computing, artificial intelligence computing**

## I. Introduction

With the development of 5G technologies, the network traffic is growing rapidly on a tremendous scale [1]. Cisco forecasts that 5G traffic will be three times the 4G traffic in 2022 and annual traffic will reach 930EB (1EB=1 billion GB) [2]. Different traffic has varying requirements for quality of service (QoS) and network resources. For example, video conferencing and telemedicine applications strictly require good real-time performance, and any unexpected delay can result in a wrong decision and cause considerable economic loss. High-quality video streaming requires substantial network bandwidth to provide a good user experience, and thus Internet service providers (ISPs) are expected to allocate suitable network resources for different traffic. In addition, analysis of the distribution of traffic types can enhance the controllability, supervision and management of the network [3], [4]. In openflow, one of the first SDN frameworks, Google clearly points out that traffic classification will become the basis of network behaviors, and all network operations such as management, scheduling and resource allocation will start from traffic classification [5].

### A. Motivation and Challenges

Due to the wide use of encryption technologies, the captured data payload is usually encrypted, preventing developers from gaining useful information from the payload to achieve precise classification [6]. So, the existing classification methods mainly identify traffic by analyzing the flow direction, protocol and conversation pattern combining with statistical features (e.g., packet size, duration, etc.). For example, Kornycky *et al.* [7] focused on classifying video streaming, FTP, email, and Skype voice flows. In their method, according to different protocols, the system calculates the number of packets, mean packet inter-arrival time, etc., for downlink and uplink flows, respectively, and then clusters the flows based on the distance given by

$$\|F_a - F_b\|_2 = \sqrt{\sum_n |F_{a\_n} - F_{b\_n}|^2} \tag{1}$$

where $F_a$ and $F_b$ refer to two flows, respectively, and $F_{x\_n}$ denotes the $n$th feature of flow $F_x$, $x \in \{a, b\}$. For example, Sun *et al.* [8] classified videos into live streaming and VOD based on the statistical features of packet length, number of packets, information offset, etc. Based on the number of peer connection in both incoming and outgoing direction for 5-minute duration, Thay *et al.* [9] provided a classification technique to classify the P2P traffic, including BitTorrent, Skype, SopCast, etc. However, when applied to online classification [10], the limitations of the above methods are gradually identified, which include:

1) Most of the statistical features $F_{x\_n}$ (e.g., duration, defined as the time between the first and last packets) are not amenable for online classification, since they can only be estimated after the end of the flow, which is a serious constraint on the real-time performance of online classification [11].

2) When congestion occurs in the network, time delay and packet losses usually increase drastically, which may result in significant errors in $F_{x\_n}$ (e.g., the number of packets, packet inter-arrival time, etc.). Consequently, the classification accuracy will drop dramatically [12]. Besides, some classification schemes use the key packets from the first few seconds of the flow to achieve online classification, but they cannot obtain the expected classification results if the key packets are lost [13].

3) The statistical features assume that the continuously arriving packets are independent, but these packets are highly correlated rather than being independent [14]. It would be desirable to explore the correlation among packets and to establish the ground truth for effective classification [15].

### B. Human's ability on online classification

So far, we have been searching for the extremely precise solution to problems. By carefully analyzing the nature of human classification behavior, a question to ask is how much

of the precise calculation is included in our classification behavior? ! For example, it's so easy for a common child to recognize apple and pear, whether it's an apple with a large part bitten off or a pear with a big green worm lying on it. Moreover, victims can identify the criminal at the first glance from various photos; old friends who have been apart for decades may still recognize each other. How do human beings achieve it? Is it intuition? Or inspiration? Whatever it is, it is certainly not precise calculation. We always seek the precise solution for problems, which may be the wrong direction!

A group of scientists, e.g., A. L. Zadeh, J. R. Hobss, T. Y. Lin, explored and analyzed the manner of human thinking and learning and proposed a new classification mechanism called *granular computing* [16]–[19]. By studying the process of human recognition of objects, Zadeh *et al.* [16] found that human divided the object into granules during the analysis. These granules are not independent of each other, but closely related. Pal *et al.* [18] also pointed out that the contents of information that human observe, measure, define, and reason are all granules. Human beings analyze issues in different granularities, and shuttle up and down in these granularities to achieve a synthetic diagnosis, which help generate a comprehensive, reasonable, and correct conclusion. This kind of learning pattern is quite different from deep learning, logical reasoning, etc. It reasons and analyzes the relationships between data, which can filter out interference and noise (such as the big bug on the pear), handle missing or incomplete data (such as the apple with a large part bitten off) etc. [19].

In dynamic networks, especially in an adversarial network environment, problems such as loss, retransmission, and disorder of packets may occur at any time. Thus there is a compelling need of a new technology to deal with missing, incomplete, or noisy traffic data for online classification. So we introduce granular computing in this paper to overcome the limitations of the existing classification methods.

### C. Contributions

The major contributions of this paper are summarized as follows.

- The concept of *flow granules* is put forward for the first time. The flow granules are generated by aggregating similar neighborhood packets. Thus, the information to be processed is no longer single, individual packet, but aggregated packets. As a result, $\mathrm{M}_{GrC}$ can effectively deal with the issues of missing data and incomplete information.
- Structure granule $\boldsymbol{\alpha}$, a novel flow feature, is initially defined in this paper. It indicates the inherent relationship between packets, and thus can be utilized to identify the traffic more accurately compared with the statistical features that assume the packets are independent.

In general, our proposed classification method $\mathrm{M}_{GrC}$ overcomes the restrictions of existing classification schemes and can achieve robust and accurate online classification.

## II. $\mathrm{M}_{GrC}$: Classification Model Based on Granular Computing

According to the researches in [16]–[19], granular computing basically has three steps: 1) define granules; 2) explore the relationships between granules; 3) analyze their relationship and obtain results. In this section, the classification model $\mathrm{M}_{GrC}$ will be established in accordance with these three steps. First, we define flow granules for the traffic in Section II-A, then explore the relationships between granules, establish the structure granule $\boldsymbol{\alpha}$ in Section II-B, and thus, finally achieve classification of traffic according to $\boldsymbol{\alpha}$ in Section II-C.

### A. Flow Granules

There is no doubt that the definition of granules is very important for granular computing. Pal *et al.* proposed different types of granules, such as crisp granules, fuzzy granules, and neighborhood granules, etc. [18]. Crisp granules, which satisfy the orthogonal relation, can greatly improve the computation speed of classification. However, there is not always a clear boundary between information, and in reality, the information may be overlapping and interweaving. In comparison, the fuzzy granules are closer to the nature of things. However, when fuzzy granules are used for classification, it is very important to establish the membership function, which is usually difficult to establish [20]. Thus, the neighborhood granules are proposed to analyze the correlation of information. Overlapping is an inherent characteristic of this type of granules. Therefore, in this paper, we define the granules for flow sequence according to the concept of neighborhood granules.

Before proceeding, we first provide an accurate definition of flow as follows. Traffic is composed of flows, and the flows aggregate into traffic. Then, flow is defined as a set of packets with the same five-tuple: $\{SrcIP, DestIP, SrcPort, DestPort, Protocol\}$, and the flow sequence is described as

$$F \triangleq \{(P_i, T_i)\,|_{i=1,2,\ldots,n}\} \tag{2}$$

where $P_i$ refers to the size of the $i$th packet, $T_i$ represents the inter-arrival time between $i$th packet and the previous packet, and $n$ is the number of packets in flow $F$. $F$ can be divided into several subflows, and the $m$th subflow $F^{(m)}$ is given by

$$F^{(m)} = \{(P_i, T_i)\,|_{i=p+1,p+2,\ldots,p+n_m-1}\} \tag{3}$$

$$s.t. \quad p = \sum_{i=1}^{m-1} n_i \tag{4}$$

where $n_m$ is the number of packets in the $m$th subflow. Based on the formation mechanism of neighborhood granules by Pal [18], we define the flow granules as

$$\aleph(x) = \bigcup_{k=i}^{j} P_i \in U \tag{5}$$

$$s.t. \quad |P_i - P_{i+1}| < Thr_v. \tag{6}$$

It can be seen from (6) that if the neighborhood packets have a similar packet size, they will be aggregated into the same granule $\aleph(\cdot)$. The flow sequence in (2) or (3) would be processed in this way, and thus we obtain $\aleph(x)|_{x=1,2,\cdots,X}$, where $X$ is the size of the flow granules. Note that the members in granule $\aleph(\cdot)$ are the similar neighborhood packets, so the calculation model $\mathrm{M}_{GrC}$ is less sensitive to missing data and can remove the noisy data as well, which is one of the basic ideas of granular computing.

### *B. Structure Granules*

Zadeh *et al.* [21] found that human analyze issues from various perspectives, and can shuttle up and down at these perspectives to make a synthetic diagnosis. Coincidentally, Mandelbrot also used the concept of scale to study the traits of things [22]. Suppose $\{F(t)\}$ to be a stochastic process, and if the measurement $\mu(\varepsilon)$ and the observation scale $\varepsilon$ satisfy

$$\mu(\varepsilon) \propto \varepsilon^{\alpha}. \tag{7}$$

Then $\alpha$ can be regarded as a feature when $\{F(t)\}$ are observed on scale $\varepsilon$. $\alpha$, called the *holder index* or *singularity index*, has been widely used in prediction of gas emission in mines, classification of hydrological and water resources, anti-interference treatment of artificial scenes, etc. [23].

According to (2) and (3), flows satisfy the definition of $\{F(t)|_{(t=i)}\}$ proposed by Mandelbrot. Therefore, based on (7), we establish the relationship between flow granules and thus establish the structure granules as

$$\boldsymbol{\alpha} \triangleq \left\{ \frac{1}{m} ln\tau_m \,|_{m=1,2,\ldots,X} \right\} \tag{8}$$

$$s.t. \quad \tau_m \triangleq \sum_{k=1}^{\frac{Z}{m}} \left| \sum_{i=1}^{m} \bar{\aleph}\left(m\left(k-1\right)+l\right) \right|^2, \tag{9}$$

where $\bar{\aleph}(\cdot)$ is the average of members in the flow granules. $m$ refers to the observation scale. $\boldsymbol{\alpha}$ shows the changing relationships between flow granules when the observation scale $m$ changes from 1 to $X$.

*Proposition 1:* Structure granule $\boldsymbol{\alpha}$ uniquely identifies the type of the network flow.

*proof 1:* Suppose there are flows $F_a$ and $F_b$. $\boldsymbol{\alpha_a}$ and $\boldsymbol{\alpha_b}$ are their corresponding structure granules. Then, we compute $\boldsymbol{\alpha_z}$ of the aggregated flow $F_Z = F_a + F_b$. According to the theory proposed by Mandelbrot, $\varepsilon$ in (7) is a continuous variable, so (8) can be obtained by sampling the observation scale $\varepsilon$ as:

$$\boldsymbol{\alpha} = \{\alpha|_{ln\varepsilon=m}\} \triangleq \left\{ \lim_{ln\varepsilon \to m} \frac{ln\mu(\varepsilon)}{ln\varepsilon} \right\}, \tag{10}$$

where $\alpha$ refers to a continuous variable, and $\boldsymbol{\alpha}$ is a vector. The members of $\boldsymbol{\alpha}$ are sampled from $\alpha$. From (7), we have $\mu_a(\varepsilon) \propto \varepsilon^{-\alpha_a}$, $\mu_b(\varepsilon) \propto \varepsilon^{-\alpha_b}$, and then

$$\alpha_z = \lim_{\varepsilon \to m} \frac{ln\left(\mu_a(\varepsilon) + \mu_b(\varepsilon)\right)}{ln\varepsilon}. \tag{11}$$

Hence we can deduce the boundaries of $\alpha_z$ as:

$$\inf(\alpha_z) = \lim_{\varepsilon \to m} \frac{\ln\sqrt{2\mu_a(\varepsilon)\,\mu_b(\varepsilon)}}{ln\varepsilon} = \frac{1}{2}(\alpha_a + \alpha_b); \tag{12}$$

$$\sup(\alpha_z) = \lim_{\varepsilon \to m} \frac{2\max\left(\mu_a(\varepsilon), \mu_b(\varepsilon)\right)}{ln\varepsilon} = \max(\alpha_a, \alpha_b). \tag{13}$$

In particular, when $\alpha_a$=$\alpha_b$=$\alpha$, we have $\inf(\alpha_z) = \sup(\alpha_z) = \alpha$, which indicates that, if flow $F_a$ belongs to the same class as flow $F_b$ , then the aggregated flow $F_Z = F_a + F_b$ will fall in the same class. If flows $F_a$ and $F_b$ belong to different classes, the $\alpha$ of the aggregated flow $F_Z$ would be neither $\alpha_a$ nor $\alpha_b$. Therefore, we prove that the vector $\boldsymbol{\alpha}$, samples of $\alpha$ with different scale $m$ as in (10), is unique. That is, $\boldsymbol{\alpha}$ can uniquely identify the type of network flows.

### *C. Calculating Differences between Structure Granules*

Structure granule $\boldsymbol{\alpha}$ describes the trajectory of bursty data at different observation scales. For a certain type of flows, they always follow a specific communication protocol and transmission pattern, so that they have similar variations reflecting the inherent traits. Due to this reason, statistical features (e.g., mean packet size, maximum and minimum packets) are used to distinguish different flows. However, as described in Section II-B, these statistical features assume the packets to be independent. In fact, they are closely related to each other. $\boldsymbol{\alpha}$ reflects the correlation between granules. The inherent relationship between packets indicated by $\boldsymbol{\alpha}$ can be utilized to identify the traffic more accurately. Therefore, based on the gray correlation, which is generally used to quantitatively measure the similarity between curves [24], we define the difference between $\boldsymbol{\alpha_a}$ and $\boldsymbol{\alpha_b}$ as

$$Dif(\boldsymbol{\alpha_a}, \boldsymbol{\alpha_b}) \triangleq 1 - \frac{2 \cdot \boldsymbol{\alpha_a}\boldsymbol{\alpha_b}^{\mathrm{T}}}{\boldsymbol{\alpha_a}\boldsymbol{\alpha_a}^{\mathrm{T}} + \boldsymbol{\alpha_b}\boldsymbol{\alpha_b}^{\mathrm{T}}}. \tag{14}$$

where $\boldsymbol{\alpha_a}$ and $\boldsymbol{\alpha_b}$ refer to the structure granules of flows $F_a$ and $F_b$ respectively. According to (14), we have $Dif(\boldsymbol{\alpha_a}, \boldsymbol{\alpha_b})$=$Dif(\boldsymbol{\alpha_b}, \boldsymbol{\alpha_a})$, and $Dif(\cdot)$ is between 0 and 1. The smaller the value of $Dif(\cdot)$, the smaller the difference, and the higher the similarity. In the extreme case, $Dif(\boldsymbol{\alpha_a}, \boldsymbol{\alpha_a})$=0, which means there is no difference between the two vectors.

Suppose there are $L$ classes $\{M_l\}_{l=1}^{L}$, and several flows $\{\cdots, F_j, F_k, \cdots\}$ in each class. The centers of classes are $\{P_l\}_{l=1}^{L}$. Note that $Dif(\cdot)$ changes with uniform distribution between 0 and 1. Therefore, the center $P_l$ is determined by

$$P_l \triangleq \min_{F_k \in \mathrm{M}_l} \{ \max_{j \neq k, F_j \in \mathrm{M}_l} Dif(\boldsymbol{\alpha_k}, \boldsymbol{\alpha_j}) \}. \tag{15}$$

How to judge whether flow $F_a$ belongs to $M_l$? Just calculate the difference degree between flow $F_a$ and the class center: $Dif(\boldsymbol{\alpha_a}, \boldsymbol{\alpha_{P_l}})$. If the difference degree is less than or equal to the threshold, then $F_a$ belongs to class $M_l$; otherwise $F_a$ does not belong to class $M_l$:

$$\mathrm{Be}(Fa, M_l) \triangleq \begin{cases} \in, if\,\{Dif(\boldsymbol{\alpha_a}, \boldsymbol{\alpha_{P_l}}) \leq T\} \\ \notin, if\{Dif(\boldsymbol{\alpha_a}, \boldsymbol{\alpha_{P_l}}) > T\} \end{cases}. \tag{16}$$

TABLE I
DATA SET CONSTRUCTION

| Day | Date | Number of flows | | | | | |
|---|---|---|---|---|---|---|---|
| | | IM | Video | Audio | WB | FTP | Email |
| 1 | Mar 12 | 5814 | 4643 | 3596 | 5725 | 2828 | 3026 |
| 2 | Apr 24 | 4791 | 5714 | 3611 | 4869 | 2634 | 2521 |
| 3 | May 16 | 4532 | 5695 | 4718 | 6345 | 2592 | 2977 |
| 4 | Jun 23 | 6815 | 6827 | 3795 | 5147 | 3423 | 2514 |

## III. PERFORMANCE EVALUATION

The traces were captured by Wireshark at a ISP of a leading Chinese network service provider located in City A in southern China (Due to the commercial secret issue, the names are omitted). As shown in Table I, taking into account the variability of networks, the data was collected in four different time periods during Mar. 2018 to Jun. 2018. Then, they were preprocessed by Linux shell scripts into fivetuple of flow sequence as described in Section II-A. The data set included six types of traffic flows: Instant message (IM), Video, audio, web browsing (WB), FTP and Email. For each IM flow, traffic was generated by sending ten messages at random times. Video flows were generated by starting a randomly selected video from the five video services, including PPlive, PPStream, TVant, UUSee, and LETV. For audio, a phone call using QQ or Wechat was made to a client that automatically accepted the call. WB flows were generated by having the laptop visit six different websites. Applications were downloaded from the campus network service center to obtain FTP flows. Emails from three mailbox servers (Sina, QQ, and AHNU) were used.

### A. Evaluating the Structure Granules of Flows

In this subsection, we use video flows generated by application iQIYI to demonstrate how to obtain the structure granule $\boldsymbol{\alpha}$. The resolution of flow sequence is set to $N = 5000$, which means there are 5000 packets in this subflow. According to (2), this video flow can be described as $F$={(54, 0.388139), (32, 0.389026), (268, 0.397229), $\cdots$, (1286, 371.084922), (1286, 371.085084), (294, 371.085922)}. The following two steps are executed in this experiment.

(*Step i*): Scanning the flow sequence to obtain the flow granules. According to (5)–(6), the members of $F$ are aggregated to form the flow granules:

$$\aleph_v(x) = \{\{54, 32\}, \{268\}, \cdots, \{1286, 1286\}, \{294\}\}$$

(*Step ii*): Observing the above flow granules at various scales to form structure granules $\boldsymbol{\alpha}$. For different observation scales $m = 1, 2, \cdots, \lceil logN \rceil$, $\boldsymbol{\alpha}$ can be obtained by (8)–(9):

$$\boldsymbol{\alpha} = \{26.614, 23.537, 22.183, 21.562, 21.242, 20.980, 20.877, 20.746, 20.628, 20.212\}$$

Besides, $\boldsymbol{\alpha}$ of IM, audio, WB, FTP and Email is also plotted in Fig. 1. We can see that flows have different characteristic of $\boldsymbol{\alpha}$.

### B. Performance of Classification

Precision, recall, and F1-score are commonly used to measure the accuracy of classification models in statistics [25]. In this paper, we also use these three metrics to evaluate the traffic classification performance.

In this experiment, 3000 flows are randomly selected from the data sets for training or testing, including video, audio, web browsing (WB), text communication (TC), FTP and email, with 500 flows for each class. We compare several state-of-the-art schemes, including Fractals [14], CHS [3], SFNN [7], and CPRF [10].

Firstly, the training and testing data were taken from the same day. As shown in Fig. 2, there is a definite consistency of performance among the F1 values for schemes $\mathrm{M}_{GrC}$, Fractals and CHS. When used with our chosen selection of a subset of the original cp features (flow duration, size of largest packet, etc.), CPRF cannot classify some of the flows that have the similar duration, number of packets, etc. However, CPRF can rapidly distinguish video flows from non-video flows using only the initial 20 packets of a flow [10]. The average F1 of SFNN is around 0.9, and that of CHS is slightly higher. CHS has combined several base classifiers, and thus achieves a superior performance and higher accuracy than SFNN, which just contains a single classifier. The average F1 of Fractals also reaches as high as 0.9. The fractal characteristics are different from the statistical features in that they capture the characteristics of traffic transmission, and thus they can classify accurately for different flows. However, the Fractals method is not suitable for online classification, because of the following reasons: (1) The process to obtain fractal characteristics is complex and time-consuming; (2) In stable network environments, the characteristics of traffic burst are stable, and thus can be used to differentiate the classes of flows. But in dynamic network environments, especially when congestion occurs, the fractal characteristics are variable due to the changing bursts of traffic, resulting in a sharp decline in classification performance as will be shown in Section III-C.

When training and testing on different days, $\mathrm{M}_{GrC}$ shows better performance than all other schemes. The proposed $\mathrm{M}_{GrC}$ exploits the correlation information between packets to implement classification. It can effectively deal with noisy data under dynamic network environment, and thus work well for each day of traffic. The highest F1 is 96.64%, and the average F1 achieves 95.86%. Even the worst F1-score is also above 95%.

### C. Adaptability to Dynamic Flows

In Section I-A, most of the schemes work well in stable network environment. So in this subsection, we simulate the bad network environment to further test their classification performance. In order to simulate the dynamic network environment, we make random adjustments of loss and delay of packets. We also make further modification, and add some packets to each flow to simulate the noisy data. The amount of changed packets is within 5%, and the intensity of change

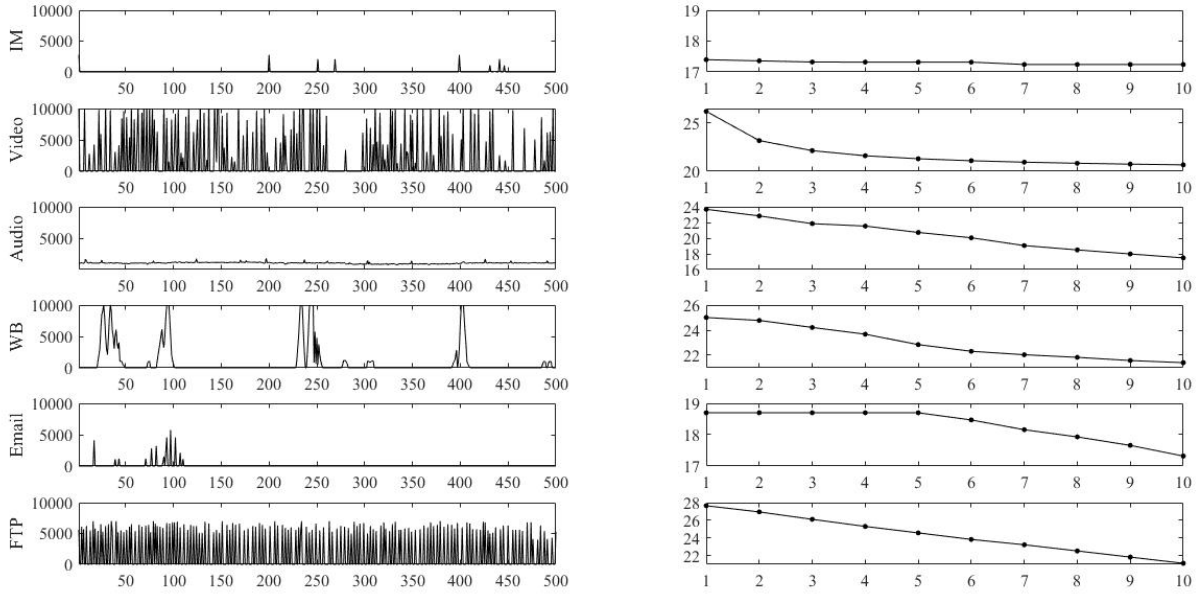


Fig. 1. Characteristics $\alpha$ of different flows.

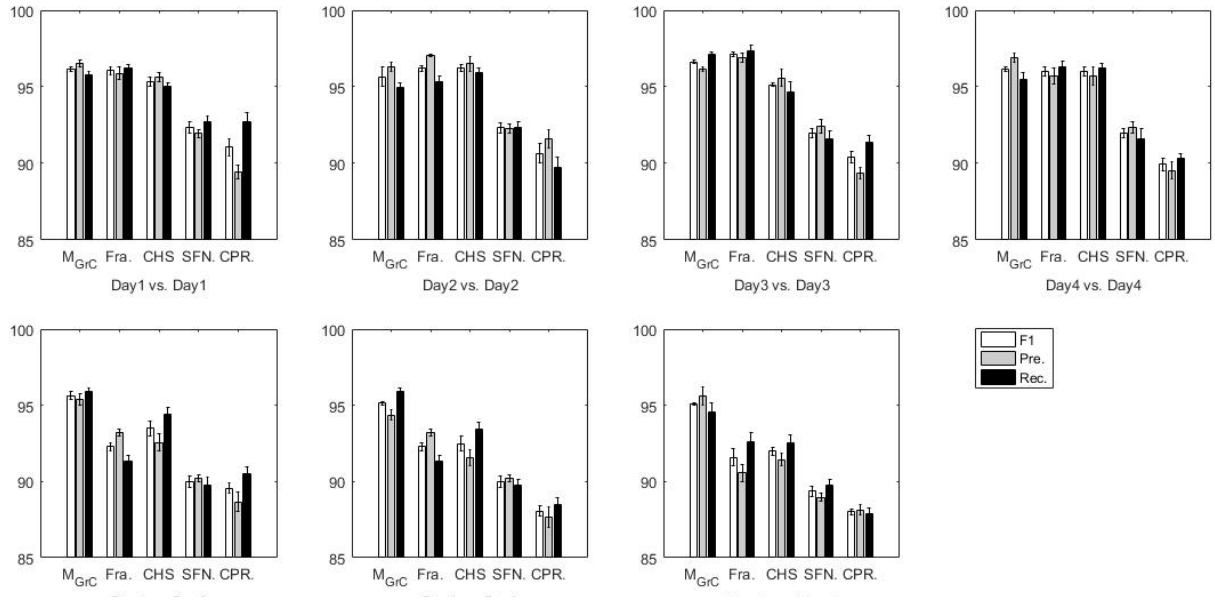


Fig. 2. Classification results.

is also controlled within 5%. Then we use these flows to test the adaptability of classification methods.

As shown in Fig. 3, the F1-scores of CHS, SFNN and Fractals present obvious decrease. These extracted statistical features and fractal characteristics, which are obtained in a friendly network environment, does not work well in an adversarial network environment. Take Video flow as an example. In a good network environment, the fractal characteristics $h(q)$ ($q$=1,2,3,4,5), referring to the degree of traffic transmission volume, are 0.356, 0.389, 0.495, 0.562, and 0.656, respectively. While in the bad network environment, they are decreased to 0.301, 0.387, 0.429, 0.522, and 0.584, respectively. Actually, the fractal characteristics $h(q)$ for the flows are always varying under different network environments, which results in unstable classification results.

In contrast, our scheme based on granular computing can achieve a better performance. Fig. 3 shows that the F1-scores of the proposed scheme are above 0.8, consistently higher than the scores of other baseline methods. $\mathrm{M}_{GrC}$ analyzes deep into the trajectory of change for different flows, and can effectively deal with noisy and missing data. Therefore, $\mathrm{M}_{GrC}$ is more suitable and robust for online classification in dynamic network environments.

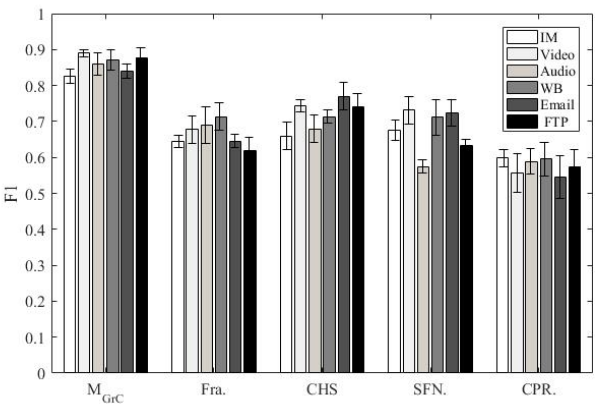


Fig. 3. Classification results in poor network environment.

## IV. Conclusions

In this paper, we conducted an in-depth analysis of traffic online classification. In order to obtain a robust and adaptive classification performance in highly dynamic network environments, we introduced granular computing and proposed a novel classification method. In the model of $M_{GrC}$, the relationship between packets is not isolated but closely correlated. Depending on the inherent correlation information, $M_{GrC}$ can identify the traffic more accurately compared with the traditional classification schemes that assume the packets are independent. Based on granular computing, $M_{GrC}$ is less sensitive to missing and noisy data, and thus can work well even in poor network environment.

However, there are some issues that need to be further explored in the future: 1) Exploring new type of granules to improve the classification performance. Note that the definition of granules is very important for granular computing. In this paper, we define the flow granules according to the concept of neighborhood granules, which prove to be effective for coarse-grained classification. In future work, we will analyze other types of granules to implement fine-grained classification. 2) The issue of threshold setting. For future fine-grained classification, it is difficult to set proper threshold, which affects the performance of the entire system. Therefore, adaptive threshold setting method needs to be investigated to achieve wonderful classification.

## Acknowledgement

This work was supported in part by the NSFC under Grant 61271233, the AHUNSR under Grant KJ2019A0491 and the NSF under Grant IIP-1822055.

## References

[1] A. O. Al-Abbasi, V. Aggarwal, and M. R. Ra, "Multi-tier caching analysis in CDN-based over-the-top video streaming systems," *IEEE/ACM Trans. Netw.,* vol.27, no.2, pp. 835–847, Mar. 2019.

[2] M. Xue, Y. Tang, L. Wu, W. Zhong, and F. Qian, "Switching stabilization for type-2 fuzzy systems with network-induced packet losses," *IEEE Trans. Cybern.,* vol.49, no.7, pp. 2591–2604, Mar. 2018.

[3] X. Luo, H. Wu, and H. Yuan, "Temporal pattern-aware QoS prediction via biased non-negative latent factorization of tensors," *IEEE Trans. Cybern.,* Early Access, pp. 1–12, Apr. 2019.

[4] A. Canovas, J. M. Jimenez, O. Romero, and J. Lloret, "Multimedia data flow traffic classification using intelligent models based on traffic patterns," *IEEE Network,* vol.32, no.6, pp. 100–107, Jul. 2018.

[5] M. Li and L. H. Chen, "Energy-efficient traffic regulation and scheduling for video streaming services over LTE-A networks," *IEEE Trans. Mobile Comput.,* vol.18, no.2, pp. 334–347, Feb. 2019.

[6] N. Carlsson, D. Eager, and V. Krishnamoorthi, "Optimized adaptive streaming of multi-video stream bundles," *IEEE Trans. Multimedia,* vol.19, no.7, pp. 1637–1653, July 2017.

[7] J. Kornycky, O. Abdul-Hameed, A. Kondoz, and B. C. Barber, "Radio frequency traffic classification over WLAN," *IEEE Trans. Parallel Distrib. Syst.,* vol.25, no.1, pp. 56–68, Feb. 2017.

[8] Y. Sun, K. Tang, and L. Minku, "Online ensemble learning of data streams with gradually evolved classes," *IEEE Trans. Knowl. Data Eng.,* vol.1, no.1, pp. 1532–1545, June 2016.

[9] C. Thay, V. Visoottiviseth, and S. Mongkolluksamee, "P2P traffic classification for residential network," in *Proc. IEEE ICSEC,* Chiang Mai, Thailand, Feb. 2016, pp. 23–26.

[10] J. Garcia, T. Korhonen, R. Andersson, and F. Vastlund, "Towards video flow classification at a million encrypted flows per second," in *Proc. 32th IEEE AINA,* Cracow, Poland, May 2018, pp. 358-365.

[11] M. Luo, X. Chang, and L. Nie, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE Trans. Cybern.,* vol.48, no.2, pp. 648–660, Feb. 2018.

[12] A. Tejero-De-Pablos, Y. Nakashima, T. Sato, and N. Yokoya, "Summarization of user-generated sports video by using deep action recognition features," *IEEE Trans. Multimedia,* vol.20, no.8, pp. 2000–2011, Aug. 2018.

[13] X. Yun, Y. Wang, Y. Zhang, and Y. Zhou, "A semantics-aware approach to the automated network protocol identification," *IEEE/ACM Trans. Netw.,* vol.24, no.1, pp. 583–595, Feb. 2016.

[14] E. Aresträm and N. Carlsson, "Early online classification of encrypted traffic streams using multi-fractal features," in *Proc. IEEE INFOCOM ICCN,* Paris, France, Sept. 2019, pp. 84–89.

[15] J. Dong, X. Li, and C. Snoek, "Predicting visual features from text for image and video caption retrieval," *IEEE Trans. Multimedia,* vol.20, no.12, pp. 3377–3388, Dec. 2018.

[16] X. Zhu, W. Pedrycz, and Z. Li, "Granular data description: designing ellipsoidal information granules," *IEEE Trans. Cybern.,* vol.47, no.12, pp. 4475–4484, Dec. 2017.

[17] X. Li, L. Fang, and Z. Lu, "A line flow granular computing approach for economic dispatch with line constraints," *IEEE Trans. Power Syst.,* vol.1, no.1, pp. 4832–4842, Nov. 2017.

[18] S. K. Pal and D. B. Chakraborty, "Granular flow graph, adaptive rule gerneration and tracking," *IEEE Trans. Cybern.,* vol.47, no.12, pp. 4096–4107, Dec. 2017.

[19] H. Fujita, A. Gaeta, and V. Loia, "Resilience analysis of critical infrastructures: a cognitive approach based on granular computing," *IEEE Trans. Cybern.,* vol.49, no.5, pp. 1835–1848, May 2019.

[20] Z. Wu, L. Zhang, and Y. Meng, "Low-rate DoS attacks detection based on network multifractal," *IEEE Trans. Depend. Secure.,* vol.13, no.5, pp.559–567, Feb. 2016.

[21] L. A. Zadeh, "Toward a generalized theory of uncertainty (GTU) - an outline," *Inform. Sciences,* vol.172, no.16, pp.1–2, Aug. 2005.

[22] B. B. Mandelbrot and J. R. Wallis, "Some long-run properties of geophysical records," *Water Resour. Res.,* vol.5, no.2, pp.321–340, Apr. 1969.

[23] I. Hernandez-Carrasco, V. Garcon, and J. Sudre, "Increasing the resolution of ocean pCO2 maps in the south eastern Atlantic Ocean merging multifractal satellite-derived ocean variables," *IEEE Trans. Geosci. Remote,* vol.56, no.11, pp. 2243–2249, Nov. 2018.

[24] F. Shen, X. Yan, and L. Li, "Unsupervised deep hashing with similarity-adaptive and discrete optimization," *IEEE Trans. Pattern Anal. Machine Intell.,* vol.40, no.12, pp. 3034–3044, Dec. 2018.

[25] C. Long, C. Yang, and J. Tao, "Edge computing framework for cooperative video processing in multimedia IoT systems, " *IEEE Trans. Multimedia,* vol.20, no.5, pp. 1126–1139, Mar. 2018.