MARS: Multiple Access Radio Scheduling for a Multi-homed Mobile Device in Soft-RAN

Guolin Sun[1], Kongmaing Eng[1], Seng Yin[1], Guisong Liu[1], and Geyong Min[2]

[1] School of Computer Science and Technology, University of Electronic Science and Technology of China
Chengdu, 611731, China
[e-mail: guolin.sun@uestc.edu.cn, engkongmaing@yahoo.com, yinseng2001@gmail.com,lgs@uestc.edu.cn]
[2] Department of Mathematics and Computer Science, University of Exeter, Exeter, EX4 4QF, U.K.
[e-mail : g.min@exeter.ac.uk]
*Corresponding author: Guolin Sun

Abstract

In order to improve the Quality-of-Service (QoS) of latency sensitive applications in next-generation cellular networks, multi-path is adopted to transmit packet stream in real-time to achieve high-quality video transmission in heterogeneous wireless networks. However, multi-path also introduces two important challenges: out-of-order issue and reordering delay. In this paper, we propose a new architecture based on Software Defined Network (SDN) for flow aggregation and flow splitting, and then design a Multiple Access Radio Scheduling (MARS) scheme based on relative Round-Trip Time (RTT) measurement. The QoS metrics including end-to-end delay, throughput and the packet out-of-order problem at the receiver have been investigated using the extensive simulation experiments. The performance results show that this SDN architecture coupled with the proposed MARS scheme can reduce the end-to-end delay and the reordering delay time caused by packet out-of-order as well as achieve a better throughput than the existing SMOS and Round-Robin algorithms.

---

1. Introduction

Nowadays, mobile Internet is the most commonly used information sharing infrastructure worldwide. The increasing amount of communication capacity and resource consumed by video services has been identified as a big challenge in the next-generation Internet. This challenge for video streaming over Radio Access Networks (RANs) is caused by the delay requirement and bandwidth limitation. In recent years, the growth of video data transmission over multi-path has become an effective way to improve the quality of user experience through enabling low latency in heterogeneous LTE/Wi-Fi networks.

In mobile networks, bandwidth aggregation in multiple radio access networks is another problem that should be addressed [1]. The latency issue is caused by bandwidth limitation. Soft-RAN is an architecture defined to manage the ubiquitous radio resources in heterogeneous RANs in cellular networks [2]. The bandwidth required for video streaming sometimes exceeds the availability in a single access point and this results to more waiting time. The effective way for solving this problem is to aggregate the bandwidth available from multiple Radio Access Technologies (RATs) around the Multi-homed Devices (MDs) in order to improve the end-to-end performance in terms of throughput, delay and out-of-order effects.

The use of multiple interfaces is to address some of the limitations of wireless medium that can also achieve other advantages such as:

- Bandwidth Aggregation: Bandwidth for multiple interfaces can be aggregated to improve or support the demands of clients that need high bandwidth to perform.
- Mobility Support: When a communication path is kept alive, the delay associated with handover is significantly reduced.
- Reliability: Reliability is guaranteed by encoded/duplicated packets that are sent through multiple paths.
- Resource Sharing: Using multiple radio access interface networks, resources can be shared between sender and receiver when the receiver requests the data in multiple paths.
- Data-Control Plane Separation: Decoupling data and control plane can achieve the traffic offloading. For example, LTE can be used for out-of-bandwidth control for Wi-Fi systems to support the data distribution.

The available bandwidth is very important to provide the Quality-of-Service (QoS) guarantee [3]. Therefore, we extend the bandwidth in the radio interface with flow splitting and then aggregate the flows at the edge switch/router, which is connected to a software defined flow manager/controller. In order to evaluate the performance of multipath video transmission, we record the end-to-end delay that measures the delay of each flow from the sender to the receiver, including the queuing delay, access delay, physical transmission delay, propagation delay and reordering delay. We also analyze and reduce the queuing delay, accessing delay, and reordering delay. The main contribution of this paper is the proposed scheduling algorithm that immigrates out-of-order effects by considering limitations on bandwidth and radio access. A relative RTT measurement method is also introduced based on Openflow protocol, which can work in software defined heterogeneous wireless networks.

The rest of this paper is organized as follow: Section II reviews some related works and introduces the multipath transmission methods that improve the latency. In Section III, we describe the proposed Soft-RAN architecture to investigate the flow aggregation and split in RANs, analyze the timing modeling for end-to-end delay and Multiple Access Radio

Scheduling (MARS). Our proposed algorithm is based on relative RTT measurements. The performance evaluation on out-of-order, throughput and end-to-end delay is presented in Section IV. Finally, we conclude our research work on MARS in Section V.

2. Related Work

Video stream transmission over heterogeneous RANs has become popular for mobile applications, such as vehicle-to-vehicle communication, Tele-medicine and the interactive video conference. However, the traditional use of one path to transmit video data with a single network interface has raised problems such as bandwidth limitation, more energy cost and large end-to-end latency. Thus, multiple connectivity has been proposed in the 5G cellular networks to meet the requirements from the scenarios where one path is not enough for streaming video [4], [5], [6], [7].

Here, we consider three main requirements for video transmission [8]: (1) Throughput for transmission with low-quality or resolution that requires a throughput between 0.5Mbps and 1.4Mbps; (2) 720p video quality requires throughput of around 5Mbps; (3) High Definition (HD) video quality requires throughput of around 11 Mbps. However, in the real situation, we face the bandwidth limitation for video streaming, which cannot meet the aforementioned requirements. Furthermore, the end-to-end delay gap between two consecutive frames is normally less then ten microseconds. However, a single path for video streaming cannot achieve it. We therefore resort to multi-path transmission.

On physical layer, multi-path transmission utilized the bandwidth aggregation technique to extend bandwidth resource in order to reduce latency, queuing and transmission delay and so on. Unfortunately, multi-path transmission introduces the packet out-of-order problem, which further leads to increased reordering delay, packet loss, and data re-transmission. In order to avoid the shortcomings of multi-path transmission, some packet scheduling algorithms have been proposed. A Sender-based Multi-path Out-of-order Scheduling (SMOS) scheme was proposed for the high-definition videophone for multi-homed devices to schedule data packets to each path to mitigate the packet out-of-order effects and improve the throughput of video streaming [9]. Moreover, SMOS has been proved to reduce the end-to-end delay for the multi-path video transmission in wired networks. Szymanski *et al.* [10] proposed an ultra-low latency scheme that guaranteed data rate for cloud service using an enhanced Internet that provides ultra-low-latency cloud services. End-to-end delays are effectively reduced to the fiber "time of flight" by using bounded normalized-jitter integer programming to meet the QoS requirement in flow level, increase the aggregation capacity of the Internet, and reduce the Internet operation cost and to improve the energy efficiency.

However, it is deficient in seeking for radio access network with multi-path transmission for video streams. Therefore, a new solution using IP packet reordering for multiple links was introduced in [11] to reduce reordering delay. The proposed system resorts to a large buffer to avoid the delay in packet reordering. The packet scheduling was also employed to reduce buffer requirements at the receiver. This solution achieved a high throughput using the proposed technique. However, there was an increase in transmission delay time. Although it provides a mechanism to handle the packet reordering issue, it lacks in a situation where the buffer is highly constrained and it cannot work accurately in a dynamic radio environment. Bandwidth aggregation provides an effective solution for video streaming in wireless network by extending the spectrum resource. Bandwidth aggregation with SCTP was implemented in [12] to solve the bottleneck that is associated with the shared flows from the same aggregation connection and also supported TCP. Bandwidth

aggregation with STCP allows the bandwidth to be aggregated over multiple interfaces. This scheme considers bandwidth aggregation and includes the packet out-of-order problem, but does not consider transmission delay. Low-latency video transmission over the lossy packet network is proposed to optimize the different picture selection using rate distortion. It aims at increasing the resilience error, which can eliminate re-transmissions. It also reduces the latency of video from 15 seconds to 10 seconds or to hundreds of milliseconds and hence guarantees the QoS [13]. However, this solution does not consider the bandwidth aggregation and the packet out-of-order problem. For real-time applications over heterogeneous RANs, the Earliest Delivery Path First (EDPF) scheduling algorithm was proposed to estimate an idealized bandwidth aggregation to eliminate delivery time on multiple paths, and also a comparison was conducted with a single path transmission using the same bandwidth [14]. This ensures the deadline of video playback. However, EDPF does not provide any mechanisms for the packet reordering at the receiver. It handles the mobility of wireless networks to a certain extent but does not fully handle recurrent wireless networks.

In this paper, we are motivated to split flow at mobile terminals, transmit video stream in multiple paths and aggregate sub-flows at the edge router in the Openflow enabled Soft-RAN. MARS eliminates the reordering delay at the receiver side to a maximum level for multi-path transmissions, which proposes a probabilistic algorithm to dispatch the packets at the sender side. The main contribution of this paper is the RTT based multi-path scheduling method in Soft-RAN to reduce the end-to-end delay and achieve ultra-low latency of video transmission.

3.   The Proposed Method

3.1 Soft-RAN Architecture

The proposed system architecture for flow aggregation and splitting is based on SDN with RANs as shown in Fig. 1. There are three elements considered to perform flow splitting and aggregation: bandwidth, QoS requirement and network statistics, such as Channel Quality Indicator (CQI) for each user and RTT measurement for each radio link. These factors are considered as theinputs to determe the action to be taken on data transfer by the SDN controller. In this paper, relative RTT measurements and bandwidth are used for packet scheduling to solve the out-of-order problem.

There are two control modules used in determining actions on the flows, network statistics and traffic analysis modules. Network statistics module constantly monitors the network status, gathers information based on protocols, which is vital to determine actions on flows. The essential identifiers for flow operations could be CQIs for each user, and the number of flows and RTTs on each radio link. As a result, the flows with the distinct characteristics of traffic, packet size and gap between two consecutive packets require bandwidth and deadline limit. They jointly determine the tag of bandwidth, QoS, and RTT for each flow.

The flow control algorithm makes decision on the actions with the mapping table and outputs actions on data plane with the inputs from the network statistics module and the traffic analysis modules. It then executes the flows aggregation and splitting. For example, a video unicast flow can be split on network devices when it requires large bandwidth and urgent deadline limitation. This video multicast flow can be duplicated in several groups of unicast flows with adaptive transmission rates, and also can be cached in network based on its re-usability. The control plane issues control messages via SSL to allocate radio resources and control flow aggregation and splitting on the network devices. The use case discussed in

this paper is assumed that the sum of sub-flow bandwidth is fixed for a continuous video flow based on its traffic characteristics and QoS requirements, whatever the number of paths is. We consider a scheme that the bandwidth allocation for multiple paths will be dynamically changed based on the distinct time-varying RTTs of each radio link.
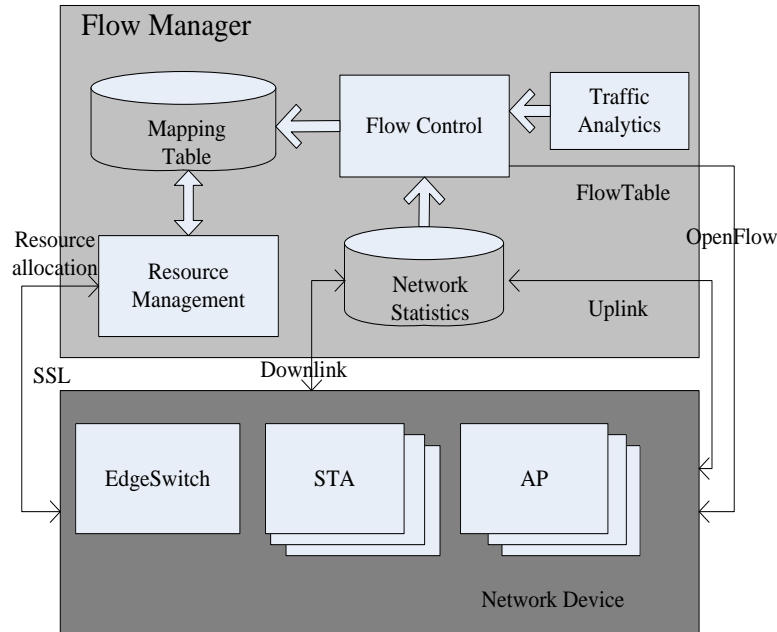

Fig. 1. The proposed architecture

3.2   Flow aggregation/split with OpenFlow

The essential structures of flow tables are already defined in the OpenFlow specifications. In this paper, a three-level table structure is designed to enable flow aggregation and splitting in soft-RAN, the basic flow table, the aggregation table and the mapping table, as shown in **Fig. 2**. The aggregation table has multiple general flow tables defined with two existing structures in Openflow specification, that is, group table and multiple table. The group tables allow the different flows to share a common action set in the OpenFlow specification V.1.2. Flow aggregation can be implemented with group tables. The multiple tables allow matching once, followed by a sequence of actions as defined in the OpenFlow specification V.1.1. The functionality of flow splitting could be implemented via the multiple tables.

The mapping table is used in the flow manager to build a mapping relationship between one flow identifier and multiple flow identifiers. The distinct data structures are defined for flow splitting and aggregation for uplink and downlink in mobile scenarios. In **Fig. 2**, RTT value is the metric for network statistics in the mapping table. The virtual MAC is only chosen as an option to replace the real physical interface address for an aggregated flow at the edge switches.
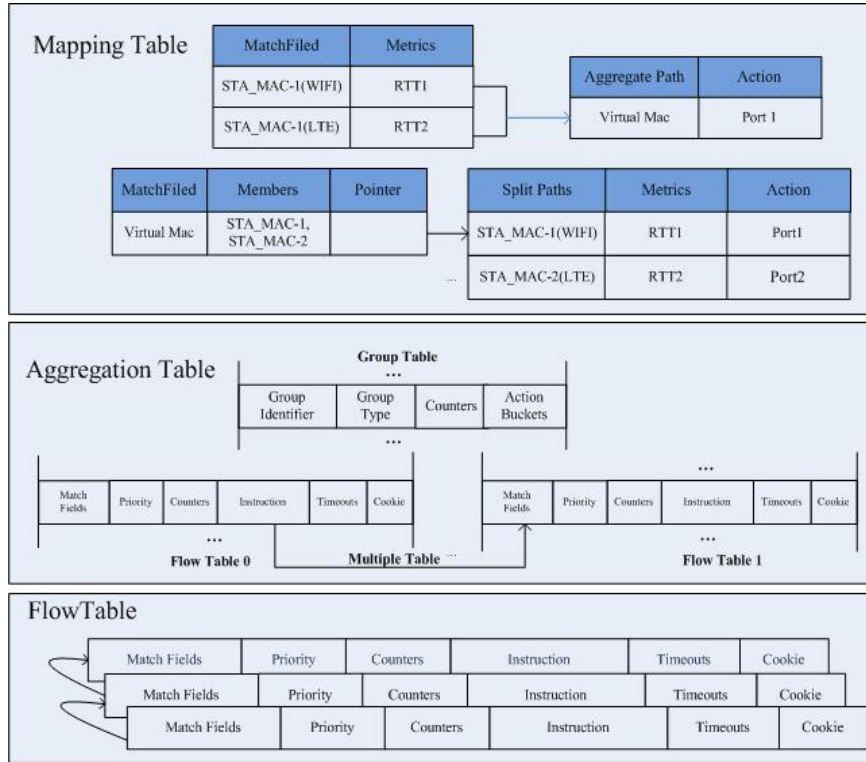
Fig. 2. The three-level flow table structure

3.3   Relative RTT Measurement with Openflow

In order to evaluate the effectiveness of each path for multi-path transmission, a relative RTT measurement method with Openflow for a multi-homed mobile device is introduced, as shown in **Fig. 3**. The relative RTTs will be updated in a fixed period to make sure that it is always valid in the time-varying conditions. Based on the idea of relative RTT measurement, an MD sends two duplicated packets from two distinct radio interfaces to two forward radio links, such as Wi-Fi and LTE, is transferred at the edge switch, and then goes back on a common reverse path. This transfer process at the edge switch can be implemented with the group table in Openflow defined in **Fig. 3**. We would not like to measure the absolute forward path delays because it requires the tight clock synchronization between sender and receiver. In trying to determine the relative delay of N forward paths, we do not consider all possible RTT combinations but N probes along different forward paths. These will return and follow a common reverse path [15]. As a result, the low-latency paths will be chosen relatively.
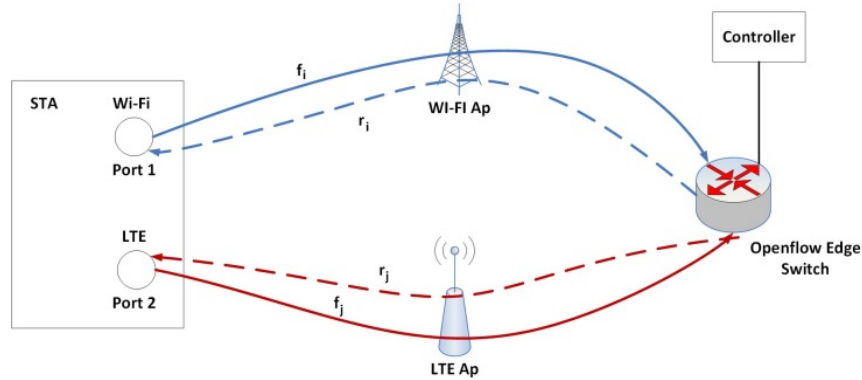
Fig. 3. Relative Round-trip Time Measurement

In **Fig. 3**, the reverse path for RTT requests is r_i or r_j. The relative RTT can be determined between the round-trip path ($f_i r_i$) and ($f_j r_j$). This is the delay difference between forward paths (f_i) and (f_j). The delay of forward paths(f_i) and (f_j) are respectively represented by(df_i) and (df_j).

$$RTT(\ f_i r_i\ )\text{-}RTT(\ f_j r_j\ ) = \ df_i + dr_i -(\ df_j + dr_i) = \ df_i - df_j, i, j = 1,2\ldots \tag{1}$$

The path with the lowest one-way delay is determined by comparing the RTTs among the forward paths. To minimize the data transmission delays for any device in multi-path transmission, we choose the radio links with small RTTs mentioned above.

3.4 Timing Model

In this paper, we aim to reduce the end-to-end delay using the multiple access radio scheduling on packets for the multi-path video transmission in a dynamic radio environment. We setup a system timing model for the end-to-end delay. This is an important QoS metric commonly used in latency-sensitive applications. The end-to-end delay is denoted as $E2E$, which is the sum of the transmission delay, the propagation delay, the processing delay, the queuing delay and the reordering delay, and is given by:

$$E2E = d_{trans} + d_{prop} + d_{proc} + d_{queue} + d_{reor} \tag{2}$$

$E2E$ includes all processing delays from the sender to the receiver. The transmission delay is introduced by the physical transmitter in the hardware. The propagation delay is caused by the distance between mobile stations to APs. The processing delay is caused by decoding or packet matching at transfer devices. The queuing delay is considered as a result of packet buffering in transfer from one hop to another hop. The access delay in MDs occurs in the air interface of RANs when sending the resource to other users. In the end, we consider the reordering delay, which happens only in the case of multipath transmission at the destination. In order to get a better understanding of end-to-end delay, a system timing model is illustrated in **Fig. 4**. In our proposed system model, the video blocks are dispatched into several paths by the packet scheduler. The controller computes the available bandwidth and the RTT of each path for each flow, and dispatches the packets into the paths based on scheduling algorithm MARS. This process introduces the queuing delay during transmission.

The source generates an initial time-stamp for each packet. The data mapping from packets in data bits to the physical radio interface is done in transmitter. During this process, two types of delay are introduced namely, the transmission delay and the radio access delay. The base station is also introduced the transfer delay and the propagation delay. The reordering delay is caused by transmiting video stream on multiple paths with the different delay characteristics on the internet and recovering the flow in order at the receiver. A system model of video transmission via multi-path over the LTE/Wi-Fi networks is illustrated in **Fig. 4**, in order to record the delay timing.
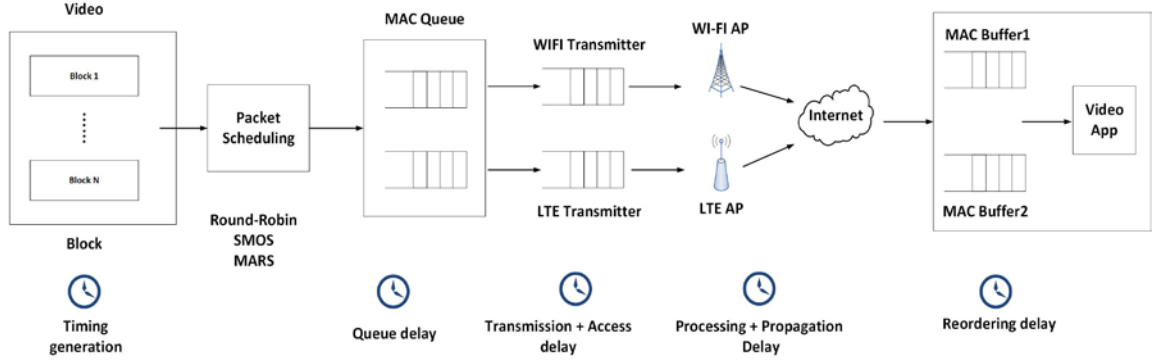

Fig. 4. The system timing model

3.5 Scheduling Algorithm

As an ideal model, the initial probability to be scheduled for each path with the equal bandwidth is assumed to be equal. However, in the real case, the available bandwidth of each path is distinct and is dependent on the resource allocation algorithms. The probability for scheduling opportunity for each path in real case is computed as a function of bandwidth based on the equation (6). As illustrated in **Fig. 5**, we assume there are N_p paths with distinct delay characteristics. The delay for each path includes the access delay and the transmission delay. The transmission delay is usually very small in few microseconds, while the access delay can be tens or hundreds of microseconds. Therefore, the transmission delay can be ignored. In the scheduling procedure, the sender schedules data packets based on the bandwidth available and the R_i values on each path per video flow. The scheduling algorithm computes the resources and decides the packet flow through different paths. The scheduling algorithm keeps monitoring the system, adjusts the resources, and makes the decisions dynamically. The period of scheduling depends on how often is the monitoring measurements updated.

Step 1: Update the equivalent bandwidth (b_i) with the relative RTT R_i measurements based on Equations (3), (4), and (5), and then compute the probability (p_{ih}) for path i.

Step 2: Schedule packets in each trunk into multiple parallel paths based on their scheduling probability (p_{ih}). Then we split the 1st trunk from the fth flow and the 2nd trunk from the fth flow, and so on. This process continues as a loop until the $(N-1)^{th}$ trunk is dispatched.

In general, for the hth trunk, we update the bandwidth (b_i) with the measurement of relative Round-Trip-Time (R_i) on each path. Then we compute (p_{ih}) with the updated (b_i) to find the probability of each path.
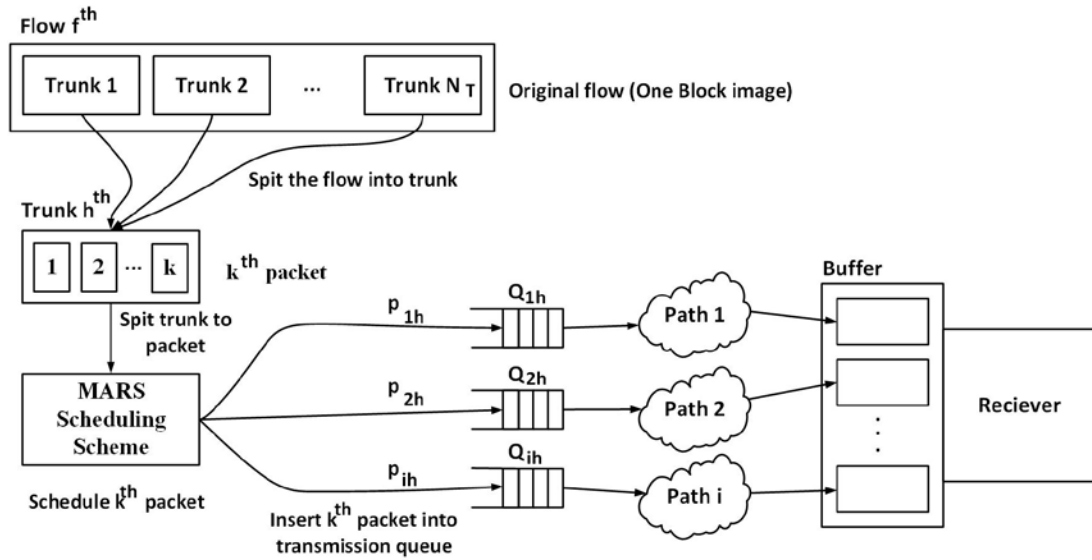

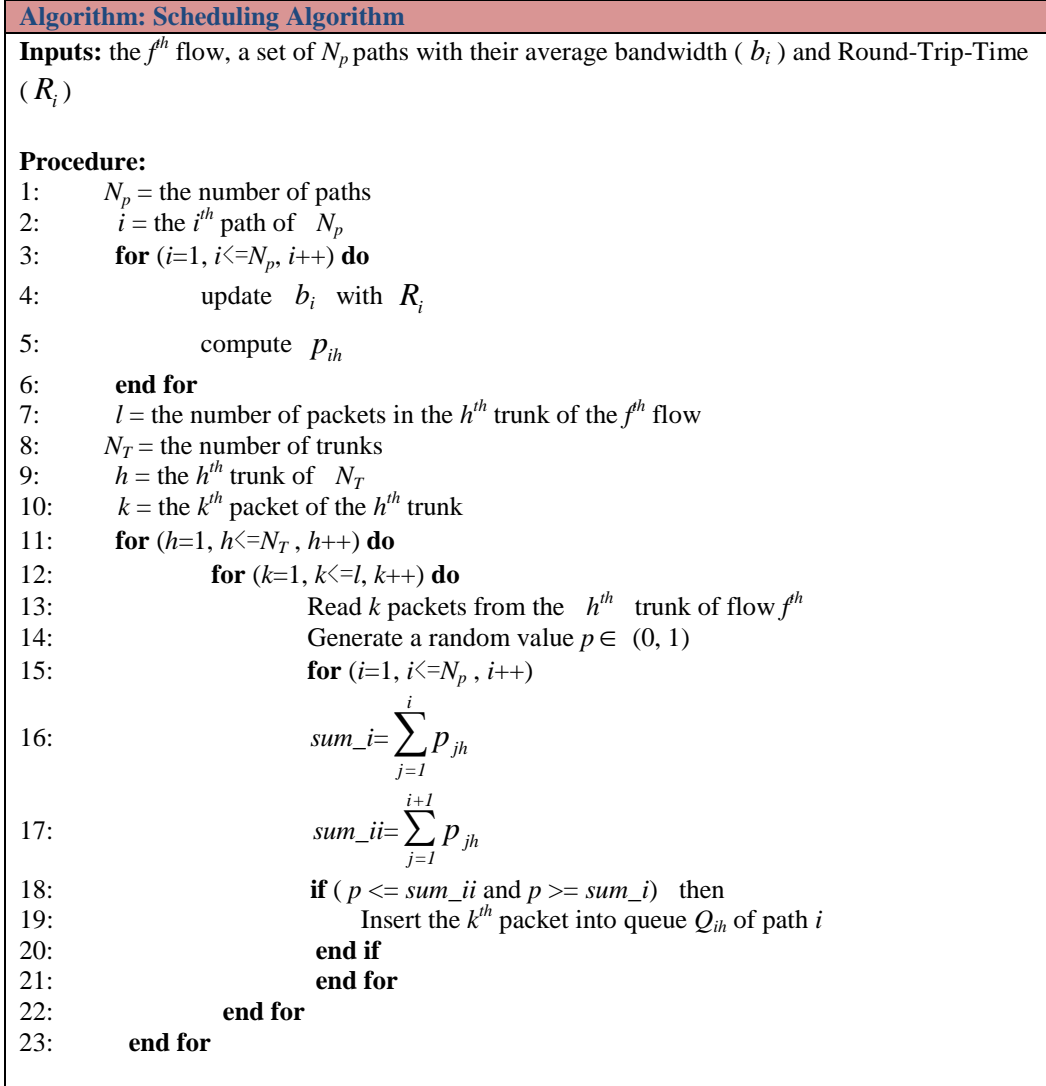
Fig. 5. Overall Architecture of MARS

The hth trunk $(1 \leq h \leq N)$ in the fth flow is scheduled in steps as illustrated in **Fig. 6**. The scheduler dispatches each trunk into multiple paths. Each trunk has its own scheduled packet by computing the probability (p_{ih}). To schedule the kth packet of the trunk, we read the kth packet from the hth trunk and generate a random value $p(0,1)$. After generating the value $p(0,1)$, we compare the random value p with the probability p_{ih} of path i to insert the kth packet into the transmission queue Q_{ih}. If p falls into the probability p_{ih} zone of the ith path, then the kth packet is inserted into the transmission queue Q_{ih} of the ith path.

To compute the probability (p_{ih}) of each path, we update bandwidth (b_i) with the relative $RTTs$ (R_i) value, as shown in Equation (3).

$$b_i = \frac{packetSize}{AccessDelay} = \frac{L_i}{A_i} \tag{3}$$

The access delay can be estimated with the relative $RTTs$ (R_i) as:

$$A_i = R_i - \frac{L_i}{b_i} \tag{4}$$

Algorithm: Scheduling Algorithm

Inputs: the fth flow, a set of N_p paths with their average bandwidth (b_i) and Round-Trip-Time (R_i)

Procedure:
1:      N_p = the number of paths
2:      i = the ith path of N_p
3:    **for** (i=1, $i <= N_p$, i++) **do**
4:          update b_i with R_i
5:          compute p_{ih}
6:     **end for**
7:      l = the number of packets in the hth trunk of the fth flow
8:     N_T = the number of trunks
9:      h = the hth trunk of N_T
10:      k = the kth packet of the hth trunk
11:    **for** (h=1, $h <= N_T$, h++) **do**
12:            **for** (k=1, $k <= l$, k++) **do**
13:                  Read k packets from the hth trunk of flow f^{th}
14:                  Generate a random value $p \in$ (0, 1)
15:                  **for** (i=1, $i <= N_p$, i++)

16:                  $sum\_i = \displaystyle\sum_{j=1}^{i} p_{jh}$

17:                  $sum\_ii = \displaystyle\sum_{j=1}^{i+1} p_{jh}$

18:                  **if** ($p <= sum\_ii$ and $p >= sum\_i$)   **then**
19:                      Insert the kth packet into queue Q_{ih} of path i
20:                    **end if**
21:                  **end for**
22:          **end for**
23:     **end for**

Fig. 6. MARS: Multiple Access Radio Scheduling algorithm

In our simulation, we mimic the R_i measurement procedure to generate random numbers from a fixed set of values in milliseconds (ms). Equation (5) is the updated bandwidth (b_i) with the R_i.

$$b_i = \frac{L_i \times b_i}{(R_i \times b_i) - L_i} \tag{5}$$

For the hth trunk $(1 \le h \le N)$ of the fth flow, the scheduling probability of the ith path is given as:

$$p_{ih} = \frac{b_i}{\displaystyle\sum_{i=1}^{h} b_i}, 1 \le h \le N \tag{6}$$

4. Experiments and Results

In this section, we evaluate the performance of the proposed multiple access radio scheduling scheme for a multi-homed mobile device in Soft-RANs. The simulation experiments are done in both the static and dynamic scenarios for multi-path transmission. We configure the scenarios with one mobile device and four links to the distinct access points. We consider the bandwidth and RTT for the distinct radio links can be dynamically changed in the dynamic scenario and can be fixed in the static scenario, but the total bandwidth for multiple paths is not changed for a video flow. In this experiment, a constant bandwidth 1.14MHz is configured as the total bandwidth for this video flow. The bandwidth for each path will be dynamically changed. The packet size for video transmission is configured as 2 KB. The RTT values can be dynamically changed to mimic the congestion varying on each link. In this experiment, the RTT generated randomly from the set (800, 300, 100, 40, 20 and 10) in ms. The scheduler periodically dispatches the packet flow into different paths based on the algorithm. Then we evaluate the performance at the receiver on reordering delay, the E2E delay gap and throughput for each scenario. The impact of packet size on the above metrics is also analyzed.

4.1 Reordering Delay

For the multiple path transmission, if the RTTs are unbalanced on each path, the out-of-order problem will happen at the receiver side. The difference between the sending sequential number (SSN) and the receiving sequential number (RSN) for packets in a video flow is shown in **Fig. 7**.
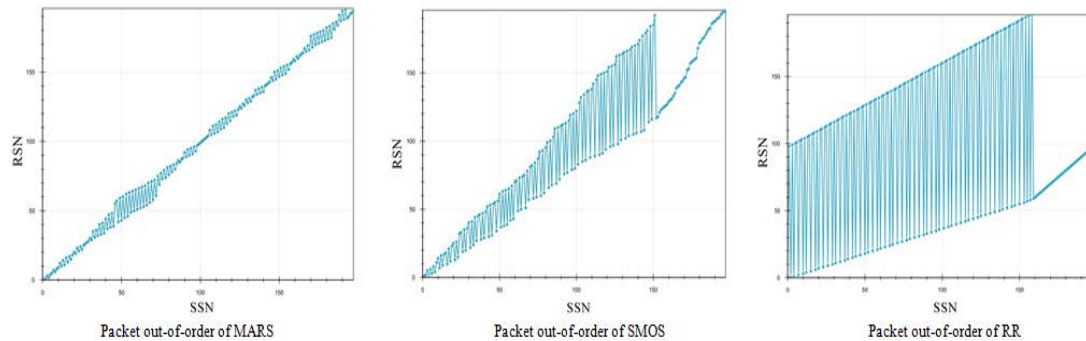

Fig. 7. The out-of-order problem for the balanced scenario and the dynamic scenario

The results with three distinct packet scheduling algorithms are shown in **Fig. 7**, such as MARS, SMOS and Round-Robin. The effects of packet out-of-order problem at the receiver are represented with the difference between SSN and RSN. In **Fig. 7**, we use a sorting function *sort* (.) to reorder packets at the receiver and count the reordering complexity in time at the receiver, which aims to compare the reordering delay in the different scenarios.

The results for the out–of-order problem in **Fig. 8** are obtained with Test Set 1 and Test Set 2. Test Set 1 is configured with the balanced RTT and the balanced bandwidth. Test Set 2 is configured with the random RTT and the random bandwidth. Test Set 1 includes the results when we setup the set of parameters with the balanced RTT and the balanced bandwidth. Test Set 2 describes the results of the experience when we change the parameters of random RTT and bandwidth. The RTT values are randomly generated from this set of values (800, 300, 100, 40, 20 and 10).
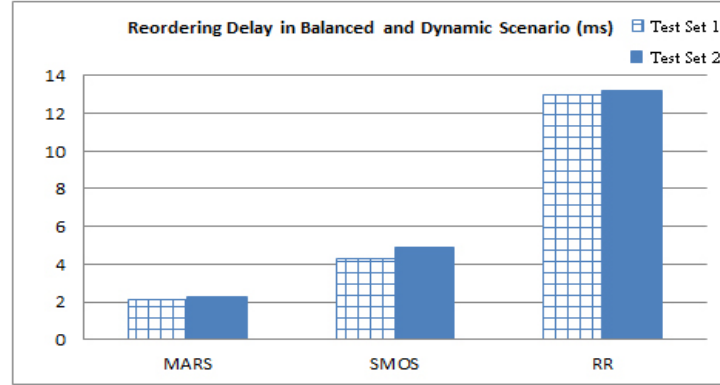
**Fig. 8.** The average re-ordering delay for the balanced and dynamic scenarios

The longer exchange time in the sorting procedure leads to a larger reordering delay. In **Fig. 8**, MARS achieves the best performance and reduces the reordering delay for both Test Set 1 and Test Set 2. Test Set 1 is configured with the balanced RTT and the balanced bandwidth. Test Set 2 is configured with the random RTT and the random bandwidth.

4.2 Static scenario

In the static scenario, we fixed one factor of a pair of RTT and bandwidth for each path in the simulation. We fixed the packet size as 2KBs and the number of paths as 2. Two specific scenarios are set up. Scenario 1 is configured with the balanced RTT and the unbalanced bandwidth, and Scenario 2 is configured with the unbalanced RTT and the balanced bandwidth.

4.2.1 Scenario 1

In Scenario 1, the delay gap between two consecutive blocks/images is measured for a video flow at the receiver. The balanced RTTs are defined as (10, 10) and (800, 800). The bandwidth for *N* paths is generated randomly. The number of paths changes from one to five. The results in **Fig. 9** show that the small RTT can introduce a smaller delay gap at the receiver. When the number of paths is equal to 3, it achieves the minimum delay gap.
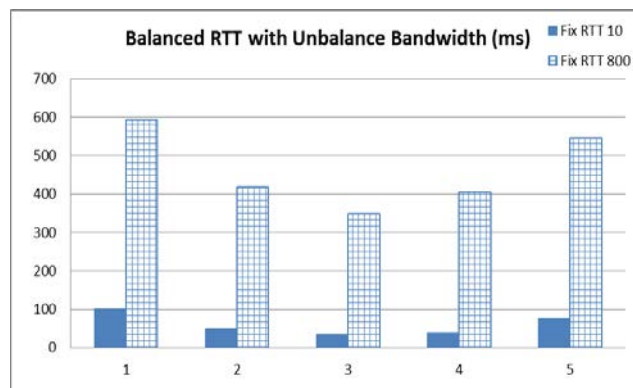
Fig. 9. The E2E delay gap for the scenario 1

4.2.2 Scenario 2

In Scenario 2, each path is set with the balanced bandwidth and the unbalanced RTTs. One path has a smaller average RTT and the other path has a bigger average RTT. Series 1 show a list of the unbalanced RTT set, {(800-10), (300-10), (100-10), (40-10), (20-10), (10-10) }. The series 2 is a list of the unbalanced RTT set, {(800-800), (300-800), (100-800), (40-800), (20-800), (10-1800)}. The results in **Fig. 10** reveal that the delay gap heavily depends on the smallest RTT in a set of two or more paths. Test Set 1 with a small RTT of 10ms leads to a small end-to-end delay gap. Test Set 2 results in a higher end-to-end delay gap because the smallest RTT increases from 10ms to 800ms, which is greater than 10ms.
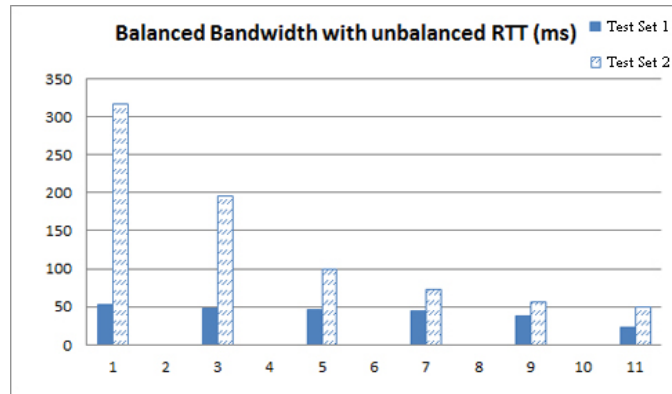

Fig. 10. The E2E delay gap for the scenario 2

4.3 Dynamic scenario

In the dynamic scenario, we set up the scenarios with the random bandwidth and the random RTT to mimic dynamic status of network changes in real case. The performance in terms of E2E delay gap, throughput, and the impact of packet size is analysed.

4.3.1 E2E Delay Gap

In the dynamic scenario, the simulation configuration is given in **Table 1**, which is also a summary of end-to-end delay for three sender-based scheduling algorithms, such as Round Robin, SMOS and MARS. The E2E delay is collected from the simulation environment, which is limited by the computer processing capability. Round Robin achieves the worst performance on the E2E delay gap because it is blind to the difference on bandwidth and RTT for each path. SMOS is better than Round Robin because it considers the resource difference on wired bandwidth only. MARS scheme is adaptive to the dynamic changes of the bandwidth and the RTT, and achieves the smallest delay gap. MARS provides a better performance with a stable end-to-end delay for video streaming with the distinct number of paths.

Table 1. The simulation configuration and the E2E Delay

Number of path	Bandwidth (1.14Mbps)	Average End-To End Delay (ms)		
		MARS	**SMOS**	**Round-Robin**
1	Random	21,458	21,458	21,458
2	Random	21,546	23,409	24,397
3	Random	21,596	26,061	28,254
4	Random	21,691	30,640	32,531

4.3.2 Throughput

When the number of paths changes from one to four, MARS is proved to achieve the best performance of throughput. In this experiment, the dynamic scenario is configured with random bandwidth and random Round-Trip-Time. The RTTs are chosen randomly form a set as (800, 300, 100, 40, 20 and 10) in microseconds. MARS achieves the highest throughput in dynamic scenario, which is shown in **Fig. 11**. SMOS scheme still can achieve a higher throughput than Round-Robin.
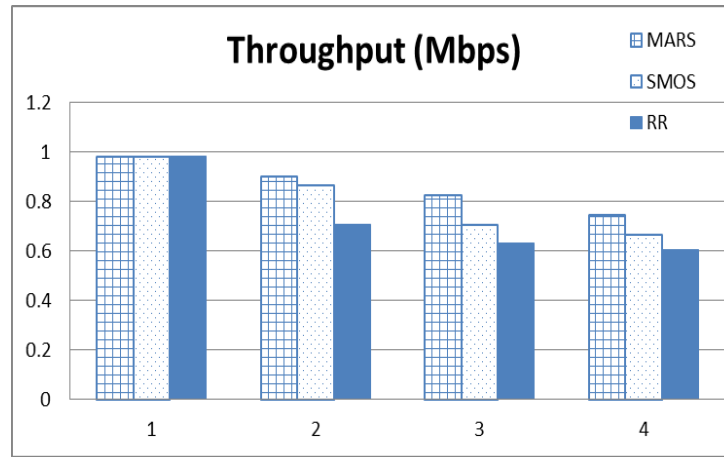

Fig. 11. The throughput in the dynamic scenario

4.3.3 Packet size

We investigate the impact of packet size on the E2E gap for the video flows. When the packet size changes from 2KB, 1.5KB, 1KB, 0.5KB, to 0.3KB, the end-to-end delay gap increases because the average queuing waiting delay increases. The result is shown in **Fig. 12**. In real case, the packet size depends on the specific application and protocol. The results demonstrate that MARS scheduling scheme achieves the smallest delay gap, and SMOS scheme is still better than Round Robin, whatever the packet size is.
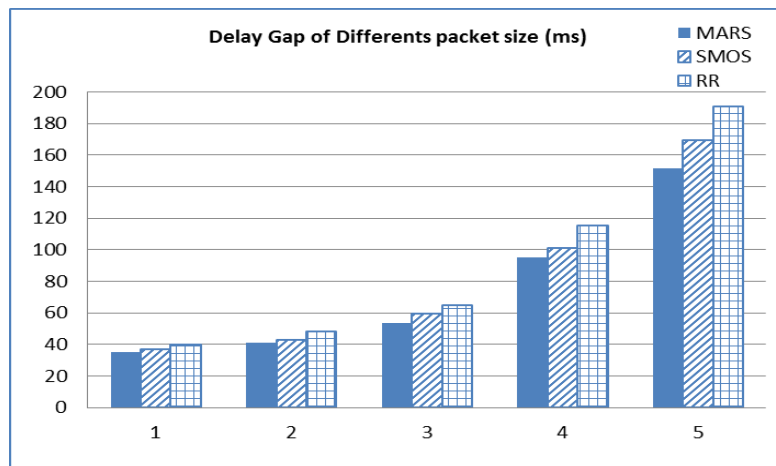

Fig. 12. The delay gap of different packet sizes

The impact of packet size on throughput is depicted in **Fig. 13**. MARS achieves the highest throughput with the packet size 2KB.
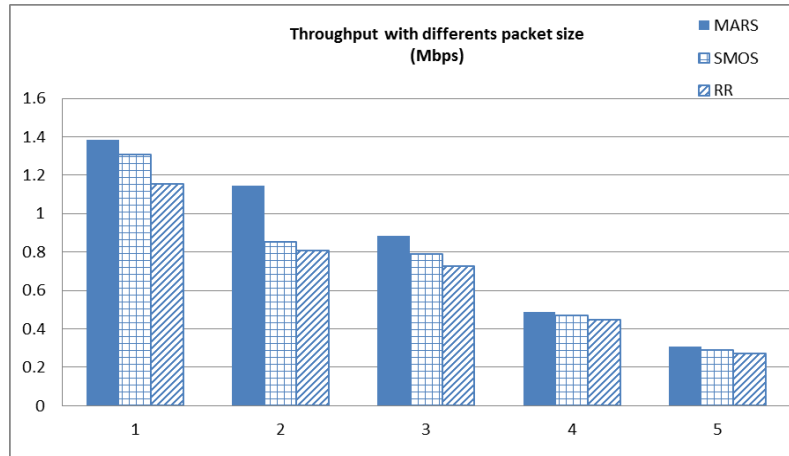
Fig. 13. Throughput achieved with distnct packet sizes

Based on the performance results above, we can conclude that the end-to-end delay gap can be reduced if flow aggregation and packet aggregation are incorporated into soft-RANs for the scenario of smart grid and vehicle-to-vehicle.

5. Conclusion

In this paper, we present a SDN-based architecture for multipath video transmission to reduce the end-to-end delay gap. A Multiple Access Radio Scheduling (MARS) scheme based on relative Round-Trip Time (RTT) is proposed for soft-RANs. The experimental results show that MARS performs better with low latency and higher throughput than SMOS and Round Robin. Our proposed scheme reduced the out-of-order problem for multi-path transmission at the receiver. The results also show that a better performance in terms of delay gap and throughput using bandwidth allocation for multiple radio connection can be achieved. The minimum RTT value in a set of RTTs for each path determines the delay gap at the receiver in the unbalance scenario. The packet size has impacts on the performance of multiple radio transmission and indicates that packet aggregation and flow aggregation may be a promising way to improve the performance in soft-RAN for the applications with *mice* flows. The scheduling algorithm has been proved to be effective and has important impacts on multipath TCP transmission and dual-connectivity in 5G. In the future work, we intend to apply this algorithm on adaptive bandwidth aggregation and splitting for multi-homed mobile devices.

References

[1] P. Sharma, S-J. Lee, J. Brassil, and K.G. Shin, "Aggregating bandwidth for multihomed mobile collaborative communities," *IEEE Transactions on Mobile Computing*, vol. 6, no. 3, pp: 280-296, 2007. Article (CrossRef Link)

[2] A. Gudipati, D. Perry, L.E. Li, S. Katti, "SoftRAN: Software defined radio access network," in *Proc. of the 2nd ACM SIGCOMM workshop on Hot topics in software defined networking*, pp. 25-30, 2013. Article (CrossRef Link)

[3] M. Jain and C. Dovrolis, "End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput," *IEEE/ACM Transactions on Networking*, vol. 11, no. 4, pp. 537-549, 2003. Article (CrossRef Link)

[4] J. Wu, C. Yuen, B. Cheng, M. Wang, J. Chen, "Streaming high-quality mobile video with multipath TCP in heterogeneous wireless networks," *IEEE Transactions on Mobile Computing*, 2015. Article (CrossRef Link)

[5] Jo M, Maksymyuk T, Strykhalyuk B, et al., "Device-to-device-based heterogeneous radio access network architecture for mobile cloud computing," *IEEE Wireless Communications*, vol. 22, no. 3, pp. 50-58, June 2015. Article (CrossRef Link)

[6] J. Wu, Y. Shang, X. Qiao, B. Cheng, J. Chen, "Robust bandwidth aggregation for real-time video delivery in integrated heterogeneous wireless networks," *Multimedia Tools and Applications*, vol. 74, no. 11, 2015 , pp. 4117-4138. Article (CrossRef Link)

[7] A. Sridharan, R.K. Sinha, R. Jana, H. Bo, K.K. Ramakrishnan, N.K. Shankaranarayanan, I. Broustis, "Multi-path TCP: Boosting fairness in cellular networks," in *Proc. of the 22nd IEEE International Conference on Network Protocols (ICNP*), pp. 275-280, 2014.
Article (CrossRef Link)

[8] W. Stallings, "Data and Computer Communications," *Pearson Education*, The Eight Edition, ISBN: 0-13-243310-9, 2007. Article (CrossRefLink)

[9] J. Liao, J. Wang, T. Li, X. Zhu, P. Zhang, "Sender-based multipath out-of-order scheduling for high-definition videophone in multi-homed devices," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp.1466-1472, July 2010. Article (CrossRef Link)

[10] T.H. Szymanski, "An ultra-low-latency guaranteed-rate Internet for cloud services," *IEEE/ACM Transactions on Networking*, 2015. Article (CrossRef Link)

[11] D. Kaspar, K. Evensen, A.F. Hansen, P. Engelstad, P. Halvorsen, C. Griwodz, "An analysis of the heterogeneity and IP packet reordering over multiple wireless networks," in *Proc. of IEEE Symposium on Computers and Communications*, pp. 637-642, 2009. Article (CrossRef Link)

[12] A. Argyriou, V. Madisetti, "Bandwidth aggregation with SCTP," in *Proc. of IEEE Global Telecommunications Conference*, pp. 3716-3721, 2003. Article (CrossRef Link)

[13] Y.J. Liang, M. Flierl, B. Girod, "Low-latency video transmission over lossy packet networks using rate-distortion optimized reference picture selection," in *Proc. of IEEE International Conference on Image Processing*, pp. 181-184, 2002. Article (CrossRef Link)

[14] K. Chebrolu, R.R. Rao, "Bandwidth aggregation for real-time applications in heterogeneous wireless networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 4, pp: 388-403, 2006. Article (CrossRef Link)

[15] E.P. Ribeiro, V.C.M. Leung, "Minimum delay path selection in multi-homed systems with path asymmetry," *IEEE Comm. Letters*, vol. 10, no. 3, pp.135-137, 2006. Article (CrossRef Link)

Guolin Sun received his B.S., M.S. and Ph.D. degrees all in Communications and Information Systems from the University of Electronic Science and Technologies of China (UESTC), Chengdu, China, in 2000, 2003 and 2005 respectively. After Ph.D. graduation, Dr. Guolin has got eight years industrial work experiences in wireless research and development for LTE, Wi-Fi, Internet of Things (ZIGBEE and RFID, etc.), Cognitive radio, Location and navigation. Before he joined the School of Computer Science and Engineering, University of Electronic Science and Technologies of China, as an Associate Professor on Aug. 2012, he worked in Huawei Technologies, Sweden. Dr. Guolin Sun has filed over 30 patents and published over 30 scientific conference and journal papers. He has acted as TPC member of several conferences. Currently, he serves as a vice-chair of the 5G oriented cognitive radio SIG of the IEEE Technical Committee on Cognitive Networks (TCCN) of the IEEE Communication Society. His general research interest is 5G/2020 oriented wireless networks, including software defined networks, network function virtualization, wireless networks.

Eng kongmaing received the Bachelor degree in Computer Science from Royal University of Phnom Penh, Cambodia, and the Bachelor degree in Electricity Engineering from Preah kossamak Polytechnic Institute, Cambodia in 2013. He is currently studying MSc in computer science at University of Electronic Science and Technology of China, due to finish in 2015. He worked as a website developer for Potential Asia Group from 2012 to 2013. His interests include communications, Internet-of-Things and video transmission.

Yin Seng received the Bachelor degree of Computer Science from Royal University of Phnom Penh, Cambodia and the Bachelor degree of English Literature from BELTEI University, Cambodia in 2013. He is studying MSc in Computer Science at the University of Electronic Science and Technology of China and is due to complete in 2016. From 2012 to 2013, he worked as a Software Engineer for Pathmazing. From 2013 to 2014, he worked as a Software Developer for Open Institute. His interest includes communications, internet of things, RFID, and video transmission.

Guisong Liu received his B.S. degree in Mechanics from the Xi'an Jiao Tong University, Xi'an, China, in 1995, M.S. degree in Automatics and Ph.D. degree in Computer Science both from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2000 and 2007 respectively. Now, he is an associated professor in the School of Computer Science and Engineering, UESTC. His research interests include cloud computing, big data, and computational intelligence.

Geyong Min is a Professor of High Performance Computing and Networking in the Department of Mathematics and Computer Science within the College of Engineering, Mathematics and Physical Sciences at the University of Exeter, United Kingdom. He received the PhD degree in Computing Science from the University of Glasgow, United Kingdom, in 2003, and the B.Sc. degree in Computer Science from Huazhong University of Science and Technology, China, in 1995. His research interests include Future Internet, Computer Networks, Wireless Communications, Multimedia Systems, Information Security, High Performance Computing, Ubiquitous Computing, Modelling and Performance Engineering.