Reliability Improving with Local Utility Scheduling and Global Overlay Routing Algorithm of Real Time Service in SAE

Olfa Ben Ahmed
Mediatron, Sup'com
University Of Carthage,
Tunisia

Zied Choukair
Mediatron, Sup'com
University Of Carthage,
Tunisia

Meriem Afif
Mediatron, Sup'com
University Of Carthage,
Tunisia

ABSTRACT
System Architecture Evolution (SAE) is the IP based core network of Evolved Packet System (EPS). In core network telecommunication service is reliable if its QoS criteria are satisfied. Our work aims to improve reliability of real time services by minimizing latency. We propose tow contributions; *Delay utility scheduling* to improve service timeliness in nodes and *Overlay Routing Algorithm* to identify the optimum path having the minimum latency between source and destination**.**

Keywords
EPS, Core Network, Dependability, Reliability, Scheduling, Routing, Network Virtualization

1. INTRODUCTION
The Evolved Packet System (EPS) includes the Long-Term Evolution (LTE) which is composed of evolved radio access network (E-UTRAN) and system architecture evolution (SAE).

LTE contains new network elements called enhanced Node Base (eNB), which provide E-UTRA user plane and control plane termination towards the user equipment UE [1].

The SAE is a flat network architecture, IP based multi-access core network that supports the operation of a common packet core network for 3GPP radio accesses, non-3GPP radio accesses and fixed accesses. The two main elements of SAE are Serving Gateway (S-GW) and Packet Data Network Gateway (PDN-GW)[2]. S-GW acts as local mobility anchor, exchanging packets with eNB, where UEs are served. It serves as routing node towards other 3GPP technologies. PDN-GW interfaces with the external PDNs. It performs IP related functions like address allocation, policy enforcement, packet classification and routing. It also acts as mobility anchor for non-3GPP access networks.

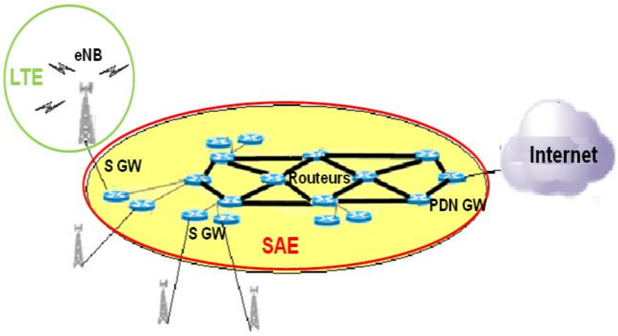

Fig 1: EPS Architecture

In contrast to the circuit network switching limited to voice services, mobile telecommunication networks offer multimedia applications with audio services, video and integrated data. In the EPS system all multimedia sessions go through the LTE access network then by the switching packet network that SAE.

We consider in this work the following main failures of mobiles services:

- Blocking and outage in LTE
- Delay of real-time service and data loss in SAE

We aim to improve telecommunication services dependability. Dependability is the discipline that quantifies the reliance that can be placed on the service delivered by a system [3] and consists of two major aspects: availability and reliability [4]. Additionally, there are several other aspects of dependability: Maintainability is the aptitude to undergo repairs and evolutions. Safety is dependability with respect to catastrophic failures. Integrity is an aspect of dependability that is more commonly associated with security.

In access network (LTE) telecommunication service is available if it is admitted by eNB and is reliable if it is still supported in handover position [5]. In core network (SAE) it is available if packets are not dropped and reliable if its QoS criteria are satisfied. In this work, we aim to improve reliability of real time services.

After having introduced our context, we focus primarily on the expression of our solutions to improve EPS dependability. The paper is organized as follows; Section 3 defines our classification of traffic. We present and discuss our tow contributions such as *Delay Utility Scheduling* and *Overlay Routing Algorithm* in Section 4 and 5. Concluding remarks are made in Section 6.

2. EPS DEPENDABILITY
To improve EPS dependability we propose an approach:

- To reduce blocking and outage rate of services in access network, we propose in [5] a new *Admission Control* algorithm using preemption bandwidth from already connected applications according to a specified "utility" policy protocol. Experimentations give results that show an improvement of eNB availability and reliability with acceptable QoS.
- To minimize latency of real time services in core network, our contributions are; a new *Scheduling algorithm* using utility delay and a new *QoS Routing* algorithm using overlay.

In this paper we present and discuss our contributions to minimize latency of real time service.

3. SLA CLASSIFICATION

According to [6] we classify the applications into three groups in term of real-time requirements:

- G1 to highest requirements services (group1 G1 :: (Real Time Requirement set to 1 in table1)
- G2 to medium requirements services; Real Time Requirement column equal to 2
- G3 to lowest requirements services; Real Time Requirement column equal to 3

In our model [5] each group customer can choose a level (high, middle, low) of SLA. The "*Service Level Agreements*" between the operator and the client is implemented as a contract which states the agreed characteristics for each level of service like bandwidth, latency and corresponding bill. We use the classification of Table1;

Table 1. Classification groups

Application	**Audio**	**Video**	**Data**	**Real Time Requirement**	***Groups***
WWW	---	---	X	2	*G2*
IP Telephony	X	---	---	1	*G1*
Multimedia Conference	X	X	X	1	*G1*
Audio Streaming	X	---	---	2	*G2*
Video Streaming	X	X	---	2	*G2*
File Download	---	---	X	3	*G3*
Electronic Mail	---	---	X	3	*G3*
Multimedia Mail	X	X	X	3	*G3*
E-commerce	---	---	X	1	*G1*
Service on demand	X	X	X	2	*G2*

4. SCHEDULING ALGORITHM

4.1 Related Work

Each network node has a scheduler. His role is to define the order of packet transmission. It determines which packets are selected for transmission. There are different approaches of scheduling:

- FIFO (First In First Out): packets are queued and served in order of their receiving. In case of a burst, the queue can be found in overload and arrival packets can be discarded without distinction of their traffic type [7][11].
- PQ (Priority Queuing): the arrival packets are placed in different queues according to their class. Packets of higher class are associated to prior queue. But when there are any packets waiting in a higher queue packets of lower queue will have to wait that can saturate the buffers of low priority [7][8].
- RR (Round Robin): like PQ scheduling packets are stored by class in queues. Then a tourniquet alternates to serve packets from these queues according to their weight [8].
- WFQ (Weighted Fair Queuing): is an equitable sharing algorithm. Arriving packets are classified and placed in their respective queues. Bits of the packets are served in a circular fashion [8][9].
- EDF (Earliest Deadline First): assigns priority according to deadline of request. The assigned priority is higher for the tasks for which the deadline is shorter. [8][10][11]. AS we consider the delays of our processes and then the delay of the transmitted packets, this algorithm does not fit exactly our needs as it does not take into account the time spent in the network in addition to the fact that the clocks in the network are not exactly synchronized. This scheduling policy is optimal only in centralized context meaning that if scheduling is possible, EDF will provide one possible scheduling. Reciprocally, this means also that if EDF does not provide a scheduling, there will be no possible scheduling.

4.2 Delay Impact on the Utility Scheduling (DUS)

We consider the service delay utility (U_{DL}) according to subscriber's contract is the difference between the agreed SLA delay and the duration spent by service data transport in the network.

we propose in DUS that each network node has three tail queues: Q_1, Q_2 and Q_3. This approach is already used, for example by PQ scheduling. The first queue for services is belonging to G1 with the highest priority. The second queue for the services of G2. The third queue for those of G3. In addition to classifying packets by priority level, we sort packets according to theirs U_{DL} in decreasing order. Hence we give priority to packets that will exceed their SLA delay if we do not act. Then we combine priority of PQ, weight of RR and delay utility ranking inspired from EDF:

- Each queue has a capacity L_i, i=1, 2, 3.
- λ_k is the arrival rate of packets of group k, $k \in \{1,2,3\}$
- We assume all packet have same size
- Each tail queue k has weight ω_k, in our first experimentations we consider a static weight of each queue (DUS-SW). In a second stage the weight of priority queue will be dynamically changed depending on U_{DL} (DUS-DW).
- The average of processing time in the server is $1/\mu$

- The average service rate of a queue $1/\mu_k$; $\mu_k = \frac{\mu * \omega_k}{\sum_{i=1}^{3} \omega_i}$

4.3 DUS-Static Queue Weight

We note:

- N: total number of packets waiting on the node $N = N1 + N2 + N3$
- N_i is the number of packets in Q_i
- We vary the number of packets waiting in the node and we calculate the service time for each group
    - For three groups service time in FIFO is equal to $(N+1)/\mu$
    - The maximum service time ($DUS_{max}SW$) is $(Nk+1)/\mu_k$, when the value of U_{DL} is the highest the packet is scheduled at the end of the queue.
    - The minimum service time ($DUS_{min}SW$) is $1/\mu_k$, when value of UDL is the smallest, the packet is scheduled in the top of the queue.

4.4 DUS-Dynamic Queue Weight

In this section we propose a dynamically change for the weight w of the prior queue focus on U_{DLmax}, U_{DLmin} and queue load:

$$w = w_f * IMRU_{DL} * IMR_{Load}$$

- w_f is a fixed weight of queue
- $IMRU_{DL}$: The inverse of margin rate utility

$$IMRU_{DL} = \frac{U_{DLmax}}{(U_{DLmax} - U_{DLmin})}$$

As shown in figure 2 $IMRU_{DL}$ increases if delay utility of packet on tail closes to delay utility of packet on front of queue; U_{DLmin} is the U_{DL} of packet in front of queue and U_{DLmax} is the U_{DL} of packet in tail of queue . When U_{DLmax} closes to zero $IMRU_{DL}$ increases also. Then if all waiting packets of prior queue ($w_f * IMRU_{DL}$) will always be greater than w_f that will improve service times of packets.

- IMR_{Load} is the inverse of margin rate load, N_{max} is the total capacity of queue:

$$IMR_{Load} = \frac{N_{max}}{N_{max} - N}$$

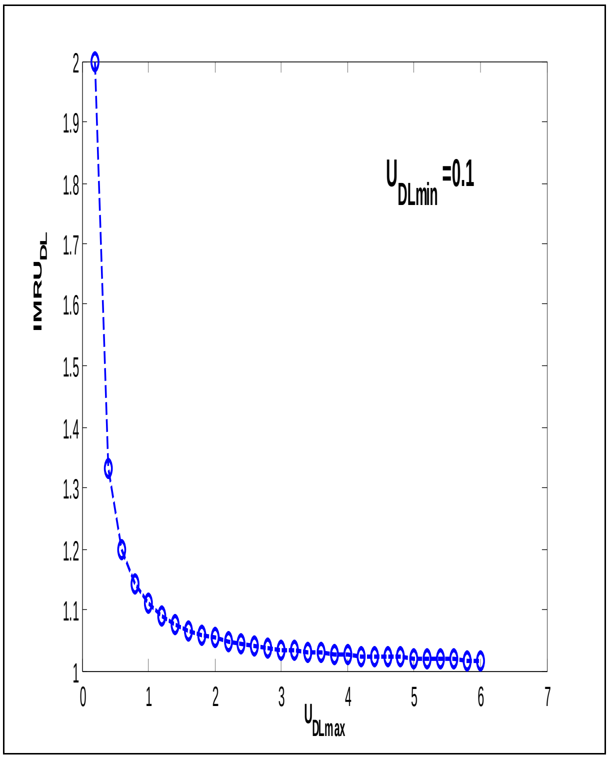

Fig 2: Increase of $IMRU_{DL}$ when all waiting packets have shortest U_{DL}

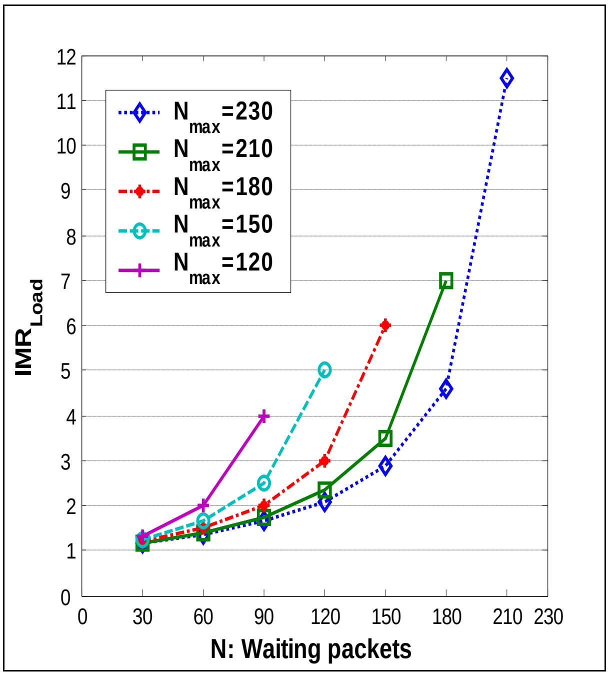

Fig 3: Increase of IMR_{Load} when number of waiting packets will saturate queue

Nmax is the total capacity of queue. Figure 3 shows that IMR_{Load} increases when load (N) of queue increase. This will increase the weight when the number of packets waiting increases more and more. In addition IMR_{Load} increases if Nmax decrease. Then the weight depend on capacity of queue.

4.5 Results Discussion

We assume that:

- N1=N2=N3=(1/3)*N
- w1=8; w2=6 and w3=2
- μ=3(packets/s)
- μ1=(μ*w1)/(w1+w2+w3);
- μ2=(μ*w2)/(w1+w2+w3)
- μ3=(μ*w3)/(w1+w2+w3);

In a first phase we vary the number of packets waiting in the node from 30 packets to 210 packets and we calculate the service time for each group with fixed weight:

- Figure 4 and figure 5 show a decrease in service time of groups 1 and 2 which have real time requirements. The results of FIFO algorithm are greater than highest service time in our algorithm DUS-SW.

- Figure 6 shows that with FIFO group 3 may have service time less than with DUS, but this does not present problem because Q3 is not real time service.

In a second phase, figure 7, we vary number of packets waiting in the node from 30 packets to 210 packets and we calculate:

- The maximum time service for DUS with static weight

- The maximum time service for DUS with dynamic weight.

- So we find in figure 7 that service time with DUS_{max}-DW is better than DUS_{max}-SW. because like explain in section 4.4 , dynamic weight increase firstly if U_{DL} of all waiting packet are close, secondly if number of waiting packet is close to total capacity load and if queue capacity decrease.

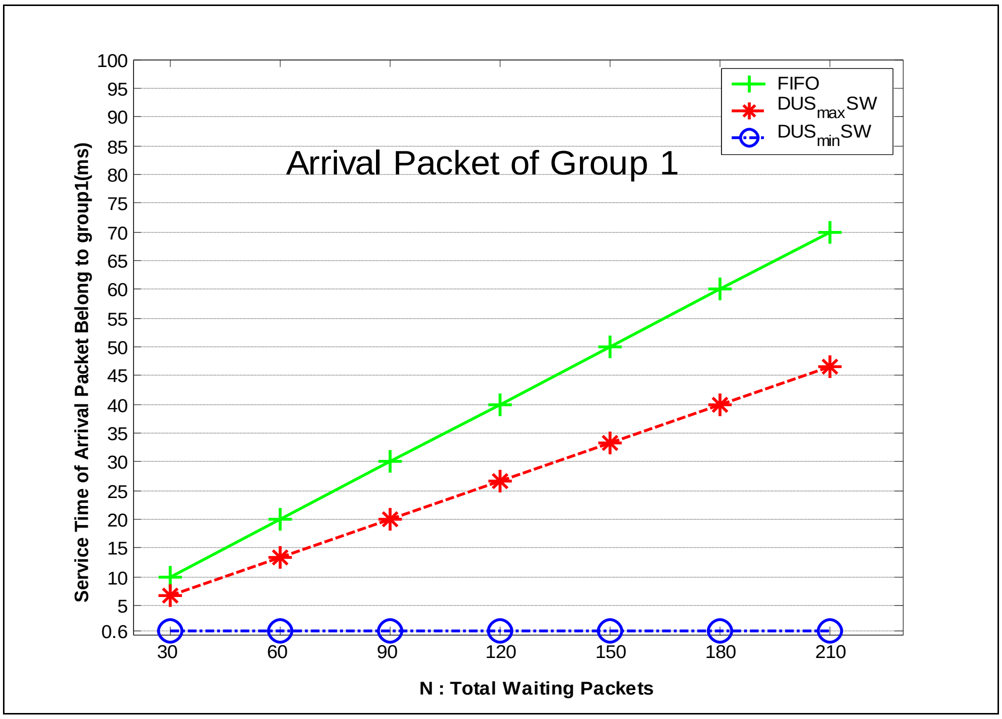

Fig 4: Comparison of service time of G1 arrival packet

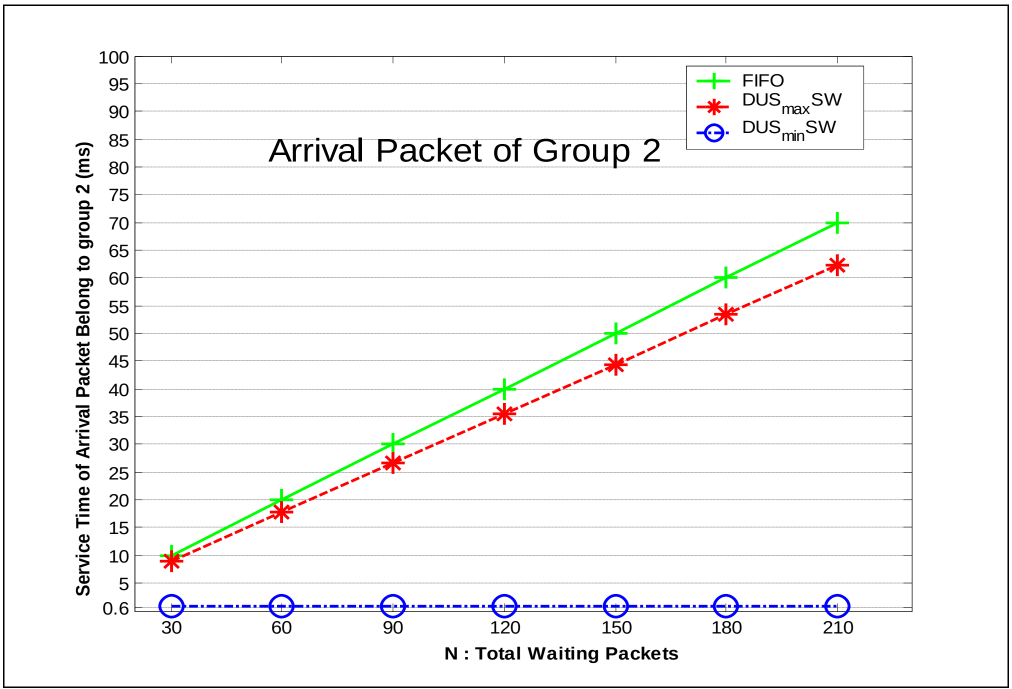

Fig 5: Comparison of service time of G2 arrival packet

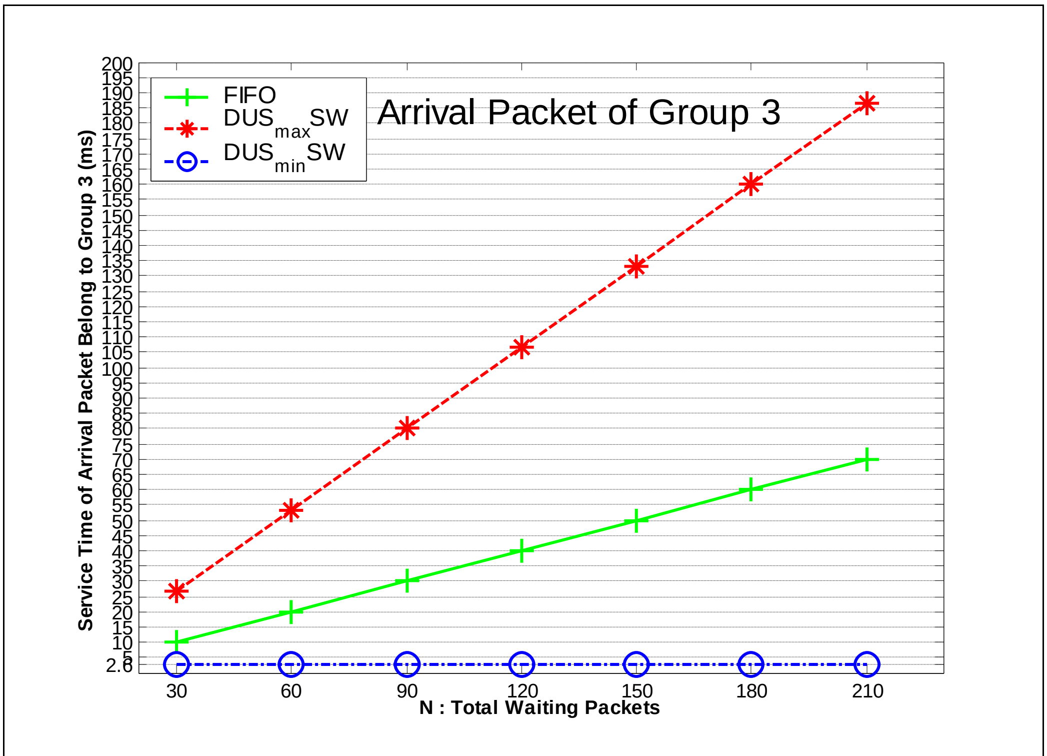

Fig 6: Comparison of service time of G3 arrival packet

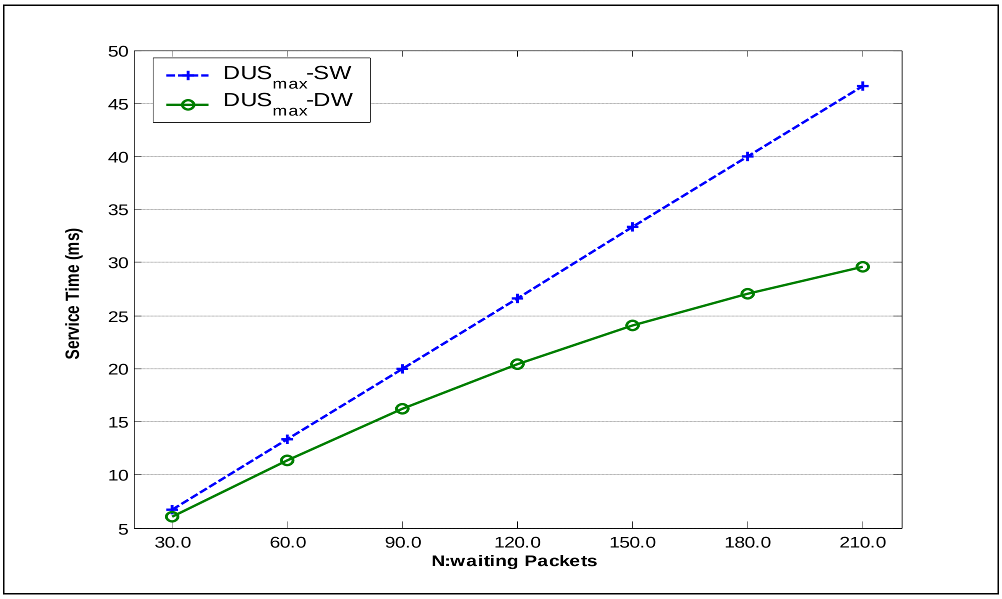

Fig7: Improvement of service time by dynamic weight

5. OVERLAY ROUTING ALGORITHM

Network virtualization was used as evaluation tools. Also the role of virtualization in network can be the separation of policy from mechanism [12]. Network virtualization consists on overlay that is a logical network built on top of one or more existing physical network [13].

OSPF is a link-state routing protocol. Each OSPF router calculates a routing table by constructing a shortest-path tree based on the minimum cost of link.

$\mathcal{N}$ is the set of node constituting a path between a source node S and a destination nod. $\mathcal{L}$ is the set of edges between nodes. The total time between S and D is TT=$\sum_{i\in\mathcal{L}} T_{p_i} + \sum_{j\in\mathcal{N}} T_{s_j}$ with T_p is the propagation time of edge and T_s is service time of node. We propose to use overlay and OSPF algorithm to have the minimum TT between S and D. because we cannot use OSPF with two criteria; Tp_{min} and Ts_{min}. We separate information into two overlays. On the first overlay we find path between S and D using OSPF algorithm according to Ts_{min}. On the second the OSPF criterion is Tp_{min}.

In figure 8 S is the node 1 and D is the node 3 we built some overlays:

- Overlay 1 contains only information about nodes service times of application group
- Then Overlay 3 contains a $Path1_{1\text{-}3}$ that is resulted of OSPF algorithm according to Ts_{min} (minimum service time). $Path1_{1\text{-}3} = Min(\sum_{j\in\mathcal{N}} T_{s_j})$
- Overlay 2 contains only information about links propagation times
- Then Overlay 4 contains a $Path2_{1\text{-}3}$ that is resulted of OSPF algorithm according to Tp_{min} (minimum propagation time). $Path2_{1\text{-}3} = Min\ (\sum_{i\in\mathcal{L}} T_{p_i})$
- In Overlay 5 we stack Overlay 4 and Overlay 3 and we chose portion having Min(TT) between each two common node of Overlay 5. The solution will be one path of this on figure 9. Overlay 6 contains the optimum path between S and D.

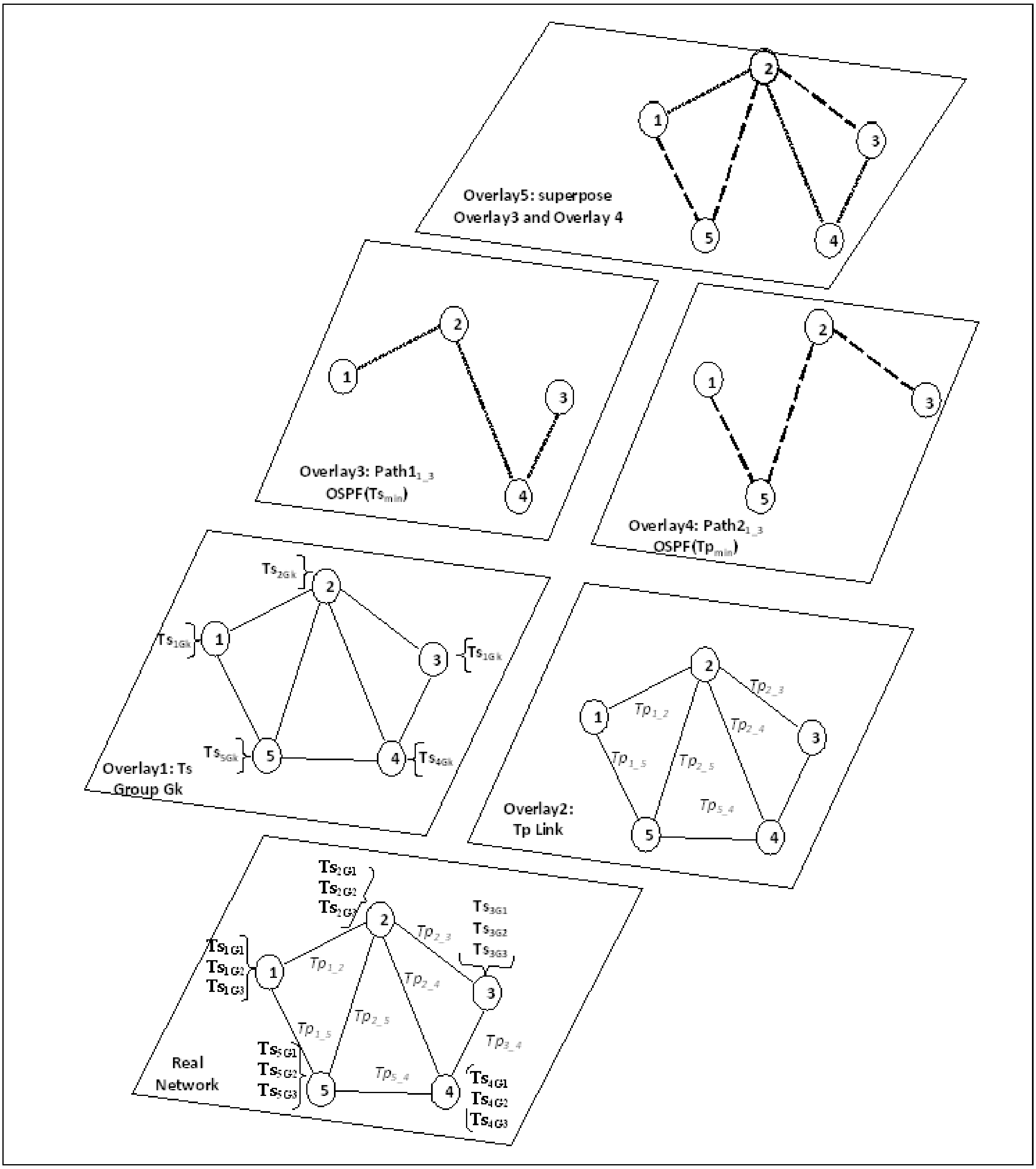

Fig8: Construction of overlays to use OSPF with two criteria

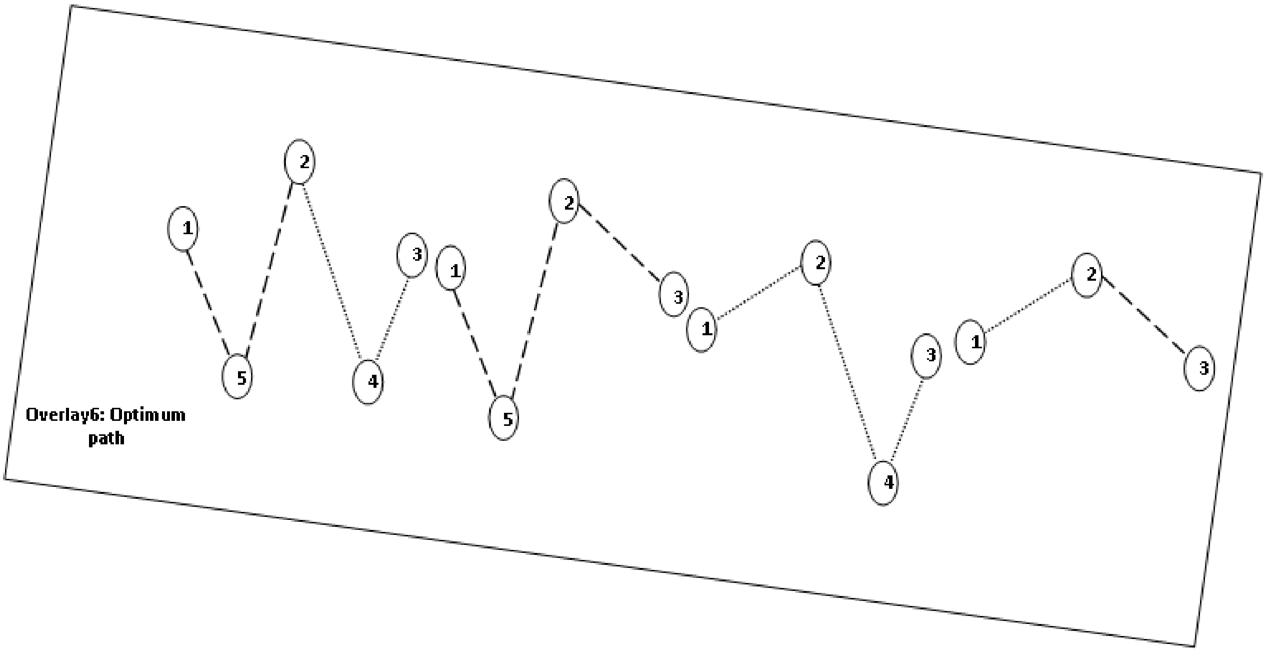

Fig9: Optimum path with OSPF according two criteria

6. CONCLUSION

In IP based network, telecommunication service is reliable if its QoS criteria are satisfied. In this paper a *Delay utility Scheduling* is presented and discussed. According to experimental results in Section 4.5, our proposition improves service times in different nodes. This work defines also an *Overlay Routing Algorithm* in which we propose to separate information in overlays and use OSPF according to two criteria such us propagation time and service time. Our solution gives the optimum path between source and destination.

7. REFERENCES

[1] S. M. Chadchan and C. B. Akki, Priority-Scaled Preemption of Radio Resources for 3GPP LTE Networks, International Journal of Computer Theory and Engineering , 2011.

[2] S. M. Chadchan, C. B. Akki, Quality of Service Provisioning in 3GPP EPS, Journal of Telecommunications Management, April 2010.

[3] J.C Laprie, J.C, Guide de la sûreté de fonctionnement, Cépadus-éditions, 1998.

[4] P.G.S James, H. David, K. Ç. Egemen, J. Abdul, P.R Justin, S. Marcus and S. Paul, Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines, Science Direct Journal, 2010.

[5] O. Ben Ahmed and Z. Choukair, Adaptation AC to Improve Wireless Access Network Dependability, International Journal of Computer Science & Information Technology, 2011.

[6] T. Janevski, Traffic analysis and design of wireless IP networks, ISBN 1-58053-331-0, 2003.

[7] C. Taddia and G. Mazzini, On the Jitter Performance of FIFO and Priority Queues Mixture, IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, 2006.

[8] W. Lan, M. Geyong, D. Kouvatsos and J Xiaolong, An analytical model for the hybrid PQ-WFQ scheduling scheme for WiMAX networks, the First International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009.

[9] J. Yong , W. Jianping and S. Xiaoxia, A packet scheduling algorithm in high performance routers, 4th IEEE International Conference on ATM and High Speed Intelligent Internet Symposium, 2001

[10] A. Iqbel, A. Zafar and B Sidduque, Dynamic queue deadline first scheduling algorithm for Soft real time systems, IEEE International Conference on Emerging Technologies, 2005

[11] K. Maalaoui and L.A. Saidane, FP/EDF versus FP/FIFO Local Scheduling and Probabilistic QoS Guarantees in a Bluetooth Piconet, IEEE 21st International Conference on Advanced Information Networking and Applications

[12] Y. Zaki,. Z. Liang, C. Goerg, and A. Timm-Giel, LTE wireless virtualization and spectrum management Wireless and Mobile Networking Conference (WMNC), 2010 .

[13] N.M.M.K.; Chowdhury, R. Boutaba, Network virtualization: state of the art and research challenges IEEE Communications Magazine, 2009.