International Civil Aviation Organization

Fifth Meeting of Aeronautical Telecommunication Network (ATN) Transition Task Force of APANPIRG

Phuket, Thailand, 9-13 June, 2003

Agenda Item 9: Review ATN Implementation activities/issues

A NEW FLOW CONTROL ALGORITHM IN ATN BASED ON ISIS PROTOCOL

SUMMARY

Abstract: ISIS is used in ATN as intra-domain routing protocol, which may create multiple equal-cost paths for a single destination. ISIS-ECMP and ISIS-OMP are methods to control the distribution of a flow based on the multiple equal-cost paths, but they can not distribute the flow according to the network load and link state accurately and timely. In this paper, a new dynamic flow control algorithm is proposed based on the equal-cost paths within an ATN area. The simulation results show that our new algorithm can adjust the flow effectively and prevent Intermediate Systems from getting congested. On the other hand, our algorithm is also a contribution to solve the problem of the minimum credit window in ATN congestion control.

Keyword: ATN, ISIS, ECMP, OMP, CLNP, LSA, RD

1. Introduction

1.1 Aeronautical Telecommunication Network (ATN) is an important network infrastructure for future civil aviation. It is composed of subnetworks, End Systems (ESs), and Intermediate Systems (ISs). The subnetworks used in ATN are X.25 subnetworks, and mobile subnetworks which include VHF subnetworks, HF subnetworks, AMSS subnetworks, MODE S subnetworks. ESs have complete seven layer protocols defined by ISO7498, and ISs, which are also called ATN routers, have the underlying three layer protocols. In network layer, CLNP PDUs are relayed by ISs to their destinations. There are three basic routing protocols used in ATN to advertise routing information. They are IDRP, ISIS, ESIS.

1.2 ISIS is the routing protocol used in a Routing Domain in ATN. The routing domain is divided into several areas. There are four routing metrics used in ISIS. They are Default metric, Delay metric, Expense metric, and Error metric. Within each area, all the ISs flood Link State PDUs (LSPs) periodically, and all ISs create the same Link State Database based on these LSPs. Then, each IS uses the Dijkstra algorithm to create a new forwarding database for each metric and refresh the corresponding old database.

2. The congestion control algorithm in ATN internetwork

2.1.1 The congestion control algorithm based on CLNP and COTP

The algorithm is based on calculating the degree of congestion and decreasing the credit window, which is similar to TCP/IP Explicit Congestion Notification algorithm. The structure of CLNP PDU is illustrated in figure 1:

Fixed	Address Part	Segmentation Part	Options Part	Data

Figure 1 The structure of CLNP PDU

2.1.2 The QoS maintenance parameter is a parameter of the Options Part. The CE bit is located in the value field of the QoS maintenance parameter, which is illustrated in figure 2:

Parameter Code	Parameter Length	Parameter Value
1100 0011	0000 0001	11xx CExxx

Figure 2 QoS Maintenance Parameter

2.1.3 The value of the CE flag is initially set to zero by the initiator of the CLNP PDU and the CE flag of the CLNP PDU will be set to 1 when the CLNP PDU passes a congested queue.

2.1.4 On the receiving transport layer, the transport entity will calculate the CE-setting rate of the PDUs received during the sampling period. If the value exceeds a given threshold, the transport entity will decrease the credit window for the next update period.

2.2 <u>The congestion threshold</u>

2.2.1 To achieve high throughput and low end-to-end delay, the following equation is used in ATN to optimize the power of a connection:

$$\text{Power}=\text{Throughput}/\text{Delay}$$

2.2.2 An output queue in an IS can be viewed as an M/M/1 queuing system. If the packet arriving rate is λ, and the packet forwarding rate is μ, the average time each packet spends in the system is given by:

$$T= 1 / (\mu-\lambda)$$

2.2.3 Since the throughput of the M/M/1 queuing system is equal to λ, the power of the M/M/1 queuing system could be calculated as follows:

$$\text{Power}=\lambda/(1/(\mu-\lambda))= \lambda(\mu-\lambda)$$

2.2.4 Obviously, when $\lambda \square \mu/2$, the Power reach the maximum value. Again, according to the M/M/1 queuing system, the average packet held in the queue is given by:

$N=\lambda/(\mu-\lambda)$

If $\lambda \square \mu/2$, then:

$N=1$

2.2.5 So, if the length of a queue on an IS exceeds one, the CE bit of a CLNP packet in a queue will be set to one to associate the COTP to regulate flow.

2.3 <u>The problem of the minimum credit window</u>

2.3.1 The congestion algorithm described above could bring some problems. The minimum credit window limit is one of the problems that the congestion algorithm can not cope with. When all transport connections through a congested node have reduced their credit window to one and the node is still in a congested state, the algorithm has no way to relieve the congestion further. A lot of packets will be discarded and retransmitted. Since there may be multiple equal-cost paths to a destination within an ISIS area, in our new algorithm, the advantage of the multiple paths is fully considered, packet flows can be distributed according to the down-stream node state, and the problem of the minimum credit window can be partly solved.

3. Dijkstra algorithm, ECMP and OMP

3.1 Dijkstra algorithm is widely used in intra-domain routing protocol. In ISIS, each IS within an area will flood the link state PDU periodically. So, every IS in the area can get the same link state databases for each metric. Through Dijkstra algorithm, the forwarding databases for each system can be established from the information of the link state database. But, it has been widely known that Dijkstra algorithm may result in multiple equal-cost paths for one destination. How can we take advantage of the multiple equal-cost paths fully?

3.2 Equal Cost Multipath (ECMP) is one way to distribute the flow when multi-path appears. It splits the flow evenly between the multiple equal cost paths. There are three method to realize ECMP: (1) round robin forwording (2) dividing destination prefixes among available next hops (3) using hash function to decide which next hop to choose. But link and node load information is not taken into account in ECMP. So, it can not adjust the flow according to the down stream load condition.

3.3 Optimized Multipath (OMP) is another way to distribute the flow among multiple paths. It samples the link load every 15 seconds and floods LSP_OMP_LINK_LOAD Opaque LSPs. But it can not regulate the link load change within 15 seconds. On the other hand, flooding Opaque LSPs increases the burden of the network.

3.4 In this paper, a new flow control algorithm is proposed. It can control the flow distribution according to the down stream node load information and avoid flooding Opaque LSPs, which will burden the network within that routing area.

4. Dynamic congestion control algorithm within an area of an RD in ATN

Some new concepts and the extension of CLNP PDU header

4.1.1 Before introducing our new algorithm, some new concepts will be introduced and the CLNP PDU header will be extended.

4.1.2 Cross Point (CP): A Cross Point is a node in an ATN area that has multiple equal-cost paths to a destination.

4.1.3 Extension of CLNP header: in order to let the downstream node know which node is the nearest CP in the upstream, a CP parameter is added to the option part of a CLNP PDU to record the information of the CP when the CLNP PDU passes through it. Then, the nearest upstream CP can be found by the congested downstream node from the CP parameter in a CLNP PDU, and the Congestion Report PDU will be sent to it. The structure of CLNP PDU is illustrated in figure 1. The CP parameter is defined in figure 3:

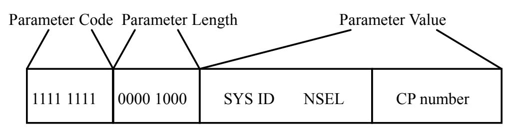


Figure 3 CP Parameter

4.1.4 Congestion Report PDU (CRP): The Congestion Report PDU is used to report the congestion information of the congested node in the downstream to the node that is the nearest Cross Point (CP) in the upstream. The Congestion Report PDU is defined to have the format as illustrated in figure 4, which is similar to CLNP PDU and only contains the fix part, address part, and option part.

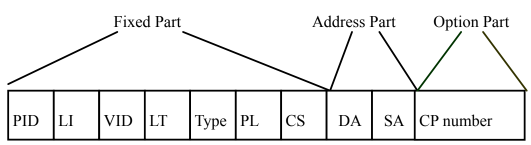


Figure 4 Structure of CRP

The values of the fix part of CRP are shown in table 1.

PID:	Network Layer Protocol Identifier	1000 0001
LI :	Length Indicator	one byte
VID:	Version ID Extension	0000 0001
LT:	Life Time	one byte
Type:		0000 0000 or 0001 0000
PL:	Packet Length	two bytes
CS:	Checksum	two bytes

Table 1 CRP Fix Part

4.1.5 In the type field, 0001 0000 indicates that the downstream node is in the congested state, and 0000 0000 indicates that the congestion in the downstream node is cleared. In address part, DA is the address of the nearest CP in the upstream of the flow and SA is address of the congested node that creates the CRP. Both DA and SA should conform to the NET structure . The option part should include the CP parameter. The CP number marks the path that passes through the congested node as shown in figure 5.

4.2 Algorithm Description

4.1.6 Let us consider a CLNP packet stream to a given destination passing through an ATN ISIS area. If the CLNP packets of the stream do not pass any CPs, the original ATN congestion algorithm will be used unchanged. For those packets that do pass CPs, our following algorithm can regulate the packet flow effectively.

4.1.7 Assuming that a packet stream in an area has the equal-cost path topology shown in figure 5, node 1 and node 4 are the CPs that are defined above.

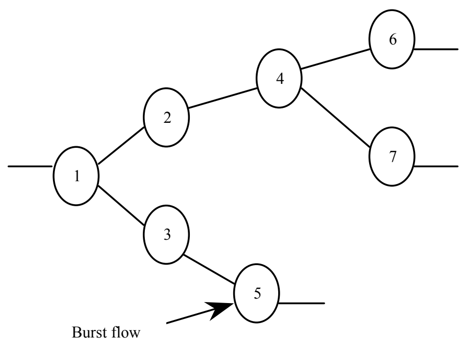


Figure 5 Flow Topology

4.1.8 When a packet of the flow arrive at the first CP in the area, for example, node 1, the CP parameter of node 1 is added to its option part. Through some flow distribution algorithm at node 1, a packet of the flow will arrive at the next CP, node 4. Its CP parameter will be replaced by the CP parameter of node 4, and so on. So, the CP parameter in a packet is always the parameter of the nearest upstream CP. When the packet leaves the last node in the area, its CP parameter will be deleted.

4.1.9 If a queue length exceeds one, the CE bits of packets on that queue will be set to one. On the same time, a CRP packet is sent back to the nearest upstream CP node, which notifies the nearest upstream CP node that the downstream of this path is overloaded. The nearest upstream CP node will lessen this path load and distribute more packet load to other paths.

> Case1: In figure 5, when a burst flow occurs suddenly at node 5, the output queue length will increase and exceed one. Then, when node 5 receives a packet from node 1, it will send back to node 1 a CRP packet to report its congestion. At node 1, through a distribution control algorithm, the flow to node 3 and node 5 will be lessened.

Case2: In figure 5, when node 4 is notified that both the path to node 6 and the path to node 7 are congested, node 4 will report to node 1 that the path from node 1 to node 4 is congested. At node 1, more packet flow will be forwarded to node 3, and so on, until the first CP of the area is notified congestion. Then the first CP will take some traditional measures, such as RED, to reduce the flow.

5. Applications of the algorithm and its simulation analysis

In this paper, the applications of the algorithm are given by modifying the round robin ECMP and hash function ECMP.

5.1 Round robin based simulation:

5.1.1 The round robin ECMP algorithm requires that the multiple equal-cost paths have the similar transit delay. In an ATN routing domain, when the delay difference of the multiple equal-cost paths is small or delay metric is chosen in the CLNP PDU QoS parameter, the round robin ECMP can work perfectly. From the simulation result, it is obvious that our new algorithm can achieve better load distribution.

5.1.2 At every CP, the improved round robin algorithm is used to divide the flow between the multiple paths. Assuming that node 1 in figure 5 has n next hops to its destination, as shown in figure 6, the node can divide the flow evenly between the n paths by using round robin ECMP algorithm.

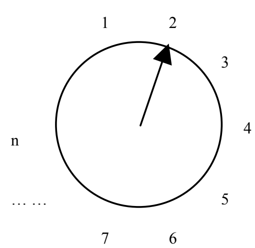


Figure 6 Round robin ECMP

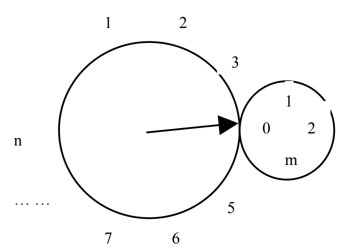


Figure 7 CP state

5.1.3 If node 1 receives a CRP from path 4, it will increase a state in path 4. When m CRPs are received from path 4, the CP will have m state as illustrated in figure 7.

5.1.4 Every time node 1 receives a packet of the flow, it will forward the packet to the path that the arrow points to. And then, the arrow will turn to the next path. If it is the turn for path 4 to forward a packet, path 4 will examine its state first. If it is in state 0, it can forward the packet. If not, path 4 will turn to next state and the arrow will move to the next path to forward the packet.

5.1.5 The simulation flow topology is shown in figure 8. Because in an ATN area, ISIS OMP only flood opaque LSP once every 15 seconds, the flow topology remains unchanged during the 15 seconds. So both OPM and ECMP can do nothing about the flow change within 15 seconds. Also, the COTP can do nothing to relieve congestion when connection credit window s have all released to one.

In the simulation, three assumptions are assumed:

(1) All COTP connections have reduced their credit window to one.
(2) There are only two flow streams in the topology: packet flow from node C to node D and packet flow from node A to node D.
(3) There are no delay in each node.

5.1.6 The simulation keeps 6 seconds. In the 6 seconds, OMP can not help solving the problem of load balancing. The output queue length of the node R2 when our new algorithm is used and that of the node 2 when the traditional round robin ECMP is used are compared. Two conditions are considered: no outburst flow appears and outburst flow does appear.

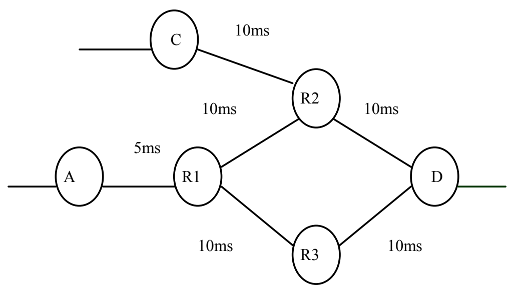


Figure 8 Flow Topology

(1) No outburst stream:

Packet arriving rate at A: λ=0.1

Average arrival time interval at A: T=10s

Packet arriving rate at C: λ=0.0625

Average arrival time interval at C: T=16s

The simulation result is illustrated in figure 9 and figure 10:

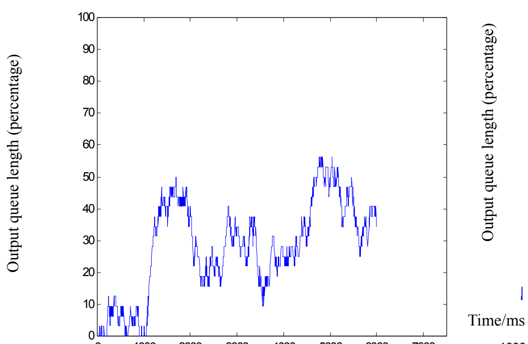


Figure 9

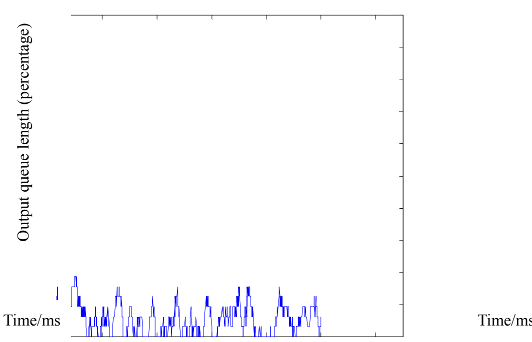


Figure 10

5.1.7 Therefore, when our dynamic flow control algorithm is not used, the length of output queue is inceasing slowly as shown in figure 9. When our new flow control algorithm is used, the percentage of the length of output queue is controlled within 20% as shown in figure 10.

(2) outburst stream appears in flow C-R2-D:

Packet arriving rate at A: $\lambda=0.1$

Average arrival time interval at A: T=10s

Packet arriving rate at C: $\lambda=0.083$

Average arrival time interval at C: T=12s

The simulation result is illustrated in figure 11 and figure 12:

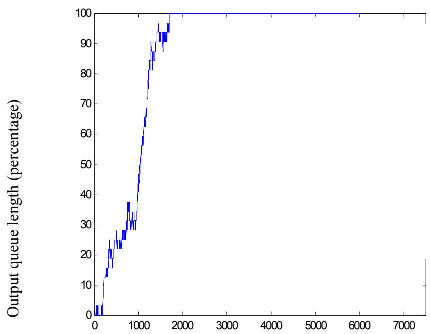
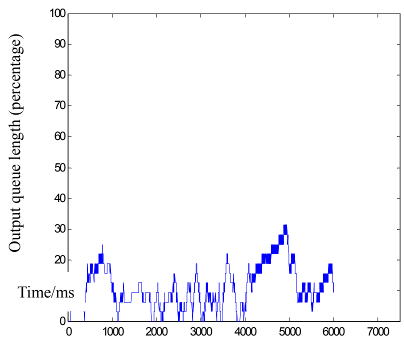


Figure 11



Time/ms

Figure 12

5.1.8 When outburst stream appears, ECMP, OMP, and the traditional ATN congestion control algorithm can not prevent the output queue from congestion. Packets are dropped due to the congestion. Yet, in our new algorithm, we controlled the queue timely and no packet was dropped when the outburst appears.

5.2 <u>Hash function based simulation:</u>

5.2.1 Hash function ECMP is widely used in link state algorithm routing protocol. Based on hash function ECMP, the dynamic flow distribution algorithm described above can also regulate the flow effectively.

5.2.2 Assuming that there are three equal cost paths in a Cross Point, as illustrated in figure 13, the CRC range is divided evenly by the three paths. CP router calculates the CRC value of the source address and destination address of a received packet by hash function, and finds a path that the CRC value belongs to. Then, the packet will be forwarded from that path.

5.2.3 When the node receives a CRP from path one, the path one will be marked once. If the node receives a packet that should be forwarded via path one, then the following equation should be used:

$$\alpha[2]= (\alpha[1]*k) \bmod 65536$$

where k is the number of the paths, $\alpha[1]= \alpha$, α is the original CRC value, $\alpha[2]$ is residue of the product of $\alpha[1]$ and k divided 65536. The CP router will forward the packet according to the value of $\alpha[2]$.

When more CRPs are received from path one, the following equations should be used:

$$\alpha[3] = (\alpha[2]*k) \bmod 65536$$

$$\dots \quad \dots$$

$$\alpha[i+1]=(\alpha[i]*k) \bmod 65536$$

where i is the number of the CRPs received. During the process of calculation, if $\alpha[j]$ ($j<k$) is in the range other than path one, the packet will be forwarded via that path. As a result, the packets that should be forwarded via path one are forwarded via all the other paths. The more CRPs the CP rerceives, the little packets the CP will forward via path one.

5.2.4 When more equal cost paths exist, as illustrated in figure 14, the algorithm described above can still work.

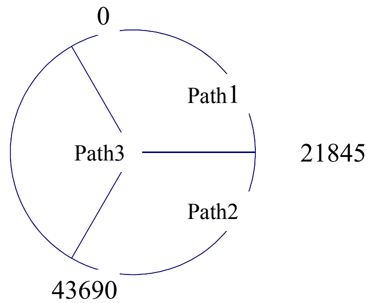


Figure 13

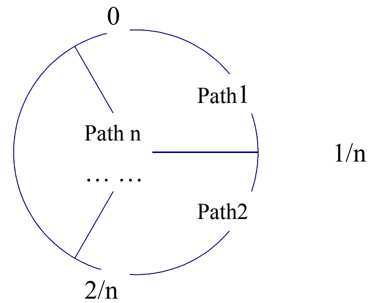


Figure 14

The simulation flow topology is illustrated in figure 8. The same simulation conditions are assumed.

(1) No outburst stream:

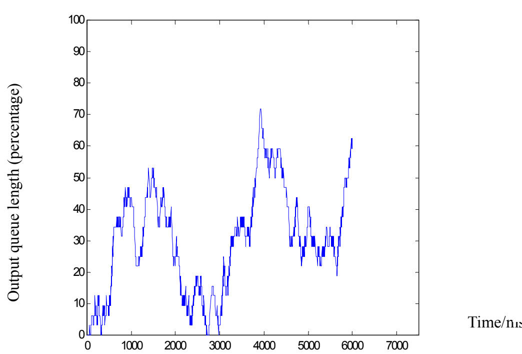
Packet arriving rate at A: $\lambda=0.1$

Average arrival time interval at A: T=10s

Packet arriving rate at C: $\lambda=0.0625$

Average arrival time interval at C: T=16s

The simulation result is illustrated in figure 15 and figure 16:



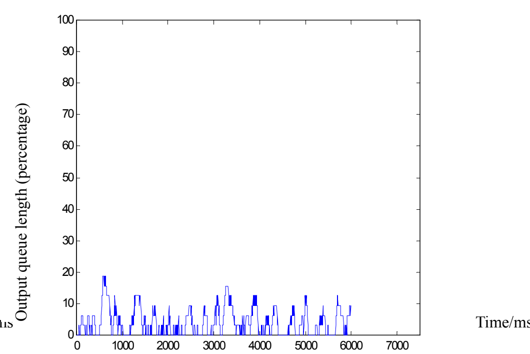


Figure 15 Dynamic flow

Figure 16 Dynamic flow

5.2.5 It is obviously that our dynamic flow control algorithm can control the queue length effectively and prevent the queue length from increasing.

(2) outburst stream appears in flow C-R2-D:

Packet arriving rate at A: λ=0.1

Average arrival time interval at A: T=10s

Packet arriving rate at C: λ=0.083

Average arrival time interval at C: T=12s

The simulation result is illustrated in figure 17 and figure 18:

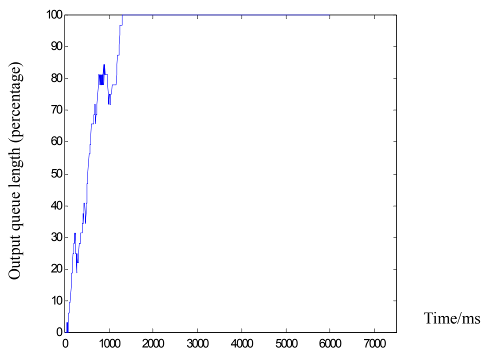


Figure 17 Dynamic flow

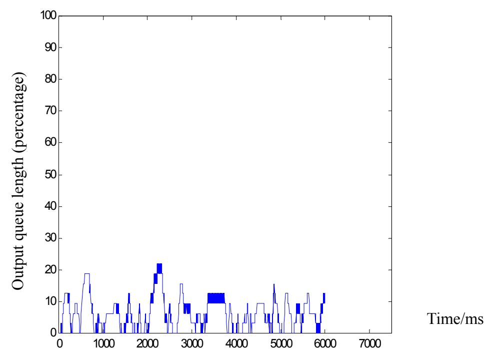


Figure 18 Dynamic flow

5.2.6 When our new algorithm is used in hash function ECMP, the queue length can be controlled and packet dropping can be prevented.

6. Conclusions

6.1 The dynamic flow control algorithm introduced in this paper has taken full advantage of the multiple equal-cost paths within an ATN area, and can distribute the flow according to the downstream load timely. From the simulation result, it is obvious that our new dynamic algorithm can do better than ECMP and OMP algorithm in ATN, and can partly solve the problem of minimum credit window in ATN congestion algorithm.

6.2 Nowadays, hash function ECMP algorithm is widely used in OSPF. Our improvement to hash function ECMP will also be beneficial to TCP/IP network.