Summary Information for LANs at Level 1 in IS–IS

Keith Sklower

Computer Systems Research Group
Department of Electrical Engineering and Computer Science
University of California
Berkeley, California 94720

ABSTRACT

A proposed minor addition to the IS–IS protocol was submitted to ANSI committee X3S3 in May of 1990, and was subsequently redistributed to other ISO members committees.

The proposed change may substantially reduce the demands of running the protocol to make it feasible to implement on very inexpensive machines and slow-speed links.

Included is a copy of the original submission, despite its colloquial rather than academic style, as a courtesy to anyone who is interested in reviewing the proposal in a timely fashion.

The author gratefully acknowledges the support of the National Institute of Science and Technology (NIST), award number 60NANB8D0830.

September 12, 1990

Summary Information for LANs at Level 1 in IS–IS

Keith Sklower

Computer Systems Research Group
Department of Electrical Engineering and Computer Science
University of California
Berkeley, California 94720

ABSTRACT

This proposes a minor addition to the IS–IS protocol which may substantially reduce the demands of running the protocol to make it feasible to implement on very inexpensive machines and slow-speed links.

1. Overview

We propose to allow routers to exchange summary information for classes of end-systems on a LAN. It is intended that the summary information replace (or greatly reduce the size of) the lists of end systems attached to a pseudo-node. However, routers could continue to advertise specific end systems in addition to the summary information.

The proposal would require a modification to the forwarding algorithm, which is only an addition to the case where the previously described lookup fails.

The proposal would modify the format of a Link State PDU (LSP) only by appending an additional optional variable length field, which would only appear in pseudo-node LSP's.

The proposal specifies remedial actions in case of a LAN partition, allowing recovery once the partition heals.

2. Example

It is our understand that nothing in the current IS–IS document prevents manual configuration of End System ID's (ESIDs). Since, over the course of the next two or three years, UC Berkeley is likely to continue running TCP/IP concurrently with emerging OSI stacks, it will have to continue to assign IP host numbers. Thus it could merely embeded the IP host number as the lower 32 bits of the 48 bit ID field.

Several ethernets at Berkeley have as many as 150 end systems on them; the proposal is to provide some way to convey that all of the end systems on a particular ethernet have IDs which (when written as a sequence of octets) are of the form:

0, 0, 0x80, 0x20, 0x96, <any>

without eplicitly enumerating each one. By doing so, we could compress out 900 octets of ESIDs from the LSP for that pseudo-node.

We do this by communicating an ID Prefix Length and ID Prefix as summary information by pseudo-node. Here the ID Prefix would be the octet sequence 0, 0, 0x80, 0x20, 0x96, 0 and the ID Prefix Length would be 40 (bits). Readers familiar with IP subnets should be familiar with this way of thinking.

3. Changes to the Forwarding Algorithm

The construction of the database PATHS (as specified in the current IS–IS document) proceeds exactly as before. The only difference is the determination of which pseudo-node is associated with a given end system. We require that an attempt be made in the same manner previously to locate an end system ID among all ESIDs explicitly enumerated by all the LSP's.

Only when that fails, do we attempt a match based on the summary information. If all the ID Prefix Lengths are the same, one could extract the leading bits of the ESID to be routed, append trailing zeroes, and then presumably hash, if that was the manner in which one looked up the pseudo-node from the ESID in the normal case. If not, one could keep a sorted database of ID Prefix and find the most specific match by comparison searching.

Note that it is still possible to determine that a host is down (or unreachable). This is done by the router advertising the summary information (as it will remember information gleaned from ESH's), rather than the first route upon entry into an area.

4. Partition Detection

If a level 1 router receives a LSP containing summary information exactly the same as its own, it should immediately send out a new LSP listing each system ID separately. If this has resulted either from misconfiguration or partition, the other router will eventually respond listing its census of associated end systems.

Thus, since exact matches are preferred, end systems which are up and present will be reachable in the ordinary way. Packets destined for end systems which have crashed or are temporarily unreachable could be detected by any router in the area which has noticed the fact of duplicity of summary information and the presence of complete census information. Even if a (non-involved) router choose to send a packet arbitrarily to either of the routers originating the summary information, either of these "last hop" routers would then have complete information and be able to send a host unreachable message.

If the LSP with duplicate summary information is due to a router which has crashed or stopped acting as a designated router, there will eventually only be one router (one set of LSPs) advertising the given summary information.

5. Structure and Encoding of PDU field

(This section is written in a style to be inserted directly into the current IS–IS draft, so it uses the language of that document without further explanation.)

– LAN Summary – ESIDs presumed on LAN

This code only appears in the Link State PDU – Level 1 generated on behalf of a pseudo-node. It summarizes a class of 6 octet system ID's which may be assumed to be found on the LAN being represented. This code may be present in addition to the End system Neighbors code. We can think of no reason to prohibit multiple summaries by LAN, although that is contrary to our intention to compress protocol information as much as possible.

* CODE – <To be assigned by editor>
* LENGTH – 2 plus a multiple of 7 octets.
* VALUE –

	No. of Octets
ID PREFIX LENGTH	1
ID PREFIX	6

x The ID Prefix Length gives the number of leading bits that many End System IDs have in common on the LAN represented by this pseudo-node.

x The Id Prefix specifies the value of those leading bits; the trailing (48 - ID Prefix Length) bits are required to be zero.

It further simplifies execution of the protocol if all ID Prefix Lengths are the same for all pseudo-nodes within a given area.

6. Expected Impact

The author of this modest proposal has spoken with employees of DEC and agrees that it is possible to build dedicated routers that can implement the protocol as is with no additions with off the shelf parts that should sell for the $15,000 range. Furthermore we conceed that the protocol, as is, takes up negligble bandwidth on 10 megabit LANs and T1 lines.

However, that can represent a lot of money of to colleges and universities. We think that colleges and small universities can begin to enter an OSI internetwork with hosts which can also function as routers, with the purchase of an additional network interface. The US Government expect to buy computers with I386 processor chips, with 4 Megabytes of memory, and 300 MB of disk for $2500, quantity in the hundreds of thousands or millions. We'ld like to have computers such as these communicate over 9600 baud modems or maybe even 64K ISDN connections.

We believe that the IS–IS protocol should be able to handle the Berkeley Campus as a single area, and the entire BARRNET (Bay Area Regional Reserach NET) as a single routing domain. Berkeley currently has 3000 hosts on about 80 ethernets. It is reasonable to expect those numbers to double in the next 3 to 5 years. (6000 hosts on 160 ethernets) By adopting this proposal, we reduce 36K octets of end system IDs, to 1.5K octets of summary ID. That's a difference of about 40 seconds at 9600 Baud.

By reducing the number of routing entries from 6000 to 160, we can also afford to use a single general next-hop-forwarding algorithm which can be shared among many protocols (such as XNS, IP, and OSI).

Let us consider having such simple routers participate in what DEC considers a large routing domain (~100000 hosts), we could imagine something like 300 hosts per lan among 350 lans. An employee of DEC counted a minimum of 10 octets per end system ID (we really think 2 4-byte pointers instead of 2 2-byte pointer would be required for more than 65K hosts); this means that a minimum of a megabyte of kernel memory would be required, and our general purpose operating system also requires that this be duplicated at user-level (although some of it could be paged out). This represents a substantial fraction of the memory of our cheap boxes. Transmitting 600K octets at 9600 baud takes a very long time (11 minutes vs 30 seconds for 350 summaries of ~50 octets total for the L1LSP). Transmitting 600K octets at 64K baud is still substantial. (A minute and 15 seconds, maybe).

7. Alternatives

While we've chosen our proposed new field to have the minimum possible length, we could live with some other possibilites. One might consider having a mask and a value rather than a prefix length.

Another alternative might be to assign a separate area number to each LAN (thus stealing bits from the area-area rather than the ID portion of the address space). We could live with this second proposal, although it seems to us that it would require changes at level 2 -- somehow communicating that a group of areas were really just 1, and the change we've proposed above require no changes at the level 2 version of the protocol.