(54) Title: INTERCONNECTED TRUNK CLUSTER ARRANGEMENT

(57) Abstract

An arrangement of trunk clusters (10) and a method for interconnecting trunk clusters wherein the interconnection method has no single point of failure, the bandwidth between trunk clusters is not limited by the throughput of a single switch, and faults are contained within each trunk cluster. A trunked interconnection structure is provided between trunk clusters. Each switch (20) of a trunk cluster has a logical port connected to a trunked port (30). The trunked port or trunk port provides a physical connection to each switch of another trunk cluster. Each trunk switch of the other trunk cluster has a logical port connected to a trunked port which in turn has physical connections to each switch of the first trunk cluster. Trunked interconnect isolates faults to a single trunk cluster and there is no single point of failure and the total throughput is not limited to any single switches capacity. This always provides a single loop free path from one trunk cluster to the other or others. Multiple trunk clusters may be interconnected using point-to-point connections. A high throughput campus interconnect trunk cluster can be used to connect each building data center trunk cluster.
# FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<table>
<thead>
<tr>
<th>Code</th>
<th>Country</th>
<th>Code</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>Albania</td>
<td>ES</td>
<td>Spain</td>
</tr>
<tr>
<td>AM</td>
<td>Armenia</td>
<td>FI</td>
<td>Finland</td>
</tr>
<tr>
<td>AT</td>
<td>Austria</td>
<td>FR</td>
<td>France</td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
<td>GA</td>
<td>Gabon</td>
</tr>
<tr>
<td>AZ</td>
<td>Azerbaijan</td>
<td>GB</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>BA</td>
<td>Bosnia and Herzegovina</td>
<td>GE</td>
<td>Georgia</td>
</tr>
<tr>
<td>BB</td>
<td>Barbados</td>
<td>GH</td>
<td>Ghana</td>
</tr>
<tr>
<td>BE</td>
<td>Belgium</td>
<td>GN</td>
<td>Guinea</td>
</tr>
<tr>
<td>BF</td>
<td>Burkina Faso</td>
<td>GR</td>
<td>Greece</td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
<td>HU</td>
<td>Hungary</td>
</tr>
<tr>
<td>BJ</td>
<td>Benin</td>
<td>IE</td>
<td>Ireland</td>
</tr>
<tr>
<td>BR</td>
<td>Brazil</td>
<td>IL</td>
<td>Israel</td>
</tr>
<tr>
<td>BY</td>
<td>Belarus</td>
<td>IS</td>
<td>Iceland</td>
</tr>
<tr>
<td>CA</td>
<td>Canada</td>
<td>IT</td>
<td>Italy</td>
</tr>
<tr>
<td>CF</td>
<td>Central African Republic</td>
<td>JP</td>
<td>Japan</td>
</tr>
<tr>
<td>CG</td>
<td>Congo</td>
<td>KE</td>
<td>Kenya</td>
</tr>
<tr>
<td>CH</td>
<td>Switzerland</td>
<td>KG</td>
<td>Kyrgyzstan</td>
</tr>
<tr>
<td>CI</td>
<td>Côte d'Ivoire</td>
<td>KP</td>
<td>Democratic People's Republic of Korea</td>
</tr>
<tr>
<td>CM</td>
<td>Cameroon</td>
<td>KR</td>
<td>Republic of Korea</td>
</tr>
<tr>
<td>CN</td>
<td>China</td>
<td>KZ</td>
<td>Kazakhstan</td>
</tr>
<tr>
<td>CU</td>
<td>Cuba</td>
<td>LC</td>
<td>Saint Lucia</td>
</tr>
<tr>
<td>CZ</td>
<td>Czech Republic</td>
<td>LI</td>
<td>Liechtenstein</td>
</tr>
<tr>
<td>DE</td>
<td>Germany</td>
<td>LK</td>
<td>Sri Lanka</td>
</tr>
<tr>
<td>DK</td>
<td>Denmark</td>
<td>LR</td>
<td>Liberia</td>
</tr>
<tr>
<td>EE</td>
<td>Estonia</td>
<td>LS</td>
<td>Lesotho</td>
</tr>
<tr>
<td>LT</td>
<td>Lithuania</td>
<td>LU</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>LV</td>
<td>Latvia</td>
<td>MC</td>
<td>Monaco</td>
</tr>
<tr>
<td>MD</td>
<td>Republic of Moldova</td>
<td>MG</td>
<td>Madagascar</td>
</tr>
<tr>
<td>MK</td>
<td>The former Yugoslav Republic of Macedonia</td>
<td>ML</td>
<td>Mali</td>
</tr>
<tr>
<td>MN</td>
<td>Mongolia</td>
<td>MR</td>
<td>Mauritania</td>
</tr>
<tr>
<td>MW</td>
<td>Malawi</td>
<td>MX</td>
<td>Mexico</td>
</tr>
<tr>
<td>NE</td>
<td>Niger</td>
<td>NL</td>
<td>Netherlands</td>
</tr>
<tr>
<td>NO</td>
<td>Norway</td>
<td>NZ</td>
<td>New Zealand</td>
</tr>
<tr>
<td>PL</td>
<td>Poland</td>
<td>PT</td>
<td>Portugal</td>
</tr>
<tr>
<td>RO</td>
<td>Romania</td>
<td>RU</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>SD</td>
<td>Sudan</td>
<td>SE</td>
<td>Sweden</td>
</tr>
<tr>
<td>SG</td>
<td>Singapore</td>
<td>SI</td>
<td>Slovenia</td>
</tr>
<tr>
<td>SK</td>
<td>Slovakia</td>
<td>SN</td>
<td>Senegal</td>
</tr>
<tr>
<td>SZ</td>
<td>Swaziland</td>
<td>TD</td>
<td>Chad</td>
</tr>
<tr>
<td>TG</td>
<td>Togo</td>
<td>TJ</td>
<td>Tajikistan</td>
</tr>
<tr>
<td>TM</td>
<td>Turkmenistan</td>
<td>TR</td>
<td>Turkey</td>
</tr>
<tr>
<td>TT</td>
<td>Trinidad and Tobago</td>
<td>UA</td>
<td>Ukraine</td>
</tr>
<tr>
<td>UG</td>
<td>Uganda</td>
<td>US</td>
<td>United States of America</td>
</tr>
<tr>
<td>UZ</td>
<td>Uzbekistan</td>
<td>VN</td>
<td>Viet Nam</td>
</tr>
<tr>
<td>YU</td>
<td>Yugoslavia</td>
<td>ZW</td>
<td>Zimbabwe</td>
</tr>
</tbody>
</table>
INTERCONNECTED TRUNK CLUSTER
ARRANGEMENT

FIELD OF THE INVENTION

The invention relates generally to network systems such as Local Area Network (LAN) systems and techniques and device models to provide high availability interfaces, and systems. And the invention relates more particularly to an interconnection of the high availability interface systems.

BACKGROUND OF THE INVENTION

It has been generally known in the network systems field how to provide redundant capabilities. Redundant packet forwarding devices are especially useful to handle the possibility of failed links and/or failed packet forwarding devices (e.g. switches, routers, bridges). However, the detection of network topology loops and the utilization of redundant paths is problematic, particularly in view of the time involved

SUBSTITUTE SHEET (RULE 26)
in detecting the path and rerouting traffic. Further, protocols which have been used are complicated to configure and manage.

Traditionally the spanning tree protocol has been use to provide both Layer 1 and 2 redundancy, in switch networks, by configuring alternate paths. When a link or switch element fails, a backup link is activated. The spanning tree algorithm defines the active and backup links in the topology. Spanning tree is designed to control the many switches of the topology and has a slow (in the tens of seconds) reconfiguration time.

Spanning tree has a significant disadvantage as to the reconfiguration time. As network systems become more complex and handle additional traffic, the reconfiguration time becomes increasingly problematic. Further, spanning tree disables the redundant path. This negates the possibility of using redundant paths for increases in throughput.

U.S. patent application serial number 09/014,547, filed January 28, 1998 (which is hereby incorporated by reference), discloses trunk clusters, namely a packet based high speed mesh which is formed of a set of loosely coupled switches, a configuration protocol and trunked network interfaces. Each switch, in a trunk cluster, provides a single "shared LAN" by interconnecting two or more links. The attached edge devices run the trunk configuration protocol and view the trunk ports as if they are connected to a shared LAN with multiple other attached devices. The maximum throughput of the trunk cluster increases with each additional switch.

Such a trunk cluster provides significant advantages including layer 1 and layer 2 redundancy, the avoidance of the spanning tree and the ability to provide additional throughput as needed (by adding additional switches to the switch cluster). The provision of such a trunk cluster necessitates further consideration such as further interconnection, interconnecting different trunk clusters or group of trunk clusters.

**SUMMARY AND OBJECTS OF THE INVENTION**

It is an object of the invention to provide an arrangement of trunk clusters and a method for interconnecting trunk clusters wherein the interconnection method has no single point of failure, the bandwidth between trunk clusters is not limited by the throughput of a single switch, and faults are contained within each trunk cluster.
According to the invention, a trunked interconnection structure is provided between trunk clusters. With this method each switch of a trunk cluster has a logical port connected to a trunked port. The trunked port provides a physical connection to each trunk switch of another trunk cluster. It is, that another trunk cluster, each trunk switch has a logical port connected to a trunked port which in turn has physical connections to each switch of the first trunk device.

Trunked interconnect isolates faults to a single trunk cluster and there is no single point of failure and the total throughput is not limited to any single switches capacity. This always provides a single loop free path from one trunk cluster to the other or others. Multiple trunk clusters may be interconnected using point-to-point connections. A high throughput campus interconnect trunk cluster can be used to connect each building data center trunk cluster.

Each trunk cluster cooperates with associated edge devices to provide a scalable logical LAN. The trunk cluster is constructed with a set of loosely coupled switches, a configuration protocol, trunked network interfaces, and optionally a reachability protocol.

Each switch in the trunk cluster provides a single "shared LAN" by interconnecting two or more links. The edge devices attached to the links run a trunk configuration protocol. These attached edge devices view each physical link within a trunked port as if the attached device is connected to a shared logical LAN with multiple other attached devices.

A logical LAN is designed to provide scalability and resilience. The set of devices (end devices, switches etc.) that interconnect to a logical LAN, called edge devices, cooperate using configuration protocols and traffic steering methods required to connect to the logical LAN.

The trunk cluster is comprised of two or more trunk switches. A single logical LAN is provided by the edge devices splitting the traffic (directing traffic flow) across the links in a trunk port. Each trunk switch provides a separate path within the trunk cluster (multiple parallel paths are provided). The two or more separate paths between edge devices allow the logical LAN to increase bandwidth by adding more trunk switches and automatically decrease bandwidth in the event of a link failure and/or in the event of a trunk switch failure.
As each trunk switch only carries part of the traffic between edge devices, each trunk switch does not need to and must not participate in any topology control or discovery protocol. Spanning tree, IGMP (Internet Group Management Protocol), and GARP (Generic Attribute Registration Protocol) packets are flooded. Unicast MAC (Media Access Controller) source addresses are learned and used to intelligently forward/filter unicast packets to minimize flooding within the "shared LAN" and increase throughput. The maximum throughput of the trunk cluster increases with each additional trunk switch.

Each MAC device of an edge device transmits a hello signal to MAC devices of other edge devices. The hello signal includes a trunk or edge device ID identifying the respective edge device of the MAC device transmitting the hello signal. Each MAC device records the edge device ID's of said hello signals received from other edge devices. These recorded edge device ID's are formed into an hello list for each MAC device. The TCMP agent of an edge device forms a trunk list for each other edge device. Each trunk list for a particular other edge device includes MAC addresses of the present edge device which received the hello signals from the respective one of the other edge devices. For example, if edge device A had three MAC devices which received hello signals from edge device B, edge device A would have a trunk list for edge device B which contained those three MAC devices. When edge device A received traffic for edge device B, edge device A would divide the traffic among the three MAC devices in the trunk list for edge device B. This dividing of traffic received by edge device A for edge device B, is according to the standard trunking convention. In this way, trunking can be accomplished through a plurality of switches in a mesh instead of just over a plurality of links where all the links start and stop at the same nodes. Also the trunking of the present invention allows links to be active and increase data rates, where previously those links would have been inactive due a redundancy decision by a spanning tree algorithm.

The various features of novelty which characterize the invention are pointed out with particularity in the claims annexed to and forming a part of this disclosure. For a better understanding of the invention, its operating advantages and specific objects attained by its uses, reference is made to the accompanying drawings and descriptive matter in which a preferred embodiment of the invention is illustrated.
BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings:

Figure 1 is a schematic diagram showing the general scheme of the trunk cluster simple redundant switch set according to the invention;

Figure 2A is a diagram showing traffic flow through a multiple link trunked port;

Figure 2B is a diagram showing traffic flow through a multiple link trunked port of a trunk cluster interconnected with at least one other trunk cluster;

Figure 3 is a state diagram showing the state of the media access controllers connected to the trunk cluster;

Figure 4 is a view similar to Figure 1 showing the logical topology as seen by the edge devices;

Figure 5 is a schematic diagram showing the scheme for interconnecting trunk clusters;

Figure 6 is a similar to figure 5 showing a high speed LAN core with multiple interconnecting trunk clusters;

Figure 7 is functionally the same structure as figure 6 shown in the context of a Campus LAN;

Figure 8 is functionally the same structure as figure 7 shown as a CORE BUILDER 9000 chassis implementation.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to the drawings in particular, the invention comprises a packet based high speed mesh 10 which is referred to herein as a trunk cluster. The trunk cluster 10 includes a plurality of switches 20 or switches A through N. The trunk cluster is constructed of loosely coupled switches 20 based on a configuration protocol, trunked network interfaces, and optionally a reachability protocol.

As shown in Figure 1, each switch 20, in a trunk cluster 10, provides a single “shared LAN” by interconnecting two or more links 22. Edge devices 24 are attached to the trunk cluster via links 22 running a trunk configuration. A plurality of links comprise a trunked port 30.

Figure 2 illustrates the traffic flow 40 through a multiple link trunked port such
as multiple link trunked port 30. Figure 2 shows transmit queue 41, receive queue 42, traffic steering means 44, steering data base 46, media access controllers (MACs) 48, physical layers 49 and connected cable (link 22). The trunked port 30 acts as a single port, but utilizes multiple physical links 22. The traffic steering means 44 is the logic that knows which MAC 48 to use when transmitting a packet. The traffic steering means 44 dequeues transmit packets, examines the packet, consults the steering data base 46 and transfers the packet to the appropriate MAC 48. The steering data base 46 is constructed by information obtained by the reachability protocol or trunk configuration management protocol (TCMP).

Figure 3 shows a state diagram of the MACs 48. From the perspective of trunking, a constituent MAC 48 can be in one of the following states.

*not in use.* A MAC 48 in the not in use state 50 has not been selected for active participation in the trunk. A MAC 48 in this state 50 should neither transmit nor receive LLC (Logical Link Control protocol-defined in IEEE 802.2) frames. This should be the initial state for all MACs 48 of a trunk. A trunk MAC 48 should remain in this state indefinitely, if the underlying network link 22 is down. If TCMP is enabled for a particular trunk, then the MAC 48 may remain in this state 50, even if the underlying network link is up, if TCMP determines that configuration errors preclude its active participation in the trunk.

*selected.* A MAC 48 in the selected state 52 has been selected for active participation in the trunk, but it has not yet made the transition into active use. A MAC 48 in this state 52 should neither transmit nor receive LLC frames. A MAC 48 in this state is eligible to make a transition to the in use state 54 as soon as TCMP confirms that all other MACs 48 connected to the same network have also been selected. If TCMP is not enabled for a trunk, then no MAC 48 for that trunk should ever enter this state.

*in use.* A MAC 48 in the in use state 54 is in active use on the trunk. A MAC 48 in this state may transmit and receive LLC frames. If a trunk is manually configured (the connections are manually assigned), then a MAC 48 will transition to this state as soon as the underlying network link is up. If TCMP is enabled (for control and configuration), on the other hand, then a MAC 48 will make a transition to the in use state as soon as it confirms that all other MACs 48 connected to the
same logical network (the same switch 20) have also been selected.

There is a separate instance of TCMP (a TCMP-agent) for each individual trunk. Initially every MAC 48 starts in the not in use state. Every TCMP "hello time", a TCMP-agent (a software based means 45 writing to the traffic steering means 44) transmits a TCMP "hello message" on every MAC 48. The hello message contains the sender's TCMP node trunk or edge device ID which uniquely identifies the instance of a trunk on a particular node. The hello message also contains the trunking MAC 48 state of the transmitting MAC 48. The hello message is transmitted to a special TCMP multicast address.

The TCMP-agent maintains a current and previous learned node trunk list (steering data base 46) for every MAC 48. Whenever the TCMP-agent receives an TCMP hello message, it adds the node trunk ID to the current list for that MAC 48. If more node trunk IDs are learned than the maximum allowable, then a list overflow indicator is set. The learned node trunk list and list overflow are set to null initially and after TCMP reselection processing.

Every TCMP reselection time, the TCMP-agent performs reselection processing. The goal of reselection is to select the set of MACs 48 which are eligible to go in use. The TCMP-agent accomplishes this by identifying the current learned node trunk list which is common to the greatest number of MACs 48 using the following steps:

1. The TCMP-agent first rejects any list which is null or overflowed.
2. The TCMP-agent also rejects any list which contains a received node trunk ID which is identical to the receiving trunk’s node trunk ID (since this indicates that there are two MACs 48 on the same trunk which connect to the same network).
3. From the remaining set of lists, the TCMP-agent selects the list(s) which is common to the greatest number of MACs 48. This biases selection towards activating the widest pipe (path - the activated links switches) possible. If there is no tie, then a list is selected. Otherwise the process continues.
4. If multiple lists are common to the same number of MACs 48, then the TCMP-agent selects the list(s) which is longest. For equal bandwidth pipes, this biases selection towards connecting the greatest number of nodes. If there is no tie, then a list is selected. Otherwise the process continues.
5. If multiple lists are common to the same number of MACs 48 and they
connect the same number of nodes, then the TCMP-agent selects the list containing
the numerically smallest non-intersecting node trunk ID. For example given two lists
of node trunk IDs, 3-4-6-7 and 3-4-8, the second list wins, because 5 is the lowest
non-intersecting ID. There cannot be a tie resulting from this step.

Once the TCMP-agent has selected a learned node trunk list, it forces MAC
48 state changes based on the selection. (If no list has been selected, then the TCMP-
agent force all MACs 48 into the not in use state). The TCMP-agent forces all MACs
48 whose current learned node trunk list differs from the selected learned node trunk
list into the not in use state. The agent forces all MACs 48 with identical lists and not
already in the in use state into the selected state. It allows MACs 48 with identical
lists that are already in use to remain in use.

Following this reselection process, the TCMP-agent will copy each MAC's
current learned node trunk list to its previous list and then reset each current list.
This allows the reselection process to be performed completely anew each period.
Note that the previous list is used during quick recovery from network link outages.

The final step of MAC 48 activation/deactivation involves further MAC 48 state
transitions based upon feedback received via the MAC 48 state field of subsequent
hello messages.

The TCMP-agent effects a MAC 48 state transition from selected to in use
once that MAC 48 has received a hello message from every previously learned node
trunk indicating a MAC 48 state of either selected or in use. As long as the most
recently received hello message from any of the learned node trunks remains not in
use, then the TCMP-agent does not allow the MAC 48 to go in use. This feedback
loop ensures that all of the MACs 48 on the same network have concurring reselection
processes. Note that a lack of concurrence might be caused by conditions such as a
stuck transmitter or various sorts of configuration errors.

The TCMP-agent also uses the feedback presented by the MAC 48 state field
to effect MAC 48 state transitions from in use back to selected. This occurs when an
in use MAC 48 receives a hello message from any previously learned node trunk
indicating a MAC 48 state of not in use.

The TCMP-agent should ignore the MAC 48 state field in any hello messages
received from any node trunks which were not previously learned. The TCMP-agent
should defer processing of such new node trunks until the next reselection process.

Figure 4 provides the view of the trunked ports of the edge devices (e.g. end device or switches) 24 as seen by the trunk configuration management protocol (TCMP). Each edge device 24 is connected to the other edge devices 24 by multiple paths (link 22 and trunk switch 20). Each trunk switch 20 only provides one “shared LAN” path such that it carries only a part of the traffic between edge devices 24. Each trunk switch 20 does not need to and must not participate in any topology control or discovery protocol. Spanning tree, TCMP, IGMP, and GARP packets are flooded.

Although each switch 20 does not participate in any topology control or discovery protocol Unicast MAC 48 source addresses are learned and used to intelligently forward/filter unicast packets. This is desirable to minimize flooding within the “shared LAN” and increase throughput. The maximum throughput of the trunk cluster increases with each additional switch.

As shown in Figure 1, the edge devices 24 connect to a trunk cluster via trunked links 22. Basic trunking and trunking enhanced with station reachability are both contemplated.

Basic trunking may be provided which requires at least one physical connection to each switch (A through N) 20 in the trunk cluster 10. This referred to as equal reachability. With equal reachability, the destination address is associated with all links 22 of the trunked ports 30. This prevents any single link attachments to the trunk cluster. If a link is lost, leaving no remaining links to that switch, then the trunk configuration protocol will eliminate links connected to that switch. Effectively removing that switch from trunk cluster. Basic trunking uses an “all or none” selection of MACs 48 in a Trunk group. This scheme denies use of a switch 20 in a trunk cluster 10 for all edge devices 24 if a single edge device 24 loses its last link to that switch. This maintains equal reachability for all stations or edge devices 24 over the links 22 in a trunk group.

An enhanced trunking scheme may also be used. The enhanced trunking scheme determines if reachability is equal across all links in the trunk group. If reachability is unequal then the station addresses that have unequal reachability are discovered, packets destined to stations with equal reachability are steered to the
transmit link using basic trunking, and packets destined to stations with unequal reachability are steered using a different method. This different method could range from static link assignment, with one destination address assigned to one physical port, to a partial trunking group, wherein one destination address can be assigned to a subset of the links. In the case of a two link trunk group, all stations with unequal reachability will have a static assignment to the only link they are reachable by. To implement this enhanced trunking, changes must be made to the TCMP and to the traffic steering method, and a station address distribution protocol is required.

Station unicast MAC 48 addresses (for the switches A-N 22) that are reachable via an edge device 22 can be discovered by one of the methods described below.

The GURP application (GARP Unicast Registration Protocol) utilizes the GARP protocol (IEEE 802.1,q,p) (GARP Unicast Registration Protocol) and distributes the reachability of unicast MAC 48 addresses via an edge device. This protocol would only need to run when the trunk configuration management protocol discovers unequal reachability (unequal unempty trunk ID lists) on its trunked links 22. The GURP messages are sent on all the links in a trunk group. The edge devices 24 note which MAC 48 addresses (for the switches A-N 22) are not reachable on all the links 22 in a trunk cluster 10. The MAC 48 addresses that have unequal reachability are then steered by a mechanism different than the basic trunking across all the links in a trunk group. If this trunk group contains two links then the MAC 48 addresses that have unequal reachability are assigned to the only link they are reachable.

Another scheme is TCARP (Trunk Cluster ARP). With this scheme MAC 48 address reachability is discovered when needed. As with the GURP scheme above all MAC 48 addresses (for the switches A-N 22) are known to be equally reachable if the trunk configuration management protocol does not discover unequal unempty trunk ID lists on its trunked links. If the edge device 24 is an endsystem, the reachability of an address is determined when the end system 24 wants to send a packet to that MAC 48 address. If the edge device 24 is a switch, the reachability of an address is determined when a MAC 48 address is learnt and entered in the forwarding database.

The reachability of a MAC 48 address is determined by sending a query on all the links in the trunk group. The edge device that provides connectivity to that MAC

SUBSTITUTE SHEET (RULE 26)
48 address replies on the link the query was received. If the replies are not heard on all the links in the Trunk group the address is determined to have unequal reachability.

The MAC 48 address is assigned to one of the links which is reachable. This assigned scheme can vary from static link assignment to a sub-trunking group.

Figure 5 shows an arrangement according to the invention wherein the trunk clusters are interconnected. The trunk cluster of Figure 5 is as shown in Figure 1. The individual trunk switch elements 20 are connected by links 22 to trunk port 30. In this way, each edge device 24 is connected to each trunk switch 20 of the trunk cluster 10. As is known in the art, the switch 20 may be provided with a plurality of ports. An logical port 31 is connected to a trunked port 30. This is provided for interconnection with other edge devices through the another trunk cluster 110.

Another trunk cluster 110 includes a plurality of trunk switches 120. In a manner similar to trunk cluster 10, trunk cluster 110 is connected via links 122 trunk ports 130. The trunk ports 130 are connected via logical ports 123 to edge devices 124.

A key feature of the interconnect system is the use of a trunked interconnect. Point to point connections or links 125 extend from each trunk port 137 to a trunk port 137 of the other trunk cluster. For example, a trunk switch 120 of a trunk cluster 110 has a logical port 131 connected to a trunked port 137. The trunked port includes links 125 or point to point connections from the associated trunked port 137 to each of the trunked ports associated with the switches 20 of trunk cluster 10. This interconnection method allows multiple trunk clusters to be interconnected using such point-to-point connections 125. This always provides a single loop free path from one trunk cluster to the other. Further, this isolates faults to a single trunk cluster and there is no single point of failure. Further with this interconnection system the total throughput from one trunk cluster to another is not limited by the capacity of any single switch.

The Interconnected Trunk Management Protocol (ITCMP) is a management protocol for the interconnection between trunk clusters (mesh switches). ITCMP is similar to TCMP described above. Every ITCMP hello time, a ITCMP-agent (a software based means 45, writing to the traffic steering means 44) transmits a ITCMP "hello message" on every MAC 48 of the logical trunked ports 137. The hello message
is as described above and includes a senders ITCMP trunk cluster ID, the trunking
MAC state of the transmitting MAC. The unique identifier is a trunk cluster ID
which is unique for the switch set. The procedure continues as noted above with
resulting MAC state changes as discussed above. The ITCMP system is quite similar
to the TCMP system described above except trunk cluster ID's are passed around with
ITCMP.

Trunk switches transform all received TCMP hello messages into TCMP relay
messages by changing the messageType from helloMessage to helloRelay. The
resulting TCMP relay messages are sent to the other Trunk Clusters.

Trunk Switches process the relay message with their TCMP agent to form a list
of MAC states. The MAC agent performs a logical AND of the output from the
TCMP agent and the output from the ITCMP agent and writes the Steering data base
with the result (see figure 2a). If an implementation can not disable only
transimition to a physical link while still allowing packets to be received on that link,
then the ITCMP agent must reflect the state output from the TCMP agent in the
ITCMP MAC state. The ITCMP agent communicates this MAC state information in
the ITCMP hello messages to prevent the other Trunk Cluster from sending packets
to a MAC in the notInUse state.

Equal reachability is preferably used for trunk interconnected trunk clusters.
The destination address is associated with all links 125 of a trunked port 137.

Figure 6 shows the same arrangement as in Figure 1 however in this situation
trunk cluster 110 forms a campus interconnect device for connecting trunk cluster 10
to trunk cluster 310, via trunk cluster 110. Here, the switches 120 of trunk cluster 110
have additional logical ports 131 which are connected to other trunk ports 137. Each
connected trunk port 137 has individual point to point links 125 to the trunk ports 137
of trunk cluster 210. This additional interconnection provides the trunk cluster as a
high throughput campus interconnect trunk cluster that can be used to connect for
example a trunk cluster 10 in one building (building 1) to a trunk cluster 210 in
another building (building 2).

Figure 7 shows the concepts of the interconnect trunk cluster 110 wherein the
switches 120 are not connected to edge devices. Instead, one logical port of each
trunk switch 120 is connected to one of two different trunk ports 137. The point-to-
point links 125 are employed as discussed with reference to Figures 5 and 6. In the arrangement shown in Figure 7, the trunk cluster 10 is connected by links 22 to trunk ports 30 however in this case the edge devices are designated "wiring closet", namely one or more connected devices 24. The trunk cluster 110 is also connected to trunk cluster 210. Trunk clusters 10 and 110 are provided in building 1 while trunk cluster 210 is provided in building 2. Trunk cluster 210 has links 222 to trunk ports 230 connected to wire closets 24.

Figure 8 shows an implementation of the interconnection arrangement discussed above. This is the CORE BUILDER 9000 chassis implementation, using the 3 COM Corp. CORE BUILDER 9000 chassis product. According to this preferred embodiment, two GA9s are provided in a CORE BUILDER 9000 chassis (CB 9000) and form the trunk cluster 10. A GA9 is a 12 port Gigabit Ethernet (GE) switch with 9 GE front panel ports and 3 GE backplane ports is connected to the chassis fabric. The two fabric modules in the chassis CB 9000 are each 24 port Gigabit Ethernet switch modules which form the trunk cluster 110. Each GEN 2 module provides front panel access to the chassis fabric ports and to form the trunk ports 137 and three connection lines to the backplane. The trunk cluster fabric 210 is based on two or more switches. Links 125 connect the GEN 2 devices 135. The trunked ports 137 are formed at the connection of each GEN 2 135 to the backplane and the switching fabric 210 that forms the trunk cluster. The GA9s in building 1 cooperate to form the trunk cluster 10. However, in building 2 the GA9s are connected by redundant links 322 to the wire closet stack. The trunk ports 230 are formed at the backplane based on the three connection lines to the backplane of the GA 9.

While specific embodiments of the invention have been shown and described in detail to illustrate the application of the principles of the invention, it will be understood that the invention may be embodied otherwise without departing from such principles.
WHAT IS CLAIMED IS:

1. A logical LAN interconnection system, comprising:
   a first trunk cluster including a plurality of switches grouped together to form a scaleable logical LAN;
   edge devices comprising one of an end system and a switch, and trunk link means connecting each switch of said plurality of switches to each of said edge devices;
   a second trunk cluster including a plurality of switches grouped together to form a scaleable logical LAN;
   edge devices comprising one of an end system and a switch, and trunk link means connecting each switch of said plurality of switches to each of said edge devices; and
   trunk cluster interconnection means providing point-to-point connections between each of said plurality of switches of said first trunk cluster and each of said plurality of switches of said second trunk cluster.

2. The logical LAN interconnection system according to claim 1, wherein said trunk cluster interconnection means comprises:
   a first trunked port connected to a switch of said first trunk cluster;
   a second trunked port connected to another switch of said first trunk cluster;
   a third trunked port connected to a switch of said second trunk cluster;
   a fourth trunked port connected to another switch of said second trunk switch cluster;
   a first link connecting said first trunked port to said third trunked port;
   a second link connecting said first trunked port to said fourth trunked port;
   a third link connecting said second trunked port to said third trunked port;
   a forth link connecting said second trunked port to said fourth trunked port.

3. The logical LAN interconnection system according to claim 1, wherein said trunk cluster interconnection means comprises:
   a campus interconnect trunk cluster including a plurality of switches grouped together to form a scaleable logical LAN;
a first trunked port connected to a switch of said first trunk cluster;
a second trunked port connected to another switch of said first trunk cluster;
a third trunked port connected to a switch of said second trunk cluster;
a fourth trunked port connected to another switch of said second trunk cluster;
a fifth trunked port connected to a switch of said campus interconnect trunk
cluster;
a sixth trunked port connected to said switch of said campus interconnect trunk
cluster;
a seventh trunked port connected to another switch of said campus interconnect
trunk cluster;
an eighth trunked port connected to said another switch of said campus
interconnect trunk cluster;
a first link connecting said first trunked port to said fifth trunked port;
a second link connecting said first trunked port to said seventh trunked port;
a third link connecting said second trunked port to said fifth trunked port;
a forth link connecting said second trunked port to said seventh trunked port;
a fifth link connecting said third trunked port to said sixth trunked port;
a sixth link connecting said third trunked port to said eighth trunked port;
a seventh link connecting fourth trunked port to said sixth trunked port;
a eighth link connecting said fourth trunked port to said eighth trunked port.

4. The logical LAN according to claim 1, further comprising means implementing a configuration management protocol for control of communication between interconnected trunk clusters.

5. The logical LAN according to claim 4, wherein said trunk link means provides one physical connection to each switch of said plurality of switches, said trunk configuration management protocol eliminating links connected to a switch with a non functioning link or a non functioning switch.

6. The logical LAN according to claim 1, wherein said trunk link means provides one physical connection to each switch of said plurality of switches.
7. A method of trunking a plurality of mesh switches, the method comprising the steps of:

   providing a plurality of mesh switches, each switch of a mesh switch being connected to another each switch of another mesh switch by a separate link and MAC device, said each mesh switch including a TCMP agent;

   each said MAC device transmitting an Hello signal to MAC devices of other said mesh switches, said Hello signal including an mesh ID identifying a respective mesh switch of said MAC device;

   each said MAC device recording mesh ID’s of said Hello signals received from other said mesh switches;

   said TCMP agent of said respective mesh switch forming a trunk list for each other edge device, each said trunk list including all MAC devices of a respective mesh switch which received said Hello signals from one of said other mesh switches;

   each said TCMP agent dividing traffic received for one of said mesh switches among said MAC devices of said trunk list for said one of said other mesh switches, said dividing being according to a trunking convention.
Physical Picture

High speed Mesh - Trunk Cluster

Switch 1

Switch N

Edge Device

Edge Device

Edge Device

Fig. 1
Fig. 5
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6): H04L 12/28, 12/56
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>US 5,088,091 A (SCHROEDER et al.) 11 February 1992, figs. 2-3; col. 1 line 18 to col. 3 line 60; col. 4 line 50 to col. 5 line 2</td>
<td>1, 4-6</td>
</tr>
<tr>
<td>Y</td>
<td>US 5,138,615 A (LAMPORT et al.) 11 August 1992, figs. 2-3; col. 1 line 19 to col. 3 line 60; col. 4 line 50 to col. 5 line 2.</td>
<td>2-3</td>
</tr>
<tr>
<td>Y</td>
<td>US 5,179,558 A (THACKER et al.) 12 January 1993, figs. 1-2; col. 1 line 19 to col. 2 line 19; col. 2 line 52 to col. 3 line 2.</td>
<td>2-3</td>
</tr>
<tr>
<td>A</td>
<td>US 5,473,599 A (LI et al.) 05 December 1995, figs. 1-2a; see abstract; and col. 1 line 11 to col. 3 line 39.</td>
<td>7</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C. See patent family annex.

Date of the actual completion of the international search: 03 JULY 1999

Date of mailing of the international search report: 18 AUG 1999

Name and mailing address of the ISA/US Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
BOB A. PHUNKULH
Telephone No. (703) 308-8251

Form PCT/ISA/210 (second sheet) (July 1992)