PROVISIONAL APPLICATION COVER SHEET

A/PRAV

Attorney Docket No.: ADAPP085A+

First Named Inventor: WILSON, Andrew

Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

☐ Duplicate for fee processing

Sir: This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c).

INVENTOR(S)/APPLICANT(S)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
WILSON	Andrew		San Jose, CA

TITLE OF INVENTION (280 characters max)

SCSI Over Ethernet

CORRESPONDENCE ADDRESS

Albert S. Penilla, Esq.
MARTINE PENILLA & KIM, LLP
710 Lakeway Drive, Suite 170
Sunnyvale, California 94086
(408) 749-6900

ENCLOSED APPLICATION PARTS (check all that apply)

X Specification — Number of Pages 74 (including drawings) — ____ Small Entity Statement

___ Drawing(s) — Number of Sheets _____ — ____ Other (specify) _____

X A check or money order is enclosed to cover the Provisional filing fees. Provisional Filing Fee Amount ($)**150.00**

X The commissioner is hereby authorized to charge any additional fees which may be required or credit any overpayment to Deposit Account No. 50-0805 (Order No. ADAPP085A+).

The inventions made by an agency of the United States Government or under a contract with an agency of the United States Government.

X No _____ Yes, the name of the U.S. Government agency and the contract number are:

Respectfully Submitted,

SIGNATURE _____ **DATE** November 3, 1999

TYPED NAME Albert S. Penilla **REGISTRATION NO.** 39,487

PROVISIONAL APPLICATION FILING ONLY

SCSI over Ethernet *e*SCSI

Inventor

Andrew Wilson

1 Introduction

This document will show how adding an appropriate light weight protocol to Gigabit Ethernet can turn it into a high quality disk subsystem interconnect.

1.1 Scope and Requirements

Because Ethernet is has price and performance points which cover everything from homes and small offices to the largest computer rooms, an Ethernet based storage subsystem could also find applicability in the same wide range of settings. However, there are some differences in the requirements, which might preclude an *e*SCSI designed for one end of the scale from working at the other. For the desktop and small server market, the basic requirement is to connect local storage class devices to computer systems, especially when the devices are housed outside of the main computer case. For the large server system market, the requirements of storage subsystem sharing and interprocess communication over campus area networks are added. The specific requirements are detailed in the next two subsections.

1.1.1 Desktop and small system requirements

1. Low latency, on the order of a couple hundred microseconds or less
2. Maximum distances of a few hundred meters
3. For disk storage, needs to support bandwidths up to those of small RAID boxes
4. For desktop connectivity, needs to support bandwidths of tapes, printers, scanners, etc.
5. CPU utilization on the order of current storage systems
6. Data must be delivered completely and without corruption.
7. Use off-the-shelf Ethernet components
8. Needs to configure automatically, i.e. can function satisfactorily with little or no system administrator intervention
9. Low cost, especially for desktop connectivity
10. The ability to jointly operate with other protocols is highly desirable.

1.1.2 Large System Requirements

1. Low latency, on the order of a couple hundred microseconds or less
2. Maximum distances of a few kilometers

3. Needs to support bandwidth of leading edge RAID boxes for at least the next 5-10 years
4. Needs to support interprocess communication paradigms
5. CPU utilization on the order of current storage systems
6. Data must be delivered completely and without corruption.
7. Able to use off-the-shelf Ethernet switches and IP routers
8. Needs easy management by system administrator
9. Needs to be cost competitive with other storage network technologies
10. The ability to jointly operate with other protocols is highly desirable.

1.1.3 Requirements discussion

As you can see, there are both similarities and differences between the two sets of requirements. Fortunately, Ethernet offers a range of compatible technologies that can cover the requirements of both small and large systems.

Requirement 1 of each list is easily met using today's Ethernet switches and switching routers. Typical values for latency are 3-20 microseconds for commercially available Gigabit Ethernet switches. Even with several switches and a couple hundred meters of cable, the round trip delay will be less than 100 microseconds in the absence of congestion.

Requirement 2 for small systems can be met with just CD/CSMA Ethernet, which can span 500 meters with fiber optics, or 200 meters diameter with copper. With full duplex, multi-mode optical fiber cables can handle several kilometers between switches, so cross country distances are theoretically possible. Thus switched (or router) based networks can handle the requirements of both small and large systems.

The bandwidth requirements imposed by future disk drives (#3) are difficult to predict, since they depend on advances in recording density and mechanical speed of the disk drives, as well as the access patterns of future applications. Figure 1 shows projected access times and sequential access rates for disk drives, assuming historical annual improvement rates continue. For applications which access data sequentially, or in very large random reads, the projections show that a single drive will be able to nearly saturate a gigabit interconnect. On the other hand, transaction processing applications will still be limited by disk access time to about 350 IOPS per drive, or one to five megabytes a second depending on request size. Even at five megabytes a second, over 20 drives could be supported by a single gigabit link. Thus, gigabit links should be adequate for small server installations, while large systems will need a mixture of 1 and 10 gigabit links.

For the types of traffic discussed in the requirement 4 of the small system section, 100 BaseT will often be adequate, and should be supported along with Gigabit. Large systems will have to support interprocess communication traffic, which consists of fairly short, time sensitive packets. Thus, Gigabit speeds are fine, but low latency and overhead protocols are required to support the IPC.

For large systems with external RAID boxes, the new 10 Gigabit Ethernet will be required. Interprocess communication doesn't require large bandwidths, but does require very low latencies (the lower the better) and efficient transport of short packets. This argues for a light weight transport protocol, but also argues for QOS and priority features such as found in IP version 6.

In order to equal the low CPU utilization of current host adapters (requirement #5 on both lists), the transport protocol will have to run onboard the *e*SCSI host adapter, ideally with special purpose hardware. To keep the hardware from becoming too complicated and expensive, a protocol with much less complexity than TCP/IP must be used. Such a protocol will be proposed later in this document.

Advocates of other storage interconnect proposals often cite the requirement for error free and complete data transfer (#6) as a reason not to use Ethernet. It is true that the basic Ethernet protocol is that of an unreliable Datagram, but addition of any of a number of transport layer protocols can produce a reliable data conduit quite suitable for storage applications. Actually, all other storage interconnects rely on similar mechanisms to mask the occasional data corruption that occurs with any physical interconnect medium, so the differences between Gigabit Ethernet (especially full-duplex Gigabit Ethernet) and other storage interconnects are not as large as they might at first appear.

This document will describe a transport protocol which provides reliable, in-order delivery between systems. Using this protocol, SCSI commands and data can be transferred reliably between initiators and targets.

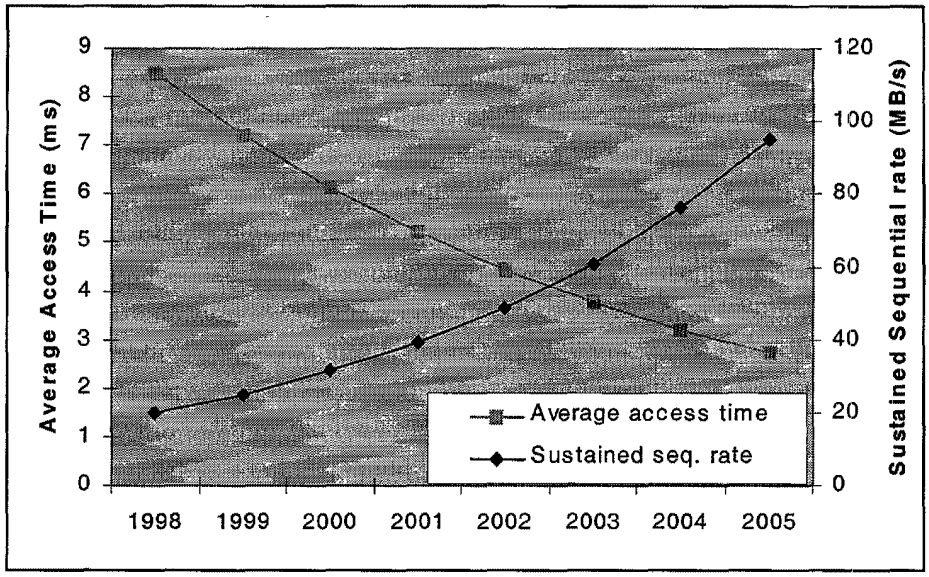

Figure 1: Projected advances in hard disk drives

In a contention and/or level 2 switched environment, requirement seven simply means that we use the basic Ethernet (original or 802.3 version) to encapsulate our eSCSI traffic. If routers are present than IP version 4 or 6 must also be used.

The requirement for small systems of automatic configuration (#8) can be readily satisfied by using switches over plain (no IP) Ethernet with the addition of an automatic discovery protocol (discussed later in this document). Larger systems will have system administrators who will want to manage certain aspects of the storage network, which can be done through use of IP and existing IP based tools. We may need proxy interfaces to *e*SCSI's native protocols, but this can be easily done.

For desktop systems, their low cost requirement (#9) can be easily satisfied with 100BaseT and 1000BaseT over cat5 UDP wiring. Larger systems will want to use fiber optics for longer distance Gigabit runs, and for ten Gigabit links. The use of of-the-shelf Ethernet switches and wiring will produce substantial cost savings due to the large volumes associated with the use of such equipment for networks.

Basic shared media Ethernet (either hub or cable) is completely oblivious to the data it is transporting, and propagates the data between stations (e.g. NICs) using addressing information contained in its own frame header. Thus, any number of protocols, including any we might invent, can be used to package the user's data. Routers, on the other hand, usually need to understand portions of those protocols in order to determine the appropriate path for the packets. Thus, routers only work with a few standard protocols, such as IPX and IP.

Between shared media and routers lie switches, which provide much of the bandwidth multiplication of routers but can use the addressing information contained in the Ethernet frame header to pass packets on to their correct destinations. Thus, Ethernet switches can work with private protocols just as well as plain Ethernet and hence meet requirement number ten.

When mixing several protocols on the same wire however, requirement ten is a little more complicated to achieve. The original Ethernet specification dedicated two bytes in the header to a type field, allowing 65,000 different protocols to co-exist on a single Ethernet. At this time there is still plenty of room for additional protocols. However, IEEE 802.3 changed that field to a length field, requiring other means to distinguish between protocols. Their solution is the addition of eight more bytes of information (3 bytes of 802.2 LLC plus 5 bytes of SNAP) to specify which protocol is encapsulated within a given frame. If we want to be fully compatible with 802.3, we will have to use these eight bytes as well.

In summary, as long as the scope of our *e*SCSI proposal is limited to desktop, computer room or campus networks, the requirements can be easily met with a combination of 100baseT, Gigabit and 10 Gigabit Ethernet and an appropriate transport protocol. We will now proceed to develop such a protocol.

1.2 Topologies

Implementations of *e*SCSI can be developed for both the desktop connectivity market and server storage market. The desktop products will use 100BaseT and some Gigabit to connect to peripherals and expansion drives through point-to-point links or switches. The Server products will use switches and routers interconnected with one and ten gigabit Ethernet to connect multiple hosts to disk and tape farms, and provide host to host communication.

1.2.1 Desktop Connectivity

The desktop topologies start with a point to point link connecting a *e*SCSI PIC to a *e*SCSI target, as depicted by the solid lines in Figure 2. Expansion to more peripherals can follow either of two paths. First of all, any *e*SCSI port can use Ethernet Hubs (or switches) to allow connection to additional targets. Thus, even a single port *e*SCSI PIC can connect to a nearly unlimited number of targets, provided that bandwidth limits are not exceeded. Secondly, Adaptec can produce a multiport *e*SCSI PIC allowing several (e.g. 4) targets to be connected without an external hub. The multiport *e*SCSI PIC could simply have a small hub built into the card, or it could actually have independent ports, providing additional bandwidth. If the independent port approach is used, then port aggregation can be used to provide higher bandwidth to high performance *e*SCSI targets. Finally, Figure 2 also illustrates how multiple hosts could share peripherals when a hub is employed.

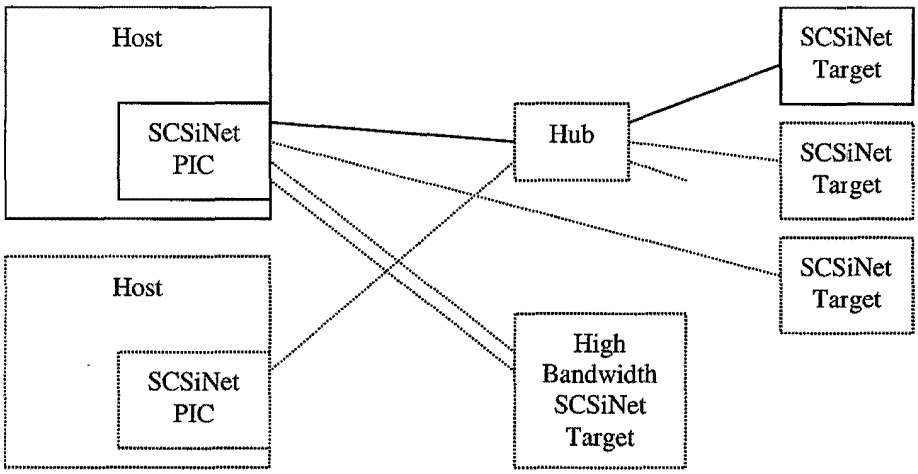

Figure 2: Possible desktop topologies for *e*SCSI

One problem with point-to-point links that do not involve a Hub or a switch is that an Ethernet Crossover cable is required. This is because traditional Ethernet NICS have their receive and transmit wires on the opposite ends of the connector from those of the Hubs. Since networks seldom consist of just two NICS there usually is a hub or switch in between, allowing straight through cables to work just fine.

But, a *e*SCSI PIC with multiple ports would normally be connected to just one target per port. Since target ports would have to be wired the same as PIC ports (to allow connection to hubs and switches), a crossover cable is required. The other option is to make *e*SCSI PIC ports configurable (hopefully automatically) as "hub" type ports or "NIC" type ports. How feasible this is has not yet been determined.

1.2.2 Combined with Network

While it would be simplest to have *e*SCSI separate from communications networks, there is a significant appeal to having a single, combined function Ethernet connector on the desktop computer. Even if several *e*SCSI connectors are provided, being able to use them for either external network or *e*SCSI, without regard to which connector is used for which purpose is a strong selling point.

For example, a single computer with PIC/NIC with two or more independent ports might dedicate one to an external network connection, and the other(s) to *e*SCSI devices. If several computers were involved, and they shared both a network connection and several desktop peripherals, then connecting them through a single small hub would be

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.