



**SIS - PILOT LONG-DISTANCE HIGH SPEED AND SECURE
DATA TRANSFER BETWEEN REPOSITORIES**

Final report

Chief Investigator: Dr. Asad Khan

Lead Investigator: Abdul Malik Khan

Version 0.1, MAY 2007

Monash University

Prepared By:

Abdul Malik Khan abdul.malik.khan@gmail.com

SUMMARY

TCP Transport layer protocols provide for end to end communication between two or more hosts. The Standard TCP (TCP Reno) is not suitable for high bandwidth, large RTT networks because of its low performance in high throughput. Standard TCP is a reliable transport protocol that is well tuned to perform in traditional networks. TCP performance depends upon the product of the transfer rate and the round-trip delay. TCP survived the days of low bandwidth, high latency, and high error rates. But for several reasons it is not able to cope efficiently with the evolving new environment. Several experiments and analysis have shown that this protocol is not suitable for huge data transfer in high bandwidth environment such as Gigabit and high-speed long distance networks with large round trip time. This is because of its slow start and conservative congestion control mechanism. Our focus on introduction to new transport protocols FAST TCP is to improve performance when compared to standard TCP in low bandwidth or small RTT networks and much better than standard TCP in high bandwidth, large RTT networks. Grid Applications need immediate deployment of modified transport protocols to utilize the huge bandwidth. In this Workpage-SI8 Dart project, we have reviewed, experimented and compared different emerging alternatives that solve this problem, in the particular context of very high speed networks.

The potential of improved access to experimental data to enrich scholarly communication between research and learning is the driving force. But this communication is viable only when the data transfers with efficient use of high bandwidth. This can only be achieved with fine tuning the TCP stack and implementation of new algorithm such as FAST TCP. While the design of FAST TCP for high-speed networks has received a considerable amount of interest, less attention has been paid to reap the benefits of these protocols. For example, network measurements have showed complex behaviors and characteristics of high speed network link traffic. Unfortunately, existing evaluation work did not capture these behaviors in their testing environments. Since congestion control algorithms are very sensitive to environmental variables such as background traffic and propagation delays, thorough performance of new TCP stack for high-speed networks require creating realistic network environments where these protocols are sure to be used.

The objective/goals of the current and future production stage will be to provide a real scenario for research community. Increased communication and computing requirements of the research organizations mean it is essential that regular monitoring of the network operations be conducted. A high bandwidth network links between repositories will be an expensive resource that requires optimization and fine tuning to utilize the available bandwidth efficiently. Therefore careful optimization, monitoring, management, and maintenance practices are key objectives of the next production stage. Further, blueprints of network and documentation will ease the tasks of performance management and capacity planning. Furthermore this production stage will be a showcase to other research organizations and educational institutions to choose this path for information sharing.

The next production stage will ensure an adequate level of understanding of network operations for data security, data integrity, high-speed data transfer between the research organizations. The SI-8 work package development will satisfy the research community for sharing precious data which will have all the infrastructure requirements and user data integrity, security, access rights and ownership issues resolved. Such an understanding will provide intelligent answers to questions of cost, performance, capacity planning and direction of growth that frequently arise throughout the lifecycle of a R&D.

TABLE OF CONTENTS

Section	Title	Page
SUMMARY	II
ACKNOWLEDGEMENTS	IX
SECTION 1	INTRODUCTION.....	1
1.1	THE OBJECTIVES OF THIS WORK PACKAGE	2
1.2	PROJECT MILESTONES.....	2
1.3	PROJECT DELIVERABLES.....	2
SECTION 2	NETWORK INFRASTRUCTURE –AARNET AND GRANGENET	4
2.1	AARNET.....	4
2.1.1	<i>AARNet background</i>	4
2.1.2	<i>Current AARNet network</i>	4
2.1.3	<i>Comparison of Links for AARNet and GrangeNet</i>	6
2.1.4	<i>AARNet Current Services</i>	6
2.1.5	<i>Collaborative areas of research on AARNet</i>	6
2.1.6	<i>AARNet: Quality of Service</i>	7
2.1.7	<i>AARNet: Network Operations</i>	7
2.1.8	<i>AARNet3</i>	7
2.2	GRANGENET	8
2.2.1	<i>GrangeNet Background</i>	8
2.2.2	<i>GrangeNet current network</i>	8
2.2.3	<i>GrangeNet current network hardware</i>	9
2.2.4	<i>New architecture of GrangeNet II</i>	9
2.2.5	<i>Services Supported on GrangeNet</i>	10
2.2.6	<i>Cost</i>	11
SECTION 3	SURVEY OF PROTOCOLS AND MECHANISMS FOR LONG DISTANCE AND HIGH SPEED NETWORKS	12
3.1	METHODOLOGY AND COMPARISON CRITERIONS	12
3.1.1	<i>Functions and features of a transport protocol</i>	12

3.1.2	<i>Comparison criteria</i>	13
SECTION 4	PERFORMANCE WITH FAST TCP	15
4.1	FAST (FAST AQM (ACTIVE QUEUE MANAGEMENT) SCALABLE TCP)	15
4.3	FTP BETWEEN SERVER AND CLIENT ON GRANGENET WITH FAST ETHERNET REPOSITORIES	17
4.3	FTP BETWEEN SERVER AND CLIENT ON WAN LINK	18
SECTION 5	BENEFITS OF FAST TCP	19
5.1	THE BENEFITS OF FAST TCP	20
5.2	EASE OF INSTALLATION / IMPLEMENTATION OF FAST TCP STACK ON LINUX KERNEL	20
5.3	THE EASE OF FAST TCP USE	21
5.4	OPERATING SYSTEM REQUIREMENTS	21
5.5	FAST TCP INSTALLATION FOR LINUX KERNEL	21
5.6	STABILITY OF FAST TCP	21
5.7	NEED TO IMPLEMENT FAST TCP	21
5.8	IMPLEMENTATION OF HS-TCP, SCALABLE TCP AND FAST TCP WITH GRIDFTP	22
5.9	IMPLEMENTATION OF FAST TCP FOR FULL PRODUCTION ENVIRONMENT	24
SECTION 6	REQUIREMENT ANALYSIS OF HIGH AVAILABILITY DATA CENTER	25
6.1	HIGH AVAILABILITY DATA CENTER	25
SECTION 7	DATA ENCRYPTORS DECRYPTORS	30
7.1	GSI AND HARDWARE LAYER 2 – LAYER 3 ENCRYPTORS / DECRYPTORS	30
7.1.1	<i>GSI -based</i>	30
7.1.2	<i>GSI –based Test bed</i>	30
7.1.3	<i>Hardware Encryption</i>	31
7.1.4	<i>Data Encryption Performance: Layer 2 vs. Layer 3 Encryption in High Speed Networks</i>	32
7.1.5	<i>Data Encryption throughput Performance test: Layer 2 vs. Layer 3 Encryption in High Speed Networks</i>	32

7.1.6	<i>Data Encryption latency Performance test: Layer 2 vs. Layer 3 Encryption in High Speed Networks</i>	33
7.1.7	<i>Data Encryption frame loss Performance test: Layer 2 vs. Layer 3 Encryption in High Speed Networks</i>	34
7.1.8	<i>Data Encryption and Performance test conclusion: Layer 2 and Layer 3 Encryption in High Speed Networks</i>	35
7.1.9	<i>Software implementation of IPSec in fast Ethernet</i>	36
SECTION 8	DATA INTEGRITY & ENCRYPTION	39
8.1	DATA INTEGRITY	39
8.2	DATA ENCRYPTION	39
8.3	ACCESS CONTROL AND ENCRYPTION	40
SECTION 9	NETWORK MODELS AND PERFORMANCE EVALUATION	42
SECTION 10	RECOMMENDATIONS	44
	<i>Recommendation-1</i>	<i>44</i>
	<i>Recommendation-2</i>	<i>44</i>
	<i>Recommendation-3</i>	<i>44</i>
	<i>Recommendation-4</i>	<i>45</i>
	<i>Recommendation-5</i>	<i>45</i>
	<i>Recommendation-6</i>	<i>45</i>
SECTION 11	TERMS OF REFERENCE	47
11.1	GLOSSARY	47
11.2	REFERENCES	48
SECTION 12	REPORT SIGNOFF	53
APPENDIX	54
SECTION 13	AARNET & GRANGENET COMPARISON	54
13.1	OVERVIEW	54
13.2	COMPETITORS	64
13.3	STRENGTHS	64
13.4	LIMITATIONS	64
13.5	INSIGHT	64

13.6	REGIONAL HEADQUARTERS	65
SECTION 14	FAST TCP INSTALLATION	65
14.1	FAST TCP FOR LINUX 2.6.7 KERNEL.....	65
14.1.1	<i>Tuning FAST Kernel</i>	67
14.2	FAST TCP FOR LINUX 2.6.15 KERNEL.....	67
14.2.1	<i>Tuning FAST Kernel</i>	69
14.3	TUNING FAST TCP FOR LINUX KERNEL –SAMPLE CONFIGURATION.....	69

LIST OF FIGURES

Title	Page
FIGURE 1: AARNET NATIONAL NETWORK IN MID 2001	4
FIGURE 2: CURRENT AARNET NATIONAL NETWORK LINKS	5
FIGURE 3: CURRENT AARNET AND GRANGENET NATIONAL NETWORK LINKS	5
FIGURE 4: NEXT GENERATION NETWORK/GRID STRUCTURES	6
FIGURE 5: GRANGENET PHYSICAL ARCHITECTURE.....	8
FIGURE 6: GRANGENET II TOPOLOGY (L2/L3)	9
FIGURE 7: NEW PHYSICAL & LOGICAL ARCHITECTURE OF GRANGENET	9
FIGURE 8: DATA TRANSFER TEST BETWEEN SERVER AND CLIENT ON FAST ETHERNET ON LOCAL SUBNETS AND GRANGE NET BETWEEN THE REPOSITORIES	18
FIGURE 9: DATA TRANSFER TEST BETWEEN SERVER AND CLIENT ON FAST ETHERNET ON LOCAL SUBNETS AND WAN LINK BETWEEN THE REPOSITORIES.....	19
FIGURE 10: DATA ENCRYPTION THROUGHPUT PERFORMANCE TEST: LAYER 2 VS. LAYER 3 ENCRYPTION IN HIGH SPEED NETWORKS	33
FIGURE 11: DATA ENCRYPTION THROUGHPUT PERFORMANCE TEST: LAYER 2 VS. LAYER 3 ENCRYPTION IN HIGH SPEED NETWORKS	34
FIGURE 12: DATA ENCRYPTION FRAME LOSS PERFORMANCE TEST: LAYER 2 ENCRYPTION IN HIGH SPEED NETWORKS	35
FIGURE 13: DATA ENCRYPTION FRAME LOSS PERFORMANCE TEST: LAYER 3 ENCRYPTION IN HIGH SPEED NETWORKS	35
FIGURE 14: NETWORK PERFORMANCE MODEL OF HIGH SPEED NETWORKS	43

LIST OF TABLES

Title	Page
TABLE 1: TABLE OF DATA TRANSFER TEST BETWEEN SERVER AND CLIENT ON FAST ETHERNET ON LOCAL SUBNETS AND GRANGE NET BETWEEN THE REPOSITORIES.....	18
TABLE 2: TABLE OF DATA TRANSFER TEST BETWEEN SERVER TO CLIENT ON WAN LINK (PENINSULA & CLAYTON).....	19
TABLE 3: TABLE OF PROTOCOLS COMPARED	23
TABLE 4: TABLE OF DATA INTEGRITY & ENCRYPTION HANDLING PROCEDURES FOR SI-8.....	41
TABLE 5: INFRASTRUCTURE	54
TABLE 6: NETWORK ARCHITECTURE.....	57

TABLE 7: TRANSPORT AND NETWORKING	58
TABLE 8: INTERNET PROTOCOL	58
TABLE 9: INTERNET PROTOCOL SERVICE	59
TABLE 10: GRID SERVICES	61
TABLE 11: MANAGED SERVICES	62
TABLE 12: CUSTOMER NETWORK MANAGEMENT / SUPPORT	62
TABLE 13: SERVICE-LEVEL AGREEMENTS	63
TABLE 14: PRICING	64

ACKNOWLEDGEMENTS

I would like to begin with expressing my gratefulness to my chief investigator, Dr. Asad khan, who has accepted me as a researcher and provided me with his invaluable guidance and support during every stage of this research work.

Also I would like to thank DART infrastructure team for providing useful feedback and insight to this work package.

Furthermore, I am thankful to DEST Australia for giving me this opportunity to conduct research.

SECTION 1 INTRODUCTION

This document provides the details of the SI 8 work package “Pilot long distance high speed and secure data transfer between repositories”. The aim of this work package is to provide secure, optimized high speed data transfers mechanisms between the repositories using the research networks such as GrangeNET or AARNet.

The DART project is benefited by this work package on the network infrastructure side with secure, high speed and optimized data transfers using the newer modified transport protocol. Every data centre node would be using the new transport protocol layer which would contribute to enhanced high speed data transfers. This would benefit almost all DART work packages.

TCP Transport layer protocols provide for end to end communication between two or more hosts. The Standard TCP (TCP Reno) is not suitable for high bandwidth, large RTT networks because of its low performance in high throughput. Standard TCP is a reliable transport protocol that is well tuned to perform in traditional networks. TCP performance depends upon the product of the transfer rate and the round-trip delay. TCP survived the days of low bandwidth, high latency, and high error rates. But for several reasons it is not able to cope efficiently with the evolving new environment. Several experiments and analysis have shown that this protocol is not suitable for huge data transfer in high bandwidth environment such as GrangeNet / AARNet networks with large round trip time. This is because of its slow start and conservative congestion control mechanism. In this DART pilot project, we have addressed the specific problem of transport of huge data transfers in grid environment (more than 100's of Gigabytes) over high latency, high bandwidth, and low loss paths. Our focus is on introduction of a totally new transport protocol FAST TCP. The performance of the new proposed protocols is comparable to TCP in low bandwidth or small RTT networks and much better than standard TCP in high bandwidth, large RTT networks. Grid Applications need immediate deployment of modified transport protocols to utilize the huge bandwidth. In this Workpage-SI8 Dart project, we have reviewed, experimented and compared different emerging alternatives that solve this problem, in the particular context of very high speed networks. We believe that these innovations are needed for a production environment which is phasing into experimental networks and requires replacing the standard TCP with an advance TCP algorithm for high performance network. The pilot stage of work package SI8 has tested the proof of concepts for high speed data transfer which requires modification to TCP stack using advance FAST TCP algorithms. The hardware and software requirement analysis of high performance machines to avoid bottlenecks at hardware and software level has been reported. Data transfer test results between the test repositories has shown improvement in bandwidth utilization and throughput with the use of FAST TCP. Further improvement is expected with the use of FAST TCP stack with Grid FTP. In very high speed context, network performances or loss can also occur due to problems on the Data repositories hardware and software. The SI-8 work package also considers these issues in the work package.

1.1 THE OBJECTIVES OF THIS WORK PACKAGE

- Survey and identification of high performance research networks AARNet / GrangeNet.
- Identify new protocols and algorithms for data transfer performance.
- Determination of a high-quality path between the collaborative sites.
- Identify High-performance data center's (Collaborative sites)
- All aspects of Data Integrity, security and ownership
- Data sets for transfer tests and other emerging data patterns
- Compatible or standard interface for data handling
- Network performance study.

1.2 PROJECT MILESTONES

- Analysis of high performance data networks.
- Requirements analysis of hardware/software for high performance data networks
- Network performance and evaluation.
- Conduct data transfer of synchronous real-time data between sites.
- Validation of the network

1.3 PROJECT DELIVERABLES

- FAST TCP software and kernel fine-tuning code
- Simulation Models
- Simulation results

The break down of the report is as follows; Section-1 describes the work package aims, objectives, milestones and deliverables. Section-2 describes Analysis and comparison of high performance data networks AARNet and Grange Net for long distance data transfer and determination of high-quality path between the collaborative partners. Section 3 provides the survey of protocols and mechanism for long distance high speed networks. Sections 4 and 5 describe FAST TCP, its implementation; maximize the utilization of available bandwidth, throughput and latency tests. And conduct data transfer tests. Section 6 covers requirements analysis of hardware/software for high performance data networks. Section-7 describes layer-2, layer-3 data encryptors and decryptors including a comparison of GSI based encryption. Security has never been tested before for large data transfers and this is seen as a tailback. Usually data security is compromised at this stage for link performance. However under DART project this is

one of the key issues and a detailed study of both Symmetric and Asymmetric encryption algorithms and key distribution infrastructure and management will be look in other dart work package.

Section-8 focuses on data integrity and encryption of offline or storage data. Section-9 covers network performance and evaluation and Validation of the network. Network models and performance evaluation: - Network performance models will be built. This model will be validated in the proposed archer project. This is essential for performance evaluation. Baseline study of link performance will include capture of traffic and network topology of full production stage network using network node managers to validate the performance evaluation model and test what-if scenarios for varying data patterns, and bottle-necks.

Finally the next sections are appendix which covers a comparison of AARNet and GrangeNet in section-13 and FAST TCP installation in section-14.

SECTION 2 NETWORK INFRASTRUCTURE –AARNET AND GRANGENET

2.1 AARNET

2.1.1 AARNet background

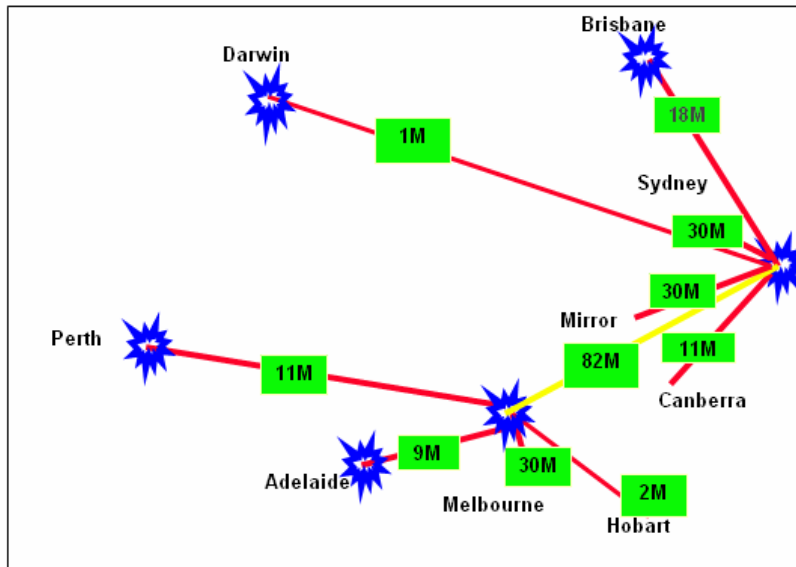


Figure 1: AARNet national network in mid 2001

AARNet was a Joint initiative of AVCC and CSIRO in 1989. AARNet initially served the universities, CSIRO and expanded to include affiliates in 1992 and extended further in 1994 to cater for ISP's. Until mid 1995, AARNet provided all Australian access to the global Internet. AARNet's role in the development of the Internet in Australia began in 1995 when AVCC sold AARNet's commercial customer base to Telstra. Later AARNet refocused its attention on providing services to universities and other research institutions, and then acquired the national ATM backbone from Optus to provide ATM and IP services. AARNet is now a separate legal entity from the AVCC and a non profit company limited by shares. Shareholders are the universities and CSIRO. [1]

2.1.2 Current AARNet network

The current generation of the AARNet network, AARNet3 provides high speed access across the country based on STM-64c (10Gbps) circuits. AARNet provides dual links from Brisbane to Perth all along through Sydney, Canberra, Melbourne, and Adelaide to Perth.

AARNet also provides dual Points of Presence in each of the Australian cities along its path. The dual 10Gbps links connects Brisbane, Sydney, Canberra and Melbourne; While only one 10Gbps & one 622 Mbps link support the Melbourne, Adelaide and Perth path.

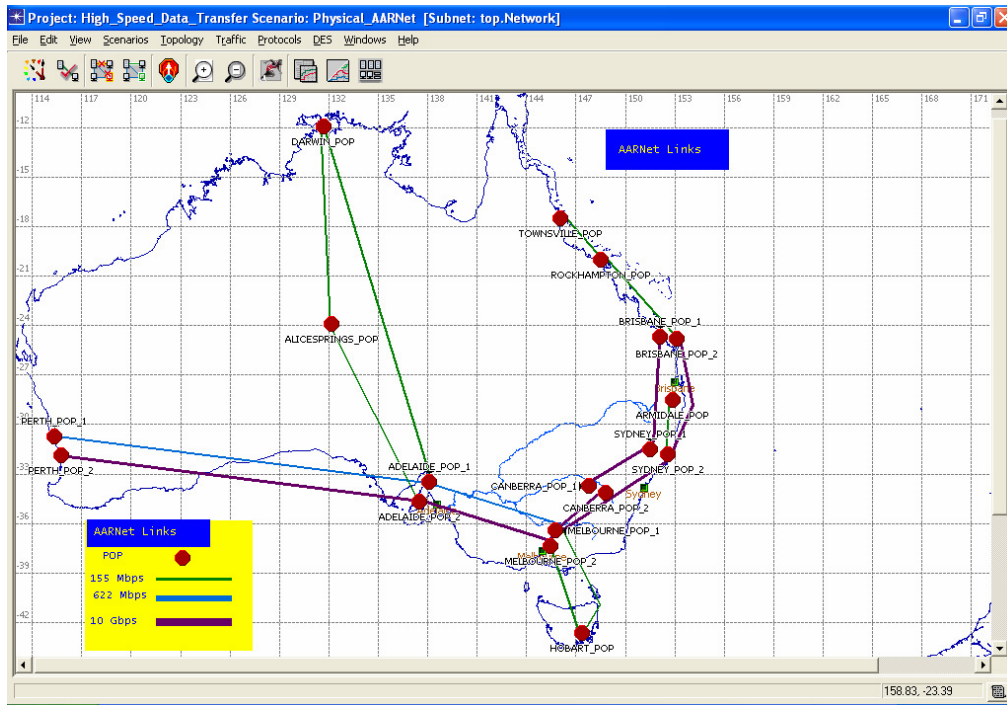


Figure 2: Current AARNet national network links

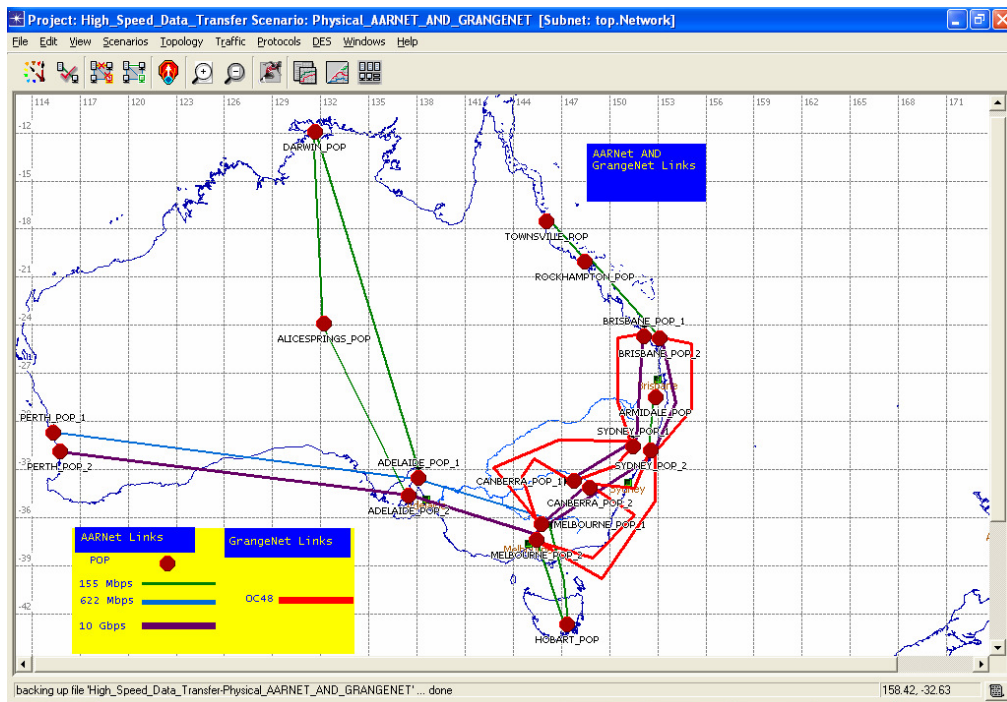


Figure 3: Current AARNet and GrangeNet national network links

AARNet has STM-64c (10Gbps) backbone deployed using DWDM(Dense Wavelength Division Multiplexing system) from Brisbane to Adelaide via Melbourne, Canberra and Sydney. This link provides multiple GigE to regional areas. AARNet has three POP's in Sydney and Perth, Dual POP's in Melbourne, Brisbane, Adelaide and Canberra. While a single POP in other regional cities. AARNet delivers high-capacity; cost-competitive Internet based network services to its members in the tertiary education

and research sector and certain non-member organizations. The AARNet3 network provides an incubator for development of advanced network infrastructure and applications with access to the global Research and Education. [1]

2.1.3 Comparison of Links for AARNet and GrangeNet

City-to-	City	GrangeNet provides	AARNet provides
Brisbane	Sydney	Dual OC48 links	Dual (STM-64c)10 Gbps link
Sydney	Melbourne	Dual OC48 links	(STM-64c)10 Gbps
Sydney	Canberra	Dual OC48 links	(STM-64c)10 Gbps
Canberra	Melbourne	Dual OC48 links	(STM-64c)10 Gbps

2.1.4 AARNet Current Services

AARNet supports basic transport of IP packets with support for large MTU size (Which helps for high data transfer), Supports Unicast, Multicast, IPv4, IPv6, Netflow used for IP accounting and flow analysis (network attacks, protocol usage, etc.), low cost connection to AARNet member users, supports QoS (diffserv), VoIP, Video over IP. Supports large MTUs in Linux clusters to maximize throughput. Implementation of MPLS Fast Failover for link protection. Uses MPLS-TE to minimize latency for Voice/Video over IP (real time applications) and Grid Services (distributed computation, visualization). Support for Grid services due to rapid growth in deployment of Access Grids nodes and support for next Generation Network/GRID Structures. [1]

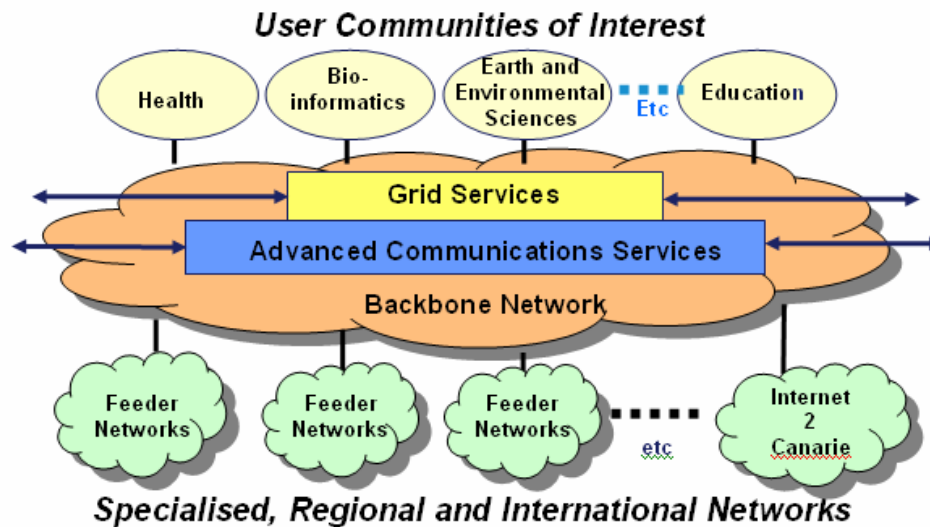


Figure 4: Next Generation Network/GRID Structures

2.1.5 Collaborative areas of research on AARNet

- Astronomy (Low Frequency and Square Kilometer Array) - very long baseline interferometry.
- High Energy Physics.

- Solid Earth and Environmental Sciences – Earth simulators and earthquake prediction
- Bioinformatics.
- Climate and oceanographic studies.
- Remote health applications and telemedicine.
- Film and media.

2.1.6 AARNet: Quality of Service

AARNet3 supports quality of service for all applications to obtain the network service it requires for successful operation. AARNet supports a wide range of quality of service options. QoS provides cost savings, DoS attacks don't affect links with QoS, prevents scavenger service to hog the bandwidth. There are a few quality of service software architectures. AARNet, like most networks, using the IETF's Differentiated Services architecture. QoS will be a key issue for providing high data throughput for data transfer between the repositories, preventing high link utilization. Precedence and type of service bits, differentiated services code point and traffic classes provision will improve our high speed data transfer. [1]

2.1.7 AARNet: Network Operations

The AARNet network is extensively monitored from each Point of Presence (POP) at each Regional Network Organization. As Net flow is deployed at each POP, traffic is measured down to each IP flow. Performance of the Network links and services are tested for availability at regular intervals.

2.1.8 AARNet3

The next generation of the AARNet network, AARNet3, has been deployed with a 10Gbps data link between Melbourne, Canberra, Sydney and Brisbane. This network will provide high speed access across the country based on STM-64c (10Gbps) circuits for serving the needs of the research and education community in Australia. Network operations at POP's are extensive with good QoS and network monitoring. AARNet delivers high-capacity; cost-competitive Internet based network services to its members in the tertiary education and research sector. This was investigated in details and a comparison of each network attribute are presented in Appendix. [1]

2.2 GRANGENET

2.2.1 *GrangeNet Background*

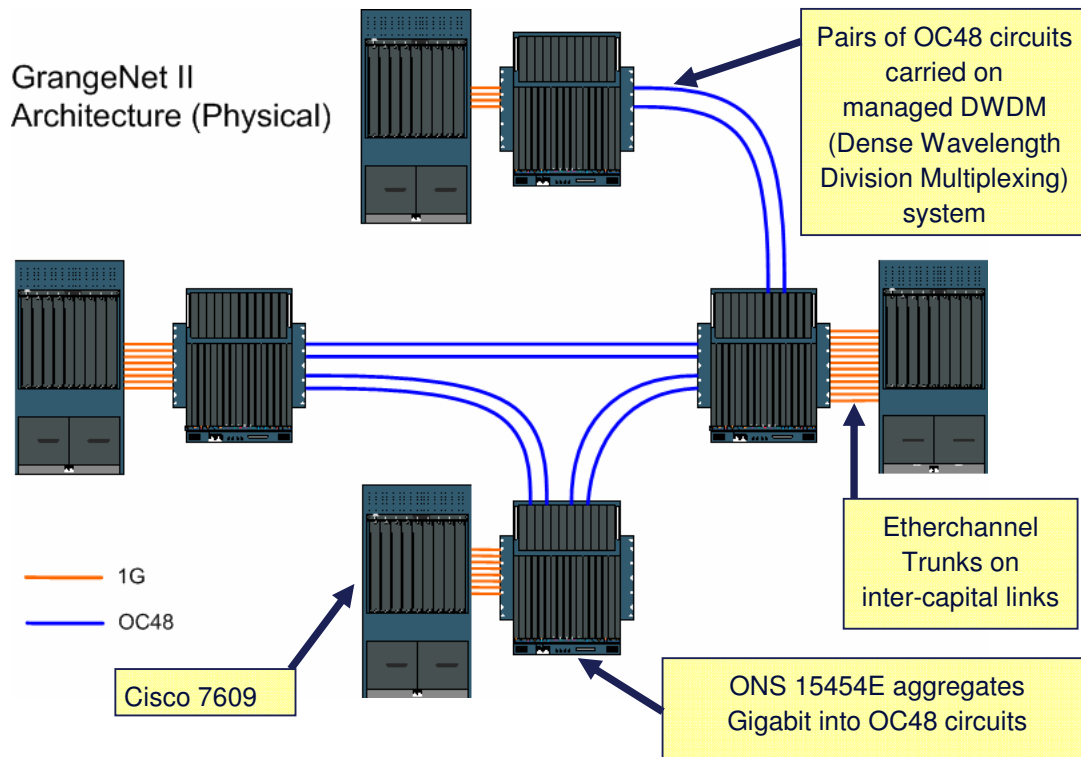


Figure 5: GrangeNet Physical Architecture

GrangeNet is a non-entity joint venture partnership of AARNET, APAC, DSTC, Cisco Systems and PowerTel. Under the initiatives of Advance Networks Program of Building IT strengths (BITS) GrangeNet was established in mid 2002. GrangeNet now supports many communities in the field of ICT, Arts, Sciences, Humanities, and Education & Engineering. GrangeNet is an experimental gigabit network supporting IPv4, Multicast and IPv6. The network is designed to provide high bandwidth with low latency and low jitter data communications. Until 2005 the hardware that supported the network was limited in its capabilities, IPv6 was processed in software and it had limited gigabit uplink ports. [2]

2.2.2 *GrangeNet current network*

The network when first commissioned had a life time until 2004 and was further extended and funded till 2006. The hardware and software were upgraded to support layer 1 –Light paths, Layer-2 VLANs and Layer-3 Routing with support for Unicast and Multicast on both IPv4 and IPv6. The original GrangeNet architecture consisted of routed point-to-point links. With advances in both router and switch architecture the upgraded network architecture is now possible to use the existing 2.5G backbone links and provide layer 1, layer 2 and layer 3 services. The routing topology (L2/L3) now supports 4 Gigabit/sec port channel trunks. Cisco 7609 handles all layer 3 (IP routing), layer 2 (VLAN) switching and Cisco ONS 15454E / ONS 16454E handles all layer 1 (TDM) switching. [2]

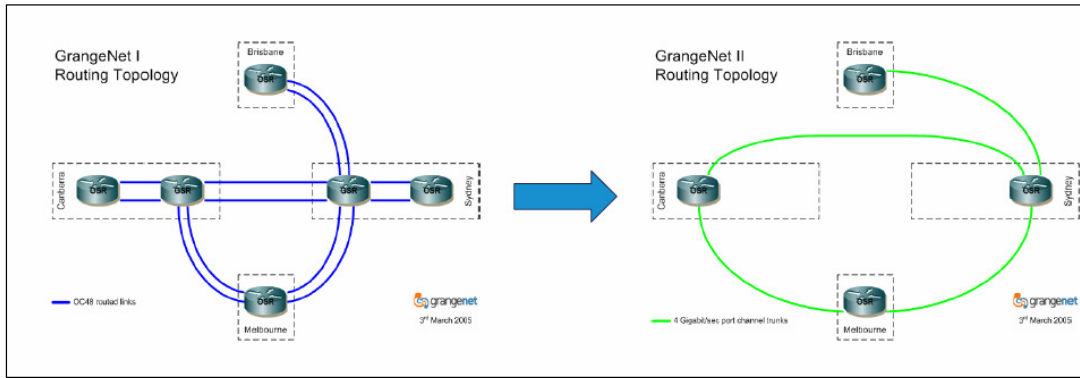


Figure 6: GrangeNet II topology (L2/L3)

2.2.3 GrangeNet current network hardware

The original Routers Cisco OSR 7609 and 12410 were replaced with Cisco 7609 which has given better performance, less jitter and reduced latency and support TDM, OC48 (2488.32 Mbps), OC192 (9953.28 Mbps), and expansion to 10 Gigabits uplinks. The Cisco 7609 has 720 Gbps switching capacity support for 1GB of DRAM. Routing support for IPv4 at 400 Mpps and IPv6 at 200 Mpps.[2]

2.2.4 New architecture of GrangeNet II

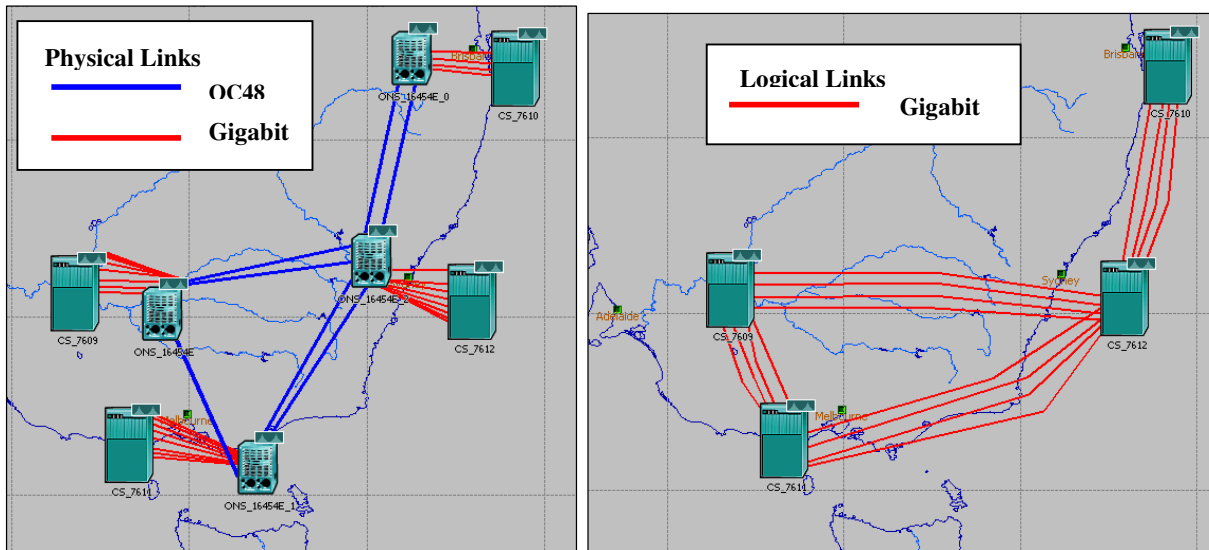


Figure 7: New Physical & Logical architecture of GrangeNet

The hardware for the new architecture is Cisco 7609 with several modules to support Layer 2 and Layer 3 support and ONS 16454E, OC48 backbone and Gigabit Interconnect for Layer-1 support. The new architecture has a Cisco range of optical transport equipment that allows ultra high speed connections between offices over fiber optic cable. The 15454Es kit is used to Time Division Multiplex multiple Gigabit Circuits onto an OC48. Gigabit circuits between 7609s are port channeled. Port channel groups between routers are 802.1q trunks (Trunks are used to carry traffic belonging to multiple VLANs between devices over the same link. A device can determine which VLAN the

traffic belongs to, by its VLAN identifier. The VLAN identifier is a tag that is encapsulated with the data. ISL and 802.1q are two types of encapsulations used to carry data from multiple VLANs over trunk links.)[2]

2.2.5 Services Supported on GrangeNet

With the implementation of GrangeNet II more services are being offered than just the traditional Research and Education(R&E) routed protocols. GrangeNet II supports three levels of service that are aligned with the lowest three layers of the OSI Reference Model.

The service offerings are:

GrangeNet Services are IPv4/IPv6 Unicast, Multicast, R&E access. They provide support for MPLS, Quality of service (QoS), Grid services, ISCSI attached storage and all Streaming data.

- *GrangeNet R&E - Layer 3 (Network Layer)*
 - GrangeNet R&E is the traditional layer 3 services that permits members to peer with both local and international Research and Education members.
 - IPv4 & IPv6
 - Unicast
 - Multicast
 - Member BGP peers with GrangeNet routers

- *GrangeNet LAN - Layer 2 (Data link layer)*
 - An extension to the current offering of MPLS is GrangeNet LAN - a Layer 2 service.
 - Traffic is carried across the GrangeNet backbone links in dedicated VLAN's.
 - No capacity or quantity constraints on the LAN service.
 - A client can request many LAN services and have these combined with an R&E.
 - 802.1q service is provided to members to select path
 - A VLAN can be extended to any POP

- *GrangeNet Lightpath - Layer 1 (Physical Layer)*
 - Require dedicated fiber to each member
 - Gig E circuits are removed from the inter capital trunks to combine on OC48 /OC198 lightpaths
 - Layer 1 throughout the network – completely transparent.
 - GrangeNet Lightpath is a Layer 1 clear channel service that transports data between member endpoints. Available in either a 1Gbps or 50 – 200 Mbps variant it provides
 - The implementation of the Lightpath service is with TDM (Time Division Multiplexing) GrangeNet backbone.
 - Data carried in this service does not travel through any of the GrangeNet routers.
 - Connectivity:- User should make arrangement / connection to two GrangeNet POPs

2.2.6 Cost

- Request for quotation (RFQ) need to be raised with GrangeNet Office.
- Unlimited bandwidth at a low price is provided by GrangeNet. GrangeNet does not charge for traffic volume between the connected sites.
- Physical connection: - the Client is responsible for all costs involved in the provision of network tails between our premises and the GrangeNet PoP, interface hardware required to connect to the GrangeNet PoP and fees due to other carriers and suppliers. Currently the AARNET will be the responsible to provide the physical connectivity and the required bandwidth between the sites and the POP at the respective cities.
- Service Level Agreement: - No service levels are offered by GrangeNet in the current state.[2]

SECTION 3 SURVEY OF PROTOCOLS AND MECHANISMS FOR LONG DISTANCE AND HIGH SPEED NETWORKS

Standard TCP (TCP Reno) is a reliable transport protocol that is well tuned to perform well in traditional networks. However, several experiments and analysis have shown that this protocol is not suitable for bulk data transfer in high bandwidth, large round trip time networks because of its slow start and conservative congestion control mechanism. In this document, we review and compare different emerging alternatives that try to solve this problem in this particular context of very high speed networks. We believe that these innovations are phasing into experimental networks and replacing the stock TCP in high performance networking applications.

Existing transport protocols have limitations when they are used in new application domains and for new network technologies. For example, multimedia applications need congestion control but not necessarily ordered reliable delivery. This combination is not offered by TCP [3] or UDP [4]. From another point of view, TCP has been highly tuned with certain assumptions in mind. For example, when a data segment is lost, it assumes that this was most likely due to congestion (i.e. too many segments are contending for network resources) [5]. But, for example, in wireless it could be because of bad reception at the location of the user. So many efforts have been proposed for improving TCP performances in such lossy systems. TCP performance depends upon the product of the transfer rate and the round-trip delay [6]. TCP survived the days of low bandwidth, high latency, and high error rates. But for several reasons it is today not able to cope efficiently with the evolving new environment.

In this section, we address the specific problem of transport of bulk data transfers in grid environment (more than 1Gbyte) over high latency, high bandwidth, and low loss paths. For a Standard TCP connection with 1500-byte packets and a 100 ms round-trip time, achieving a steady-state throughput of 10 Gbps would require an average congestion window of 83,333 segments, and a packet drop rate of at most one congestion event every 5,000,000,000 packets (or equivalently, at most one congestion event every 1 2/3 hours) [7]. This is primarily due to its congestion avoidance algorithm, based on the Additive Increase Multiplicative Decrease (AIMD) principle. A TCP connection reduces its bandwidth use by half immediately when a loss is detected (multiplicative decrease), but takes 1 2/3 hours to use all the available bandwidth again in this case if no more loss is detected in the meantime. Apparently Standard TCP does not scale well in high bandwidth, large round-trip time networks. A lot of effort is being spent on improving performance for bulk data transfer in such networks. To solve the aforementioned problems, two main approaches are proposed: One focuses on a modification of TCP and specifically the AIMD algorithm, the other proposes the introduction of totally new transport protocols. This is a very active research area in networks. In very high speed context, performance or loss can also occur due to problems on the host (sender or receiver side). We do not consider these problems in this document.

3.1 METHODOLOGY AND COMPARISON CRITERIONS

3.1.1 Functions and features of a transport protocol

Transport layer protocols provide for end to end communication between two or more host. The references [8 to18] presents a tutorial on transport layer concepts and

terminology and a survey of transport layer services and protocols. It classifies the typical services provided by a transport layer. A transport service abstracts a set of functions that is provided to the high layer. A protocol, on the other hand refers to details of how a transport sender and a transport receiver cooperate to provide that service.

3.1.2 *Comparison criteria*

Here we list the criteria we concentrate on when we review and compare these protocols.

- **Performance:** Standard TCP is not suitable in high bandwidth, large RTT networks because of its low performance in throughput. Therefore, the performance of new protocols should be comparable to TCP in low bandwidth or small RTT networks and much better than TCP in high bandwidth, large RTT networks.
- **Congestion Control:** Existence of congestion control mechanisms is critical in avoiding congestion collapse. It is important to include reasonable congestion control mechanism if the transport protocol will be used in Internet or other best effort public networks. However, is congestion control still necessary in private networks where quality of service is guaranteed?
- **TCP friendly:** The term “TCP-friendly” or "TCP-compatible" means that a flow that behaves under congestion will behave like a flow produced by a conformant TCP. A TCP-compatible flow is responsive to congestion notification, and in steady-state uses no more bandwidth than a conformant TCP running under comparable conditions (drop rate, RTT, MTU, etc.). If we strictly abide by this requirement all the time, we will be disappointed again in less congested, large RTT networks. In this document, we only evaluate whether the protocol is TCP friendly in high-congested networks. All protocols in this document are not and should not be TCP friendly in LFP (Long Fat Pipe) networks.
- **Intra-Protocol Fairness:** There are two kinds of fairness: inter-protocol fairness and intra-protocol fairness. The former is the fairness when the protocol competes with TCP connections. The latter is the fairness among the connections using the protocol. The inter-protocol fairness is the same issue as TCP-compatible. Intra-protocol fairness will be compared among protocols.
- **Easy to deploy:** When we have plenty of bandwidth in underlying networks, what applications need immediately is to deploy transport protocols to utilize the huge bandwidth. In the protocols we are comparing, some need to modify or rebuild the operating system kernel, others are just a user level library which applications can call immediately.
- **Predictable:** When we say a protocol is predictable, applications should be able to predict its performance based on current network conditions such as available bandwidth, round-trip time, etc. The purpose is two-fold. Firstly users can tell whether the transport protocol is running correctly by comparing the prediction and actual performance. Secondly protocol developers can systematically identify the factors that influence the overall performance and

predict how much benefit any potential enhancement in the protocol might provide. Usually predictability is provided by creating a mathematical analytical model for the protocol.

- **Target Usage Scenario:** “One size fits all” is good but also difficult to accomplish. Before the network speed grew beyond 10 Mbps several years ago, TCP is almost a “One Size fits all” transport protocol. Now it’s time to find other solutions for bulk data transfer in LFP networks. These solutions have different preconditions or assumption on underlying networks. Some protocols don’t implement congestion control and can only be used in private or QoS-enabled networks. Other seems to be able to coexist with each other and with TCP traffic.
- **Deployable in Internet:** TCP variants want to take the place of the current standard TCP. In order to be deployable in the Internet or public networks, they create mathematical analytical models and carry out simulations to prove the fairness and have not the tendency to cause congestion collapse and therefore adopt sophisticated congestion avoidance algorithms. However, the main motivation of reliable UDP variants is that they are easy to use with good throughput. Usually they are used by a very small amount of users who own a lot of bandwidth. Their applications run on such private networks and seek to utilize the bandwidth as much as possible. They don’t intend to substitute TCP in the Internet.
- **Easy to deploy:** Usually reliable UDP variants provide a C or C++ user space library which high performance applications can call. The users don’t need to modify or reconfigure the operating system. Instead, TCP variants need to patch and rebuild the kernel, which only system administrators can do.
- **Efficiency:** Generally TCP variants are implemented in kernel space whereas reliable UDP variants are implemented in user space. Kernel mechanisms are more scalable and provide better efficiency. [8-18]

SECTION 4 PERFORMANCE WITH FAST TCP

File transfers and replications where a large file need to be simultaneously transferred or replicated to different repositories sites is an integral part of data intensive grid environment. Currently adopted data transport mechanisms such as GridFTP use regular TCP stack and is mainly created for point-to-point file transfer. In this section we implement FAST TCP stack instead of regular TCP stack and presents FAST TCP as an improved new TCP protocol useful for long distance high-speed links. A data transfer is one of the most critical components of a data-intensive grid computing environment.

The need for this arises in various areas of data analysis such as high-energy physics, bio-informatics, climate modeling and astronomy. For example, terabytes and petabytes of data produced by CERN have to be shared and accessed by the high-energy physics community around the world. In addition to grid data environments, data replication is the key part of various data-sharing applications such as digital libraries, persistent archival environment and content distribution. Various architectures are being proposed and developed to manage data replication (Data Grid [19], European Data Grid [20], Datafarm [21]). Data Grid Reference Architecture (DGRA) [22] covers important architectural components and their functionalities. One of the key parts in data transfers is the replica catalog that manages the mappings for files from the hierarchical namespace to one or more physical file locations, thus providing an efficient and transparent file sharing on a Grid. Managing and coordinating the data movement process is the crucial performance issue. Current strategies use data locality, access time and pattern to decide whether to move computation to data source or vice versa.

In addition to these strategies, the network (transport) mechanism used in the actual movement of the data plays an equally important role in the overall performance. The access time in data replication in general depends upon how the network resources are utilized by the data transport mechanism. Currently, GridFTP is an accepted data transport choice of the Grid community (other data transport tools, such as SABUL [23], fast TCP [24], BLAST UDP [25], are also available). GridFTP is designed for point-to-point reliable data transport based on file splitting and opening multiple parallel TCP streams. However, it will result in performance degradation (increased latency) for data transfer over the links if/when some of network resources are bottlenecked if maximizing the use of available bandwidth is not implemented. FAST TCP is one such protocol proven to have improved the available bandwidth and is well suited for high-speed and long distance network links. [26]

4.1 FAST (FAST AQM (ACTIVE QUEUE MANAGEMENT) SCALABLE TCP)

FAST is built on the idea of TCP Vegas. TCP Vegas was introduced as an alternative to the standard TCP (TCP Reno). Vegas does not involve any changes to TCP specification.

It is merely an alternative implementation of TCP and all the changes are confined to the sending side. In contrast to the standard TCP, which uses packet loss as the measure of congestion, Vegas source anticipates the onset of congestion by monitoring the difference between the rate it is expecting to see and the rate it is actually realizing. Vegas' strategy is to adjust the source's sending rate in an attempt to keep a small number of packets buffered in the routers along the path. Although experimental results show that Vegas

achieves better throughput and fewer losses than standard TCP, Vegas lacks a theoretical explanation of why it works. At engineers of net labs of Caltech university develop a model of Vegas and show that Vegas can potentially scale to high bandwidth network in stark contrast to the standard TCP. Further, they show that Vegas can become unstable at large delay. Also error in RTT estimation can distort Vegas and can lead to persistent queues and unfair rate allocation. They show that by augmenting Vegas with appropriate Active Queue Management algorithm like Random Exponential Marking (which requires modification in the router), it is possible to avoid the abovementioned problems. FAST TCP aims at solving those problems by modifying just the TCP kernel at the sending hosts. Detailed description of the algorithm and implementation of FAST TCP is well published. [24]

Supported operation mode is Memory to memory (general transport) with no authentication. Its Congestion Control Algorithms is FAST TCP is built on the algorithm used in TCP Vegas. Fair bandwidth allocation is one of the main objectives of FAST TCP but the detail about the mechanism is published well.

TCP Friendly was motivated by their earlier work, which developed a TCP/AQM congestion control system to achieve high utilization, low delay and dynamic stability at the level of fluid-flow models. But the algorithm used in FAST TCP and the theoretical explanation is well published. FAST TCP test results conducted by Caltech University show a very efficient use of bandwidth in large delay and large RTT network link can be optimized. [24]

FAST TCP was demonstrated in a series of experiments conducted during the Super Computing conference (SC2002). The demonstrations used an OC192 (10Gbps) link between Star-Light (Chicago) and Sunnyvale, the Data-TAG 2.5 Gbps link between Starlight and CERN (Geneva), an OC192 link connecting the SC2002 show floor in Baltimore and the TeraGrid router in StarLight Chicago and Abilene backbone of Internet2. Using default device queue size ($txqueuelen = 100$ packets) at the network interface card and the standard MTU of 1500 bytes, the default Linux TCP (2.4.18), without any tuning on the AIMD parameters, achieved an average throughput of 185 Mbps, averaged over an hour, with a single TCP flow between Sunnyvale in California and CERN in Geneva via StarLight in Chicago with a minimum round trip delay of 180 ms. This is out of a possible maximum of 973 Mbps to the application, excluding TCP/IP overhead, limited by the gigabit Ethernet card, and represents a utilization of just 19%. Under the same experimental conditions, using the default device queue size ($txqueuelen = 100$ packets) and the standard MTU of 1500 bytes, FAST TCP achieved an average throughput of 925 Mbps (Utilization 95%), averaged over an hour. Even with a device queue size of 10,000 packets, the standard TCP was able to achieve a throughput of only 266 Mbps (Utilization 27%). With 2 TCP flows sharing the path, standard TCP was able to achieve 48% utilization ($txqueuelen = 10,000$ packets) whereas FAST TCP was able to achieve 92% utilization. With 10 flows, FAST TCP achieved a throughput of 8,609 Mbps (utilization 88%), averaged over a 6-hour period, over a routed path between Sunnyvale and Baltimore, using the standard MTU. The results using the standard Linux TCP Implementation for 10 flows are not shown.

In all the experiments described above, the bottleneck was either the gigabit Ethernet card or the transatlantic OC48 link. The experiments conducted using Intel's pre-release experimental 10- gigabit Ethernet card on a single flow from Sunnyvale to Chicago using

standard MTU, FAST TCP sustained just 1.3 Gbps. They claim this was due to the limitation in the CPU power at the sending and receiving systems.

Target Usage Scenario: Though it is intended to solve TCP's limitation in high bandwidth large-delay environments, it is expected to perform well in conventional environments too.

4.3 FTP BETWEEN SERVER AND CLIENT ON GRANGENET WITH FAST ETHERNET REPOSITORIES

A test bed between a fast Ethernet subnet at Monash university network and Fast Ethernet subnet on University of Queensland was setup to perform the file transfer test using the standard TCP stack and FAST TCP stack.

As the amount of information available over the networks has been increasing, so have the methods by which this information can be obtained. No longer is direct usage of FTP the only, or even the most frequent, method of obtaining data; we now have several tools with modified or newer TCP algorithms like FAST TCP to transfer data with site-specific interfaces. Many organizations provide data formats for representing scientific data such as W3C. The following data format gives an insight on some of the W3C data format specifications.

CDF (Common Data Format), FITS (Flexible Image Transport System) GRIB (GRid In Binary) HDF (Hierarchical Data Format) NetCDF (Network Common Data Form) VICAR (Video Image Communication and Retrieval) Planetary Data System (PDS) Miscellaneous graphics formats storing graphics files -- TIFF, GIF, JPEG, FLI, CGM.. SAIF (Spatial Archive and Interchange Format) SDTS (Spatial Data Transfer Standard) HDS (Hierarchical Data System) The Medical File Standard (MedFileS) CXF provides representation of chemical substances and queries, including atoms, fragments, molecules, and reactions JCAMP is a draft standard for spectra data (IR & NMR) CIF (Crystallographic Information File) is becoming standard in the crystallography world and related fields Digital Elevation Model (DEM) and many more. For data transfer test purposes we have used data formats with the two types of data representations such as lose less compression and losy compression. We have used zip, rar, tar, avi, vob and mpeg format of data.

FTP between Server to Client on fast ethernet locally and Grangenet between the repositories								
File Size in Bytes	FAST TCP Transfer speeds KB				TCP Transfer speeds KB			
	Low	Average	High		Low	Average	High	
1073676288 (1 GB)	10900	11278	11304		7606	8934	9466	
590559232 500 MB	10768	11080	11577		7933	8966	9548	
440481792 440 MB	10103	11181	11512		7894	9077	10593	
270893056 270 MB	10990	11029	11553		6891	8914	10400	

Table 1: Table of Data transfer test between Server and Client on fast Ethernet on local subnets and Grange Net between the repositories

Using Iperf tool the link throughput between the server and client on fast Ethernet on local subnets and Grange Net between the repositories is 92 Mbits/sec with FAST TCP enabled and 77 Mbits/sec with standard TCP stack. The peak and off-peak latency and throughput remained the same throughout the tests.

Transfer speed:-

With FAST TCP

11304x1024x8 = 92602368 bits/sec = 92Megabits/sec

With standard TCP

9466x1024x8 = 77545472 bits/sec = 77 Megabits/sec

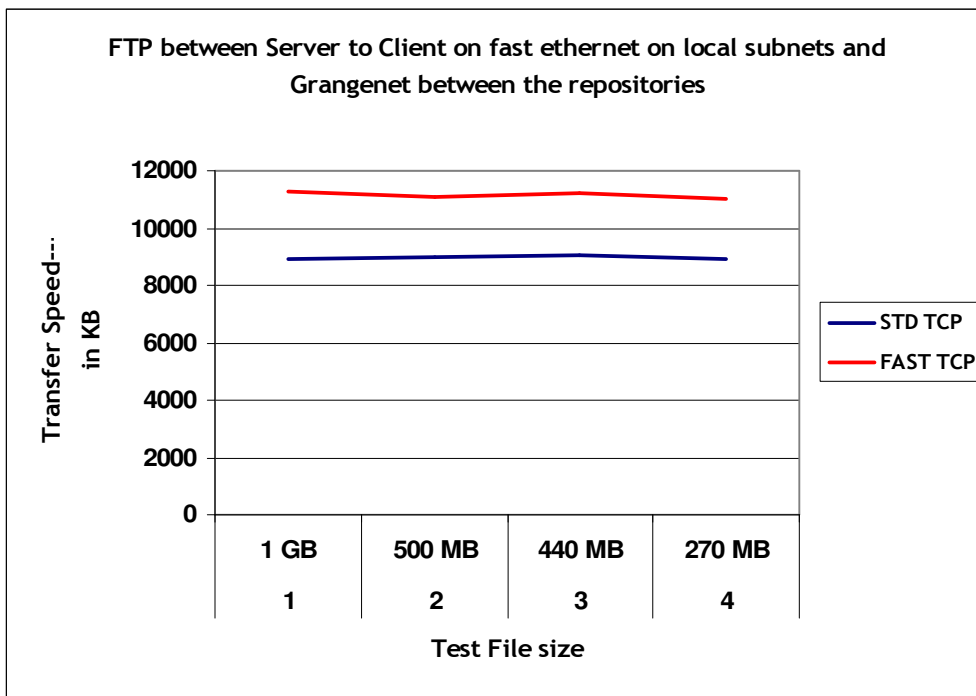


Figure 8: Data transfer test between Server and Client on fast Ethernet on local subnets and Grange Net between the repositories

4.3 FTP BETWEEN SERVER AND CLIENT ON WAN LINK

FTP between server to client on WAN link (Peninsula & Clayton)							
File Size in Bytes	FAST TCP Transfer speeds KB			TCP Transfer speeds KB			
	Low	Average	High	Low	Average	High	
1073676288 (1 GB)	1400	1500	2071	602	1055	1700	
590559232 (500 MB)	1445	1666	2074	667	1103	1769	

440481792 440 MB	1488	1676	2098		689	1213	1800	
270893056 270 MB	1492	1632	2100		634	1237	1793	

Table 2: Table of Data transfer test between server to client on WAN link (Peninsula & Clayton)

Using Iperf tool the link throughput between the Clayton and Peninsula between the test points is 37.9 Mbps during weekends (of-peak) and 18.7 Mbits/sec during the week days. (peak).

Average transfer speed:-

With FAST TCP

2071x1024x8 = 16965632 bits/sec = 17 Megabits/sec

With standard TCP

1700x1024x8 = 13926400 bits/sec = 14 Megabits/sec

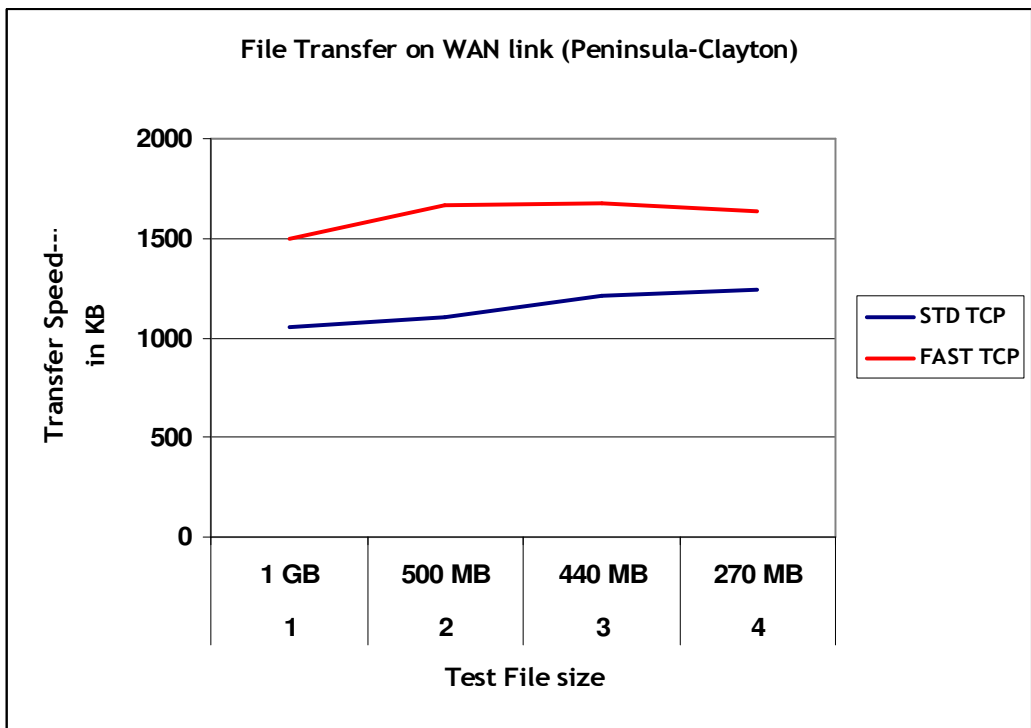


Figure 9: Data transfer test between Server and Client on fast Ethernet on local subnets and WAN link between the repositories

SECTION 5 BENEFITS OF FAST TCP

The file transfer tests using FTP protocol conducted on the above two scenarios using standard TCP and FAST TCP show the performance boosts of 20-30% in FAST TCP

scenario. FAST TCP is an alternative congestion control algorithm in TCP. It is designed for high speed data transfers over large distance, e.g., tens of gigabyte. Our current implementation is in TCP and FAST TCP on Linux platform. The experimental results showed performance improvement with FAST TCP and on the use of high-speed long distance GrangeNet.

5.1 THE BENEFITS OF FAST TCP

TCP is useful whenever the availability of resources and the set of competing users / available bandwidth vary over time unpredictably, yet efficient sharing is desired. In order to get high throughput end-to-end between data points / applications, we must have a network infrastructure that can provide large raw capacity, and a protocol that can make efficient use of the high capacity infrastructure. The efficiency of the (congestion control algorithm in the) current TCP implementation drops steadily, and the protocol eventually becomes a performance bottleneck itself, as the network infrastructure scales up in capacity. FAST TCP aims to remove this bottleneck: it is scalable to networks with large bandwidth-delay product. We must note that FAST TCP does not solve the infrastructure problem; if the underlying hardware (network resources & Datacenters) have low performance and speed, no improved TCP stack implementation can increase the throughput beyond the limit imposed by the underlying hardware. FAST TCP has been optimized to provide great performance improvement over the current TCP implementation at Gbps speed. In low speed networks, e.g., if the bottleneck in the end-to-end path is the 10Mbps or 100Mbps Ethernet card, we expect the current TCP implementation to be quite efficient, so there is not much to improve. If the performance of the current TCP implementation is poor even at such speeds, then FAST TCP may provide significant improvement depending on the reason for the poor performance. [24]

5.2 EASE OF INSTALLATION / IMPLEMENTATION OF FAST TCP STACK ON LINUX KERNEL

FAST comes as two separate patches, an operating system (OS) independent patch and an OS dependent patch. The OS independent part contains the FAST algorithm and is covered by a patent and the Caltech open source agreement. The OS dependent part simply interfaces hooks into the TCP implementation of the particular OS to the separate FAST algorithm. The OS dependent part is not restricted by the Caltech open source agreement but covered under the GPL (Gnu Public License). Currently, FAST is available as a patch to the Linux kernel. We need to have the correct version of the Linux kernel source installed on our machine (that is version 2.6.7 or 2.6.15). The procedure is to install/upgrade to new kernel 2.6.7 / 2.6.15 and recompile with FAST OS independent patch and FAST OS dependent patch; along with Fast Header and Module support. The current recompiled working version is on Linux version 2.6.15.

Configure FAST requires to set up the TCP send and receive buffers for high performance, set up the network device buffers for high performance. We also need to modify certain network device configuration parameters. The FAST patch is preconfigured with default parameters which are modified for our situations. Detailed fine-tuning parameters with shell script have been written for this implementation.

5.3 THE EASE OF FAST TCP USE

Once FAST TCP stack has been implemented; it will act and behave just like any TCP in the sense that any application, such as FTP, SSH, SCP and all other file transfer protocols which uses standard TCP will now use FAST TCP. We don't need any special programs/scripts/activation to use FAST TCP. This is straight forward simple implementation.

5.4 OPERATING SYSTEM REQUIREMENTS

Currently FAST TCP is available to Linux Operating System with kernel support for 2.6.7 and 2.6.15. I have implemented and tested on Scientific Linux 4.2 and 4.3 versions and on Fedora core 4.0 versions. It supports all Linux flavors provided the kernel version is upgraded to 2.6.7 and 2.6.15 and the FAST TCP patches are recompiled with these kernel versions.

5.5 FAST TCP INSTALLATION FOR LINUX KERNEL

A step by step installation procedure is explained in Appendix as section 14

5.6 STABILITY OF FAST TCP

I have implemented and used on five different Linux versions and done Data transfer test over three month period. So far I have not faced any problems and the new TCP stack seems to work fine.

5.7 NEED TO IMPLEMENT FAST TCP

TCP Transport layer protocols provide for end to end communication between two or more host. The Standard TCP (TCP Reno) is not suitable for high bandwidth, large RTT networks because of its low performance in high throughput. Standard TCP is a reliable transport protocol that is well tuned to perform in traditional networks. TCP performance depends upon the product of the transfer rate and the round-trip delay. TCP survived the days of low bandwidth, high latency, and high error rates. But for several reason it is not able to cope efficiently with the evolving new environment. Several experiments and analysis have shown that this protocol is not suitable for huge data transfer in high bandwidth environment such as Gigabit and high-speed long distance networks with large round trip time. This is because of its slow start and conservative congestion control mechanism. Our focus on introduction of new transport protocols FAST TCP is to improve performance when compared to standard TCP in low bandwidth or small RTT networks and much better than standard TCP in high bandwidth, large RTT networks. Grid Applications need immediate deployment of modified transport protocols to utilize the huge bandwidth. In this Workpage-SI8 Dart project, we have reviewed, experimented and compared different emerging alternatives that solve this problem, in particular context of very high speed networks.

5.8 IMPLEMENTATION OF HS-TCP, SCALABLE TCP AND FAST TCP WITH GRIDFTP

The standard protocol for network data transfer remains TCP. However, TCP's congestion avoidance algorithm can lead to poor performance, particularly in default configurations and on paths with high round trip times. Solutions to this problem include careful tuning of TCP parameters [27, 28], TCP protocol improvements [29], multiple "parallel" TCP connections [28, 35], and the substitution of alternative protocols [30-33]. We want to support such alternatives.

This report [34] has investigated how High-Speed TCP, Scalable TCP and FAST TCP behave with a real Grid application, GridFTP. Other studies have shown the results of HS-TCP in a simulated environment [27], but it is useful to see what happens in a real environment with real applications. In this report[34] results have been compared using standard TCP with HS-TCP, with respect to congestion window. Congestion window recovers from loss much faster using HS-TCP compared to standard TCP. This faster recovery results in higher throughput for the network application. It is a well known fact that standard TCP Reno does not scale to large bandwidth delay product networks. Therefore modification or implementation of standard TCP stack such as HSTCP / FAST-TCP / Scalable TCP with GridFTP will result in better utilization of bandwidth and higher throughput. [34, 35]

Data GRID project at UK particle Physics in 2003 used GridFTP to transfer large amounts of data between Mass Storage Systems in Europe and world wide. The demonstration sent large data from CERN to NIKHEF/SARA at high data rates on European NRN/Geant networks and were able to achieve high throughput with Tuned TCP parameters, Interface txqueuelen, TCP buffer size to match the $BW * rtt$ and with Different TCP stacks; Standard TCP, Fast TCP and Scalable TCP. This experiment has highlighted the fact the use of GridFTP with other TCP stacks in projects: DataTAG, Mb-NG, UK- Star- Nether- Light.

Experimental studies at data transport and data access of earth science data over high speed networks using GridFTP on current high performance data transport protocols such as SABUL/UDT, FAST, etc., can move data significantly faster than can be done with regular TCP. But the performance is significantly dropped when accessing data from disk. Hence implementation of modified TCP stack along with high performance computing machines is essential for higher throughput and link performances. [36]

	HS TCP	Scalable TCP	FAST TCP	XCP	CADPC/PTP (Congestion Avoidance with Distributed Proportional Control) Performance Transparency Protocol (PTP)	GridFTP
Principle	Modify TCP response function when congestion window is larger than a threshold.	More aggressive in congestion control.	Based on TCP vegas.	Decouple efficiency control and fairness control. Former uses MIMD and latter uses AIMD.	Congestion control based on feedback from routers.	Multiple parallel TCP streams
Operation Mode	Memory-to-memory. Kernel space.	Memory-to-memory. Kernel space.	Memory-to-memory.	Memory-to-memory.	Memory-to-memory.	File transfer protocol
Authentication	No.	No.	No.	No.	No.	GSI and Kerberos
FTP syntax	No.	No.	No.	No.	No.	Yes.
Congestion Control	Yes.	Yes.	Yes.	Yes.	CADPC	Yes.
Need to modify router software	No.	No.	Much Better if using AQM routers (All recent routers support)	Yes.	Yes.	No.
Fairness	Need more investigation	No.	Yes.	Yes.	Max-min fairness	Yes.
TCP Friendly	No when new TCP response function is triggered.	Need more investigation.	Yes.	Yes.	No.	Yes.
Predictable Performance Model	Yes.	Yes.	Yes.	Yes.	Yes.	No.
Simulation and Implementation	Both	Implemented	Well Published	Both.	Both.	Implemented

Table 3: Table of protocols compared

5.9 IMPLEMENTATION OF FAST TCP FOR FULL PRODUCTION ENVIRONMENT

The full production environment needs to implement all stages of development from the Dart project. This would require the following plan of action:-

- The High performance data centers with recommended requirement of hardware and software. These should comply with the specifications that would be outlines in the Dart project report. The activity for this stage would be to raise a RFQ with the detailed specifications and acquire the hardware.
- The targeted deployment environment need to be Linux based.
- Install the required O.S, configure and fine-tune the high performance data center server. Modify the standard TCP stack with FAST TCP to maximize the link usage.
- Configure/Install other network storage links and other devices such as (Tape backup's, SAN links/devices, UPS systems, RAID-5 etc)
- Install any hardware solution for Data integrity and Data Security-processing component. This would be a new security architecture hardware implementation such as layer-2 layer-3 or IPSec or SSL/TLS data encryptors / decryptors. Or configure open source software solutions for data integrity and security.
- Collect and collate all the data patterns to do real-time data test transfers between the high performance points. We need to keep in mind that these tests have been performed in DART as proof-of-concepts with existing computer nodes; which do not have the required hardware /software requirements. This can be seen as the major difference between the test-bed and full-production environment, and the end results would be highlighted.
- The High-Performance data center server would require a bigger pipe to connect to the point-of-presence (POP) for the AARNet or Grange Net which provides the access to high-speed research networks.
- Rigorous testing of high speed data transfers will be performed with network management and monitoring tool such HP OpenView or any other equivalent network management software that is made available will be used to capture traffic traces and network link topology. This is required to validate the network models built in Dart. This model will work as a baseline study for future performance evaluation. This study is needed to get insights on performance of network with introduction of new data patterns, bandwidth bottlenecks, and what-if scenarios.

SECTION 6 REQUIREMENT ANALYSIS OF HIGH AVAILABILITY DATA CENTER

6.1 HIGH AVAILABILITY DATA CENTER

There are many types of data centers ranging from corporate data centers in a wide range of industries, banks, telecommunication facilities, and Internet hosting facilities. Data centers are also found in other institutions such as research organizations, universities, national laboratories, and government facilities.

The work package development identified many areas where significant efficiency gains could be achieved through adoption of current best practices, better application of existing technology, and research into new technological solutions. These areas are as follows:

Activities aimed at understanding the Data Center Market – The size and growth rate of the market as well as local concentrations of data centers will be of interest to planners and implementers in production environment.

The benefits of obtaining energy utilization benchmarks – By monitoring and comparing the utilization and consumption of a variety of data centers, operators and designers will be able to learn what is possible to achieve.

Identify and promote the best practices – Adopting current best practices in existing or new data centers to provide significant improvement in the short term and plan for long term data centre requirements.

Improving data center facility systems' efficiency - Facility systems containing resources such as processing power, storage, access grid facility, Uninterruptible Power Supplies (UPS) are far from optimal.

Improving the interface between building systems and IT Equipment-The systems that house and support electronic equipment in data centers are typically not designed to optimize the efficiency of the building infrastructure systems they interface with.

Improving the efficiency of IT Equipment - Energy use in data centers is dominated by the servers, storage networks, routers, and switches that are used to process, store, and transmit data. Lastly, the electronic equipment used by this industry is continually evolving.

As enterprise data centers plan through consolidation phases toward next generation architectures that increasingly leverage virtualization technologies, the importance of very high performance Ethernet switch/routers will continue to grow. The high volume of Ethernet products continues to spur rapidly declining prices and a constant stream of enhancements/innovations including, 10 GbE WAN PHY, Ethernet MAN/WAN services, iWARP RDMA/TOE NICs, 10GBase-T for 10 GbE over twisted pair, and the next generation of the Ethernet bandwidth hierarchy at 40 or 100 Gbps.

These developments testify to the unparalleled flexibility and resilience of Ethernet and TCP/IP to respond to challenges from competing technologies. As Ethernet evolves over the next two to three years, it is likely to significantly enhance its standing as the

switching fabric of choice for data center IPC and as a highly competitive SAN fabric. As these developments unfold, data center architectures based on a unified switching fabric can be implemented without any of the performance compromises that exist today.

Fabric unification at the highest level of performance for all data center applications will serve as the basis for future waves of data center consolidation and architectural evolution. In today's high performance data center, where information transformed into answers is the currency, information technology is regaining its role as a strategic asset that enables competitive advantages. Continued data center evolution will ensure businesses are prepared to meet the new challenges and opportunities of the emerging answer economy.

Requirement analysis of high availability data center secondary repositories servers should be combined with resource management software, Capacity-on-demand services, maximize Data/Application uptime and to meet the following performance goals.

- The need for high availability
- Information technology has become a pervasive presence in the research environment. A computer system that is not functioning will have a negative impact on the research. Data repositories are integral parts of and research activity. Therefore, the cost of the measures taken to ensure that data access remains available and response time's stay within a specified range must be commensurate with the financial impact on the research.
- Causes of Downtime
- High Availability vs. Fault Tolerance: Fault tolerance differs from high availability by providing additional resources that allow an application to "ride through" a failure without interruption. Many of the high-availability solutions on the market today actually provide fault tolerance for a particular application component. Disk mirroring, where there are two disk drives with identical copies of the data, is an example of a fault-tolerant component. If one of the disk drives fails, there is another copy of the data that is instantly available so that the application can continue execution. However, once such a failure occurs, the system becomes vulnerable to the failure of the single remaining disk drive, which now has the only copy of the data and represents a single point of failure. Action should be taken as soon as possible to create a mirror of the remaining disk drive. However, this process may have a negative impact on system performance, depending on where the processing to re-mirror the drive takes place. A fully fault-tolerant solution requires that all the resources that the application is dependent on be replicated, including the application process itself. This requires an independent processor (not part of the same symmetrical multiprocessing system) and a copy of the memory that the application uses. In the worst-case failure scenario, one in which the processor or memory fails, the replicated version of the application continues to execute. Other failures simply require the application to use alternate resources (disk drives, disk adapters, communications devices). As a result of this complete hardware and process replication, fault-tolerant systems are significantly more expensive than highly available systems. A fault-tolerant system would be used in a situation where no downtime can be tolerated at

all, such as an air-traffic-control system, an emergency-response system or financial trading systems. In evaluating a fault-tolerant system, particular attention should be paid to the repair process. While the system may be capable of riding through a failure, to ensure that a subsequent failure will not bring the system down, the failed component must be repaired. The question to ask is whether the repair can take place while the system is still running critical applications – the reason for the system in the first place.

- Reducing Complexity to Increase Availability
- Virtualization to increase Availability: Increase Virtualization is a software technology that makes better use of resources. Virtualization provides the ability to aggregate the same type of resources into a single logical pool.
- Reducing Downtime Due to Hardware failures: Many hardware failures first manifest themselves as operating system crashes. Most operating systems will automatically attempt to reboot following a crash. During reboot, built-in self-test, power on self-test and other initialization routines will detect failed components and eliminate them from the configuration. After these components are removed, if there are enough resources for the system to operate, the operating system will reboot. For processor errors in a symmetric multiprocessing (SMP) system, this is a real advantage. Although the system will have less capacity and lower performance (for a two-processor SMP, this translates to 50 percent of its original capacity; for a four-processor SMP, this would be only a 25 percent reduction), the system may still be capable of executing the most critical applications. Instead of having the system idle until repair personnel arrive with the appropriate spare parts (usually measured in hours), the system can continue to operate, albeit in a degraded mode. Furthermore, the time when the system is taken down for repair can be chosen to minimize the impact of a complete outage. This feature is supported by many vendors on their systems and goes by a variety of names, including processor/memory failover. With the capacity-on-demand capabilities that are frequently found in midrange and larger servers, even the performance degradation due to a processor failure may be eliminated.
- Memory: Undetected memory errors can have serious consequences. At best, they can simply lead to system crash. Error correction code (ECC) memory is offered on virtually all servers. With ECC, a checksum is calculated and stored in additional bits appended to the data. When the data is subsequently read, a checksum is again calculated. If the checksums match, then the data is correct. If the retrieved data is incorrect, the algorithm by which the checksum is calculated can correct a single bit that has changed — the most usual case. Multibit errors are detected but cannot be corrected. With the increase in the size of caches, users should also look for ECC protection there and on internal data paths as well. Chipkill memory is an enhanced ECC memory technology that protects systems from multibit errors caused by the failure of the entire dynamic random-access memory. Chipkill memory is analogous to the redundant array of independent disks (RAID) for disks. In fact, Chipkill has also been called RAID for memory. Chipkill memory is available on a number of server systems offered by a variety of vendors and sometimes may be called by a different name. Memories are susceptible to soft or temporary

errors caused by cosmic radiation. To decrease the likelihood of a memory failure caused by a multibit error, many servers provide memory scrubbing. Memory scrubbing uses idle processor cycles to crawl through memory to correct single-bit errors to prevent single-bit errors from becoming multibit errors that can cause a system outage. However, there is a downside to memory scrubbing. If a double-bit error is encountered in a little-used area of memory during the scrubbing operation, it may cause an operating system crash.

- **Environmental Systems:** Many servers are configured, or can optionally be configured, with redundant power supplies and fans. If a power supply fails, there is additional power capacity for system operation. Similarly, a redundant fan, perhaps operating at higher speed, will continue to provide adequate cooling for system operation when one fan fails. For both of these failures, the failed component needs to be repaired / replaced to ensure the same degree of failure resilience in the future.
- **Input/Output Devices:** Input/output (I/O) device failures are often intermittent. In many cases, an I/O operation that has failed will often be successful on a retry. The issue is how many times to retry the operation before declaring a failure. While more retries may ensure eventual success, I/O operations that require many retries before being successful will significantly degrade server performance.
- **Disk Drives For high-availability systems:** The RAID levels of interest are RAID 1 and RAID 3/5. RAID can be used to mask disk drive and disk controller failures. For high-availability systems, the RAID levels of interest are RAID 1 and RAID 3/5. RAID 1 is simply mirroring. If one drive in a mirrored set fails, the data can be retrieved from its mirror image. RAID 3/5 provides a more cost-effective method for protecting data. With RAID 3/5, a record is striped (written) across several drives, with an additional drive keeping ECC-like information. If any one drive fails, the entire record can still be retrieved. The ECC-like information is used to re-create the portion of the data that is on the failed drive. RAID 3/5 is a very cost-effective way to ensure against loss of data due to drive failures. On drive failure, however, the performance penalty may be significant until the drive is replaced and the RAID set reestablished. RAID can be implemented in either hardware or software. With RAID controllers (the hardware implementation), the incremental processing to write to multiple drives is embedded in the controller without requiring any additional CPU cycles. If the controller fails, however (this occurs much less frequently than a drive failure), all drives on the controller are lost, including any data copies. The real advantage of the software solution is that it is possible to choose to put the copy disk on a different controller, thus also protecting against controller failure.
- **Event Recording and Failure Analysis**
- **Hardware Repair:** provide hot-pluggable components
- **Hardware/Software Upgrades: Reducing Downtime Due to Software:** - These failures are caused by programming errors in the applications themselves as

well as in the underlying software — database, file system and operating system. Reducing Downtime Due to Hardware failures: Purchase additional/replaceable components.

- Maintenance agreements
- Clustering: Clustering for high availability ensures that an operating system crash, for any reason, does not cause lengthy application outages.
- Reducing Outages Caused by Environmental Stresses, Failures and Disasters: Keeping computer equipment running during power fluctuations and short-term outages can be accomplished using uninterruptible power supplies (UPSs). Use of Disk-mirroring technologies.
- Data protection: - Data-protection strategies. Provide RAID 1 (mirroring) or RAID 5 ways of protecting data.
- Price and Performance issues
- Study Selection Guidelines
- Clustered Solutions
- Fault-Tolerant Solutions
- Technology Leaders: compare in detail what different vendors offer for high-availability hardware, software and services.

SECTION 7 DATA ENCRYPTORS DECRYPTORS

7.1 GSI AND HARDWARE LAYER 2 – LAYER 3 ENCRYPTORS / DECRYPTORS

7.1.1 *GSI-based*

The SRB system currently employs different authentication mechanisms for its clients and servers: challenge response mechanism, GSI authentication based on the GSSAPI, and Kerberos authentication. Data can be encrypted and sent over unsecured lines without much overhead. The Grid Security Infrastructure (GSI) employs certificate mechanisms when used in Public Key Infrastructure (PKI) where a public and a private key pair are used to encrypt and decrypt information. The client and server are authenticated on every connection given the size of the certificates; this may imply an overhead of several kilobytes. In a typical exchange, some 14 messages are sent, varying in size between ~50 to ~2000 bytes. Tests have shown that full authentication on every connection does not scale well to a large number of connections per unit time: every connection takes about 28 ms to set up. These 28 ms are only due to the TLS/GSI authentication overhead, since no further processing was done in this test. Full authentication is therefore not a viable option. [37]

In the use of only GSI and TLS, one could use asymmetric keys and full GSI authentication only when requesting a subscription. As this time, a symmetric key could be exchanged securely, and this key is used to encrypt any data sent to the listening port on the client (subscriber) side. This further message exchange can then use a conventional block cipher. Since only the two parties/peers involved avail over the (symmetric) key needed to read the data, you do not need any further authentication. In fact, it is just extending the SSL session over multiple connections. The connection overhead is almost zero. The typical time needed to set up a TCP connection is approximately 70 μ s, about 380 times faster than a GSI authenticated connection. The crucial factor then is encryption speed. Since you need to establish the encryption context only once, the overhead involved in context generation can be neglected. On a typical system (PIII, 1.2GHz, memory-to-memory), the encryption speed is 37 sec/GByte, i.e., 34 ns/byte. The conventional block length is 8 bytes for the most popular block ciphers (CAST, IDEA, Blowfish, etc.)[37]

7.1.2 *GSI-based Test bed*

Test performed on a PIII 1.2 GHz dual-processor system over the loop back interface. All tests were memory-to-memory when relevant.

With PLAIN TCP CONNECTIONS

#connections made	clock time [ms]
0	2
1	3
2	3
100	9
1000	74
10000	687

With GSI CONNECTIONS

The GSI connections were self-secured, with message integrity and confidentiality enabled. The code was based on the sample implementations from Globus_io (connect and listen).

#connections made	clock time [ms]
100	2675
1000	27056

The average time per connection in the infinite limit is ~ 27056 μ s.

BULK CRYPTOGRAPHIC CAST TEST

A sparse-file of zeroes was encrypted using the CAST algorithm and the results were:

#bytes encrypted	clock time [ms]
1.07 MB	53
10.4 MB	385
104 MB	3722
1.0 GB	37171[1]

7.1.3 Hardware Encryption

However with the advent of the High Speed Data Encryption Processor (HSDEP) for both symmetric and asymmetric key cryptography; this has introduced new hardware encryptors and decryptors. These devices can deal with high speed encryption/decryption. The rapid growth of networking is driving high-bandwidth data transfers across the globe. All data transfers are carried over networks like LAN, WAN, and ATMs, which are interconnected with routers, switches, bridges, and other network equipment. The growth of virtual private networks (VPNs) and IP security solutions (IPSec) has raised the demand for secure, high performance data transfers. The question is how secure are these transactions and will this security feature be a bottleneck in networks? To protect privacy and safeguard sensitive data transfers between the peers it is recommended to use Data Encryption Standard (DES) and Triple DES algorithms [38], [39]. There are several encryption algorithms like DES, AES, Blowfish, RC4/RC5, International Data Encryption Algorithm (IDEA) and others, each having strengths and weaknesses. DES is the most widely used open standard cryptosystem, offering excellent performance. Above all, it is the recommended encryption standard by U.S. government agencies, making it the de facto industry standard. Several high-speed DES hardware implementations have been reported in technical journals [41-43]. Sandia National Labs announced their fastest DES custom ASIC operating at 6.7 Gb/s. [40],[41] The current implementations have surpassed all the existing DES implementations and offers up to 15 Gb/s, almost twice the performance of any other existing DES solution. The most directly comparable prior design implemented in a field programmable gate array (FPGA) has complete loop unrolling and encrypts at 3.05 Gb/s [42]. Concerns about DES vulnerability are driving further standardization efforts, so any encryption hardware that is deployed today might become obsolete in a few months. In contrast, an encryption engine residing in an FPGA could be updated in the field with a new encryption algorithm when it becomes available. Recently, the US Department of Commerce selected Advanced Encryption Standard (AES) as a replacement for DES. AES is not yet in widespread use. An FPGA solution is ideal in this case where compatibility with DES is required now and an upgrade to AES will be desirable in the future.

7.1.4 Data Encryption Performance: Layer 2 vs. Layer 3 Encryption in High Speed Networks

SafeNet is one such company that has developed the family of devices for encrypting high-speed communications. They offer data encryption at the Physical (Layer 1), Data Link (Layer 2), and Network layers (Layer 3), providing the ideal solution for implementing high-speed encryption without changing the existing network infrastructure.

The development of IPv6, which has room for cryptographic information in its header, has been a slow process, with its widespread deployment still a long way to go. In the meantime, the use of IPSec with IPv4 has become the standard for securing data transfer over the network. While IPSec is secure, it effectively shrinks the bandwidth available, requiring approximately 50 bytes per packet in overhead. The overhead tax can be costly for the enterprise, especially when considering non-routed, point-to-point solutions, where it is not essential for cryptography to be performed at the network layer. A study was performed at Rochester Institute of Technology (RIT) in an attempt to determine the performance gains achieved by encrypting traffic at layer 2 as opposed to layer 3. [47]

7.1.5 Data Encryption throughput Performance test: Layer 2 vs. Layer 3 Encryption in High Speed Networks

- The layer 2 encryption was conducted using the SafeNet SafeEnterprise SONET Encryptor (SSE) on an OC-48 line. The SafeNet SSE encrypts the entire payload of the SONET frame, including the encapsulated IP traffic therein, without adding any overhead.
- The SafeNet SSE was compared to the performance of a Cisco Catalyst 6509, encrypting with IPSec and running packet over SONET on an OC-48 line, equipped with a VPN accelerator card.
- Both the layer 2 and layer 3 topologies were subjected to the same array of tests, including throughput, latency, and frame loss.
- Each test was run at various frame sizes, ranging from 64 to 1420 bytes.
- In the throughput testing, the layer 2 SafeNet SSE was able to encrypt at line speed, achieving the same throughput as an unencrypted baseline SONET link (Figure 1).
- The Cisco 6509 with VPN accelerator achieved 5-25% of this throughput at small frame sizes (64-256 bytes) and maximized throughput in larger frame sizes (1024-1420 bytes) at about 40% of the baseline throughput.
- It was later determined that overhead was not the only factor influencing the significantly reduced throughput of the Cisco VPN Accelerator. The VPN accelerator card had virtual interfaces operating at **1.9 Gigabits/s**.
- While a SONET OC-48 line carries a maximum of 4.8 Gigabits/s. Thus the VPN card used in testing was only capable of encrypting about 40% of the line. Two more teamed VPN accelerators would have been needed to encrypt the entire OC-48 line. [47]
- The theoretical maximum throughput of packet over SONET with IPSec was noticeably less than that of the layer 2 encrypted throughputs (Figure 1).

- If the entire OC-48 line had been encrypted using IPSec, adding the average 50-60 bytes of overhead seen in testing, the throughput would have been about 20% less for smaller frame sizes (64-256 bytes) and 5-10% less for larger frame sizes (512-1420 bytes).
- At the time of testing, equipment was unavailable to run these tests in the lab environment. They could not find a converged IPSec device capable of encrypting an entire OC-48 line without using multiple encryption blades in the Cisco 6509s. This would have added significantly to network cost and configuration complexity.[47]
- The SSE SONET encryptors were able to encrypt the full baseline throughput.
- The Cisco 6509s in the baseline configuration were unable to reach 100% throughput because of the additional processing and encapsulation associated with a high speed packet over SONET link.
- If this baseline limitation were removed, the SSE should be able to encrypt 100% of a SONET link at all frame sizes because of the synchronous method of operation and the lack of encryption overhead. The SSE uses the overhead bits in SONET to place encryption information, thus leaving 100% of the data stream available. Even with the bandwidth limited by the baseline topology, the SSE was still able to encrypt more traffic than a device using IPSec theoretically ever could. An IPSec device, because of the additional overhead, could never reach the performance level of the SSE.[47]

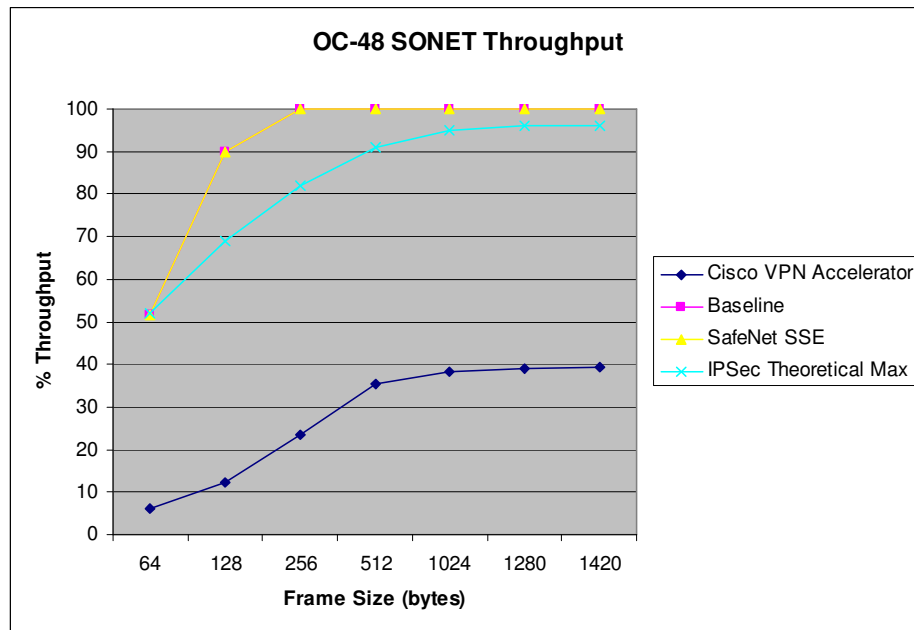


Figure 10: Data Encryption throughput Performance test: Layer 2 vs. Layer 3 Encryption in High Speed Networks

7.1.6 Data Encryption latency Performance test: Layer 2 vs. Layer 3 Encryption in High Speed Networks

- In latency testing, the layer 2 encryption outperformed the layer 3 encryption by a significant amount (Figure 2).

- Layer 2 encryption caused approximately 1-2 μ s in latency in addition to the baseline (less than a 1% increase) at all frame sizes.
- Layer 3 encryption caused a 40% increase in latency at smaller frame sizes, and up to a 60% increase in latency at larger frame sizes, as compared to the unencrypted baseline.

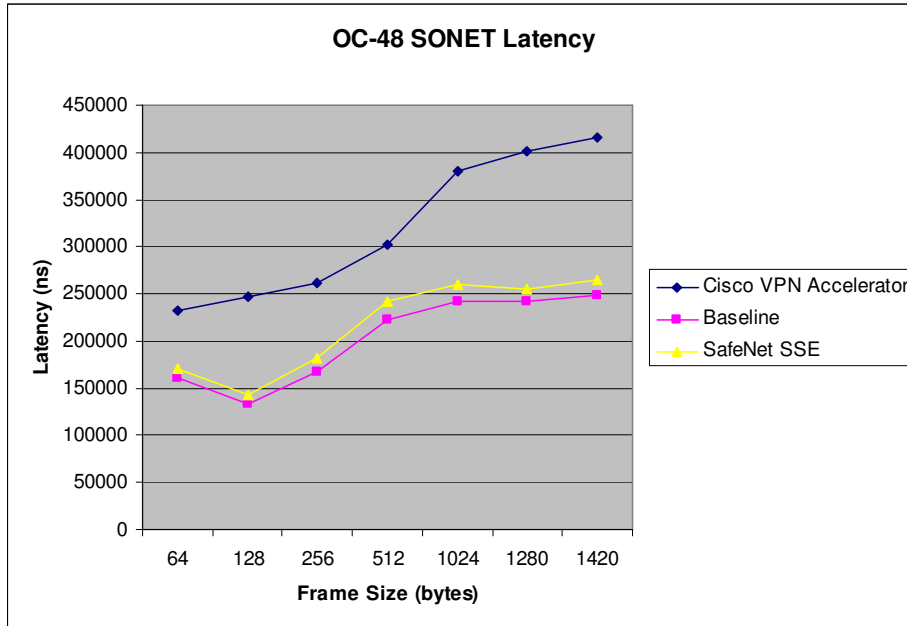


Figure 11: Data Encryption throughput Performance test: Layer 2 vs. Layer 3 Encryption in High Speed Networks

In the case of point-to-point high speed networks, layer 2 SONET encryption generates much better performance in comparison to layer 3 IPsec encryption on the same link. The encryption of traffic at line speed, addition of constant minimal latency regardless of frame size, and minimal frame loss make layer 2 encryption a highly desirable solution. Enterprises that need to secure a point-to-point link are likely to achieve better encryption performance by shifting the traditional encryption with IPsec at layer 3 to the overhead-free encryption of frame payloads at layer 2. [47]

7.1.7 Data Encryption frame loss Performance test: Layer 2 vs. Layer 3 Encryption in High Speed Networks

The frame loss tests showed better performance at layer 2 as well.

- The SafeNet SSE showed frame loss greater than 10% only at 60% throughput of 64-byte frames. This frame loss can be attributed to the baseline topology and not the SafeNet SSE.
- Adding the SSE to the baseline topology showed no additional frame loss (Figure 3).
- The Cisco 6509 with VPN accelerator dropped greater than 10% of frames at 64-256 byte frame sizes for throughput of 25% and greater.

- Adding layer 3 IPsec encryption introduced a significant amount of frame loss in this case (Figure 4).

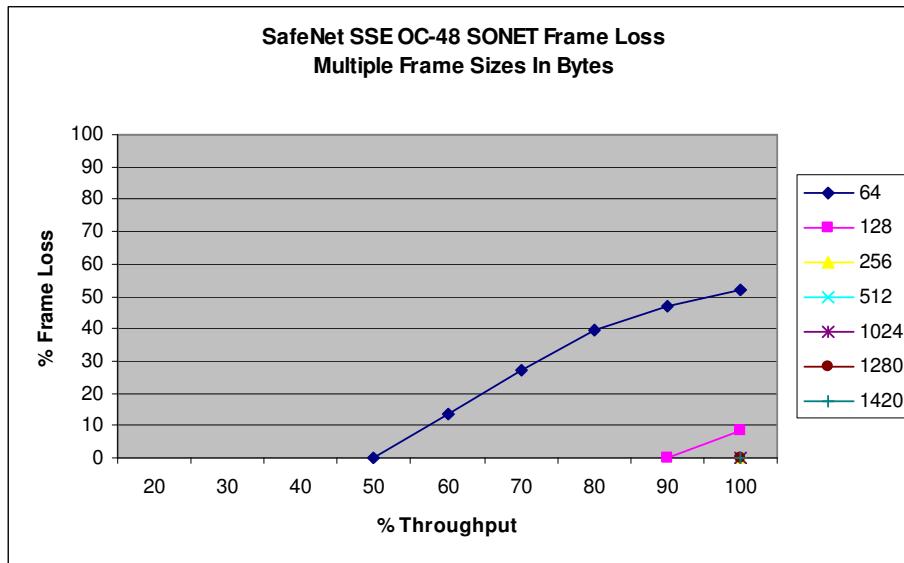


Figure 12: Data Encryption frame loss Performance test: Layer 2 Encryption in High Speed Networks

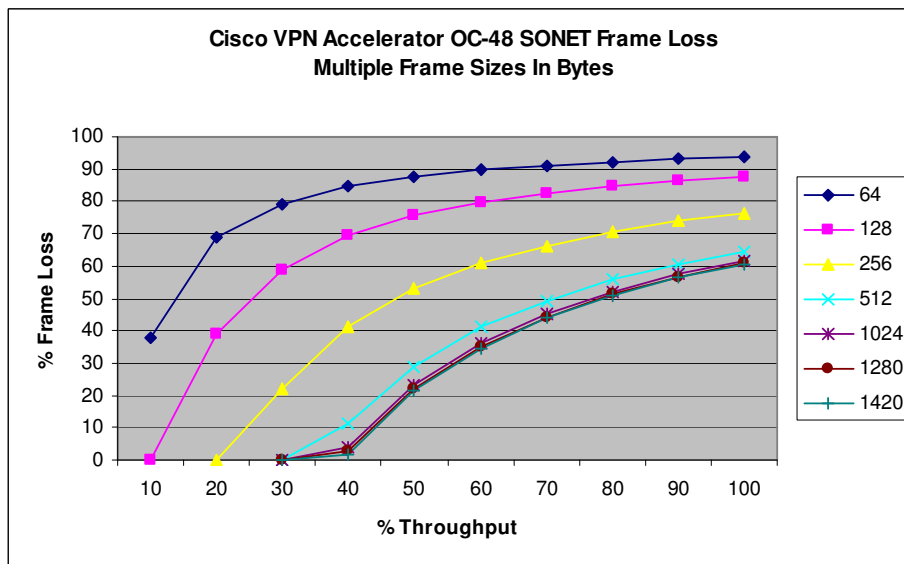


Figure 13: Data Encryption frame loss Performance test: Layer 3 Encryption in High Speed Networks

7.1.8 Data Encryption and Performance test conclusion: Layer 2 and Layer 3 Encryption in High Speed Networks

In the case of point-to-point high speed networks, layer 2 SONET encryption generates much better performance in comparison to layer 3 IPsec encryption on the same link. The encryption of traffic at line speed, addition of constant minimal latency

regardless of frame size, and minimal frame loss make layer 2 encryption a highly desirable solution. Enterprises that need to secure a point-to-point link are likely to achieve better encryption performance by shifting the traditional encryption with IPSec at layer 3 to the overhead-free encryption of frame payloads at layer 2. [47]

This report has been generated from the reference documents and white papers available at SafeNet Inc (www.safenet-inc.com) This was a study by the Rochester Institute of Technology (RIT), it was determined that Layer 2 encryption technologies (in this case, SONEt was tested, but this holds good for Ethernet as well) provide superior throughput and far lower latency than IPSec VPNs, which operate at Layer 3. In part, the RIT study states: “in the case of point-to-point high speed networks, Layer 2 SONEt encryption generates much better performance in comparison to Layer 3 IPSec encryption on the same link. The encryption of traffic at line speed, addition of constant minimal latency regardless of frame size, and minimal frame loss make Layer 2 encryption a highly desirable solution. Enterprises that need to secure a point-to-point link are likely to achieve better encryption performance by shifting from traditional encryption with IPSec at Layer 3 to the overhead-free encryption of frame payloads at Layer 2”. [47] However if Layer 2 encryption is not an option and Layer 3 is the only available solution for encryption then wire-speed IPSec using the CipherOptics SG1002 can be one of the solutions that provides full-duplex, Gigabit speed protection — No encryption bottlenecks; enables up to 1.9 Gbps full-duplex Gigabit Ethernet with AES encryption with virtually no latency. [49-50].

Cisco IPSec VPN services module is a high-speed module for the Cisco Catalyst 6500 Series Switch and the Cisco 7600 Series Internet Router also provides integrated IPSec VPN services. The IPSec VPN Services Module can be used in WAN edge; the VPN module provides VPN termination services on the WAN aggregator router. These can be Integrated with network infrastructure Incorporating VPN into the Catalyst 6500 Switch and 7600 router. Incorporating the latest in encryption hardware acceleration technology, the IPSec VPN module can deliver up to 1.9 Gbps of Triple Data Encryption Standard (3DES) traffic at large packet sizes (500 bytes+) and 1.6 Gbps of 3DES traffic at average packet sizes (upto 300 bytes). [51]

7.1.9 Software implementation of IPSec in fast Ethernet.

This section evaluates the performance of data transmission for large sized data. As an actual application, large MPEG-2 video transmission was selected. For large sized data, when we apply the authentication (AH) and encryption (ESP), the throughput degrades to 1/6 comparing with the throughput without AH or ESP. With AH and ESP, we obtained about 10 Mbps for UDP data transmission and about 6 Mbps for a simple TCP transmission.

The next generation Internet has to achieve the scalable and reliable data transmission. The IPv6 (IP version 6) [55] and IPSec (IP Security) [52] is a core protocol suite for it. IPv6 and has a 128bit address space that is enough to cover all worldwide networks and equipment, and IPSec technology provides a essential functions for reliable and secured data exchange over the Internet. The purpose of this paper is a performance evaluation of data transmission with the IPSec over IPv6 networks. For large sized data, when we apply the authentication (AH) [53] and encryption (ESP) [54], the throughput degrades to 1/9 comparing with the throughput without AH nor ESP. With AH and ESP, we obtains about 10 Mbps for UDP data transmission and about 6 Mbps for TCP

transmission. Also, the throughput was compared with the data transmission with IPv4. The degradation of throughput at the end system, due to the use of IPv6 instead of IPv4, was significantly small. The performance evaluation in this test shows, Even with an ordinary PC platform, we can perform the secure and reliable MPEG-2 video data transmission using the AH and ESP over IPv6 network. The performance degradation or improvement of data transmission throughput due to the introduction of IPv6 is quite significant for end system.

We shall discuss the rough overview of IPSec, and describe the performance evaluation of bulk data transmission over TCP and UDP. We will describe the performance evaluation of data transmission with applying the IPSec.

IPSec (IP Security) RFC 2401 [52] describes the architecture framework of IPSec (IP Security). IPSec protocol suite provides the functional suite for secured and reliable data exchange over the Internet. IPSec has the following two functions, i.e., Authentication and Encryption.

The authentication is a user validation of the communication peer. In order to validate (authenticate) the user, IPSec defines the AH (Authentication Header) in RFC2402 [53]. AH field contains the digital signature calculated by the sender node. The receiver node validates the digital signature in the received AH field. When the AH contains the valid digital signature, the received packet is correctly delivered to the corresponding application module. On the other hand, when the AH does not contains the valid digital signature, the received packet is discarded. With the AH, the receiver can receive the packets only from the authenticated node/user.

Encryption: The encryption is to allow data to be able to be read by only person(s) who has (have) the correct encryption key. In order to encrypt the user data in the IP packet, IPSec defined the ESP (Encapsulating Security Payload) in RFC2403 [54]. ESP field contains the encryption parameters to identify the encryption scheme between the communicating peer nodes. The payload of IP packet contains an ESP header, which has the encryption parameter, and the encrypted user data. In other word, the encrypted user data is encapsulated into the container, whose header is an IP header and an ESP header.

Both for IPv4 and for IPv6, the IPSec is independent from type of data transmission medium. Also, the application does not care whether the IPSec is applied to or not. For IPv6, IPSec is defined as a mandatory option, i.e., every node has to have the IPSec function. We have a concerning with regard to the performance of IPSec. As well known, the required processing for security functions are not light rather would be large. When the execution of security function (i.e., IPSec) requires very large processing power, we could not obtain an enough throughput for large data transfers FTP application. Or, we have to implement the special hardware to handle those security functions such as Data encryptors or decryptors. Performance Evaluation of large data transmission test using end host connected through the two routers. All nodes had a fast Ethernet interfaces.

The end-to-end throughput is evaluated in the following cases.

- ✓ NONE : without IPSec
- ✓ AH1 : only with AH (Authentication Header) using HMAC-SHA1 160bits
- ✓ AH2 : only with AH using KEYED-SHA1 160 bits

- ✓ ESP1 : only with ESP (Encapsulated Security Payload) using 3DES-CBC 192 bits
- ✓ ESP2 : only with ESP using BLOWFISH-CBC 192 bits
- ✓ AH/ESP 1 : both with AH1 and ESP1
- ✓ AH/ESP 2 : both with AH2 and ESP2

This is even when we apply the IPsec technology to provide the secured and reliable data transfer communication over the test bed. The current ordinary PC platform can not handle the large MPEG data with the full IPsec functionality without the 100 % CPU and memory usage. However, the anticipated implementation of fast and high performance data centers is an expected improvement. But of course we expect to have roughly 1/6 network throughput degradation. For large sized data, we obtain about 10 Mbps for UDP data transmission and about 6 Mbps for a simple TCP transmission. The end-to-end throughput was again about 10Mbps.

SECTION 8 DATA INTEGRITY & ENCRYPTION

This section highlights data integrity and data encryption. This section was generated by research material provided in the references with the view point of data transfer mechanisms between the data repositories over large distances for handling high latency networks with different TCP stacks. An overall view of Data handling requirement for SI-8 is presented in this section.

8.1 DATA INTEGRITY

In general, one thinks of hardware and networks as reliable. That is, they work most of the time and when they do fail, they provide an error indication. However, if the data is moved on regular basis, we find over time that silent (undetected) data corruption can and will occur. Over the operational life of any data storage and data transfer operations, we have occurrences of silent data corruption where data has been incorrectly transmitted or stored without any error indication from the hardware or operating system. Problems have ranged from failed storage processors, an improperly/faulty seated card, to software error. Perhaps the most surprising case involved a network router that intermittently corrupted packets in a way that was not detected by the TCP protocol checksum. It is a recommended read for any one involved with transmitting large volumes of data over TCP/IP. The reference [56, 58] gives a good insight on the reliability of the TCP protocol. In order to mitigate this problem, data storage systems must implement an end-to-end capability for verifying the integrity of each granule stored. For example, a checksum should be computed when a granule is created, travel with the granule throughout the system, and be verified when the granule is written to or read from the repository. The modified TCP/IP protocols suite should support data provider's generated file checksums and should verify these checksums on ingest and re-verify the checksums each time a file is read from the repository. In addition, file checksums are included with the metadata distributed along with each granule order so that users have the option of re-verifying file integrity after transfer of data. [57]

8.2 DATA ENCRYPTION

Encryption is a powerful secure technology, should be used in three circumstances: when the repository data is on move: should be enforced when access controls aren't specific enough: and when encryption is mandated.

Many enterprises are turning to encryption to protect their data at rest due to new regulations and industry initiatives. Research has shown that 85 percent of large enterprises will initiate encryption projects by the end of 2006 in response to regulations and industry initiatives. Although encryption tools have advanced in performance and management significantly in the past few years, particularly with dedicated network appliances, encryption is still a difficult and costly proposition.

Research has recommended encryption only in three specific circumstances: Enterprises should encrypt data that moves, for separation of duties when access controls are not granular enough, or when encryption is mandated. Any other use is a waste of resources and won't result in any security benefits. But encryption isn't always the answer, and it is most effective as part of an overall data security strategy. In particular,

it's important to understand how encryption works with (or without) access controls. [59, 60]

8.3 ACCESS CONTROL AND ENCRYPTION

Access controls are the first line of defense in protecting data of any type. Access controls are built into every file system, operating system and many major applications, and they are effective at controlling who can read and manipulate data. Access controls in all modern systems are very secure and not usually prone to direct exploitation. Typically, an attacker has to compromise a user account that has access to the data and, thus gains access without compromising the access control itself. Encryption is more like putting the data in a nearly invincible lockbox. Unless someone has the key, the person can't access whatever's inside the box. But anyone with the key (even if they stole it) can open the box and do whatever they want with what's inside, including pulling it out of the box and sending it around unprotected. Thus, access controls are effective at protecting data in a secure home, while encryption is most effective in protecting data that's subject to movement. Today, the two often work together to resolve some of the problems with distributing keys for encryption. Encryption keys are linked to access control lists and managed by a central encryption application. Instead of giving individual users the encryption keys, the central application stores the keys and decrypts the data based on the now expanded access controls. Encryption becomes transparent to the user by controlling encryption through the access controls. But encryption serves a different purpose than access controls and is not effective when used incorrectly. Hence research at Gartner has envisioned the three laws of encryption. [60, 61]

- Law 1:- Encrypt Data That Moves (Physically or Virtually)
- Law 2:- Encrypt for Separation of Duties When Access Controls Aren't Granular Enough

As effective as access controls are, in many cases, they are not granular enough to protect sensitive data from certain kinds of access. In particular, it's often difficult to protect data from administrators using access controls. The administrators need access to the files or data to carry out their regular job duties, but there are many cases that we don't want them to see what's in the files, or portions of the data. Encryption is also effective to restrict access in shared repositories without having to restructure the access controls for the entire repository. In the following situations, access controls may not be granular enough to enforce appropriate separation of duties, and encryption can improve security:

- ✓ Sensitive database fields
- ✓ Sensitive files in shared environment
- Law 3:- Encrypt When Someone Tells You to, even if it won't improve security

First we would assess the three laws to determine if encryption is appropriate. Data in a secure repository with appropriate access controls and no additional separation of duties requirements does not need encryption. Encryption will not protect data. [60, 61]

Key areas	Approach	Action	Advantages
Centralized Control	Flexible Scheduling	Replication is scheduled	Maximum flexibility for increased data protection
Network Optimized	Variable Bandwidth Throttling	Network-aware bandwidth throttling is set using the actual Speed of your network. Handling high latency networks with different TCP stacks	Full control of limited network bandwidth avoids performance impact
	Data Compression	Compress data to improve network performance	Compression minimizes network traffic and reduces data transfer times
	Byte-level Incremental	Replicate only changed bytes files to minimize network traffic (move less data)	Further reduces network traffic required to move data, keeping the network highly available.
Strong Security	Tiered Security	Combines data encryption with certificate authentication for data transfer. And Data integrity by providing checksum	Ensures high security for transmitting sensitive data over networks
	Certificate Authentication	Strong authentication between hosts utilizing digital certificates	Guarantees data is only sent to authorized parties
	Data Encryption	Selectable encryption up to 128-bit encryption. Or turn off encryption for improved performance over a LAN	Choosing the level of security that's right for the data being transmitted

Table 4: Table of Data Integrity & Encryption handling procedures for SI-8

SECTION 9 NETWORK MODELS AND PERFORMANCE EVALUATION

This is the section describing work accomplished to build model and perform performance evaluation of AARNet and GrangeNet. A summary of the accomplished tasks is given below:

1. Construction of OPNET models covering the AARNet and GrangeNet networks. The models are consistent with all the drawings published on the web sites. This phase of the project revealed several inconsistencies due to drawings published in different stages of the network growth in stages.
2. During the first phase of the project we worked on Grange Net's several drawings documenting individual parts of the network link. The first phase aggregated information from all drawings into one consolidated network model. In this first phase we discovered that there was no readily available and easily accessible documentation about the local LAN segments.
3. Construction of a detailed OPNET simulation model covering the subnet of the DART partners was difficult and lacked detailed network diagrams. This step was difficult and traffic trace and topology was difficult to capture and models were developed by using OPNET to explicitly model the traffic flow dynamics for each application source/sink. Timing sequence diagrams. This requires detailed description of how the various application clients and servers interacted.

The detailed simulation study of the GrangeNet and AARNet revealed the following:

1. The only potential bottlenecks are the Local LAN's. However, high speed links do not seem to exhibit high utilization levels in the presence of local background traffic such as email, sap, web etc.
2. An interesting observation is that routers latencies are negligible (few tens of microseconds). The main sources of delay latencies are the low speed links connecting the local sites to their main backbone gateways. These links could be over-subscribed.
3. On the backbone side the links exhibit very low utilization levels (below 3%). This implies that the links to the services themselves are not bottlenecks.

This project gave us the opportunity to study in detail the structural and logical aspects of the entire GrangeNet and AARNet network, as well as to evaluate the performance of one local LAN subnet for a variety of traffic scenarios. On the basis of this study, we have some recommendation put forth in the following section.

SECTION 10 RECOMMENDATIONS

Recommendation-1

From the comparison study of the two major research networks i.e. AARNet and GrangeNet, it is recommended we use GrangeNet for the following reasons.

- The network has primarily support for educational and research and development organizations.
- It provides high speed dedicated network links between the DART partners.
- Traffic on its network does not have any commercial association.

Recommendation-2

Implement FAST TCP on all repository's data centers.

The congestion control algorithm in the current TCP has performed remarkably well and is generally believed to have prevented severe congestion in the networks for the past two decades. The TCP stack worked well for legacy networks on the LAN side and also on the WAN side for low speed network links such as T1, T3, and E1 E3 etc. It is also well-known, however, that as bandwidth-delay product continues to grow, when we move on to high-speed networks and long distance connectivity the current TCP implementation will eventually become a performance bottleneck. In this work package we recommend an alternative congestion control scheme for TCP, called FAST. FAST TCP has three key differences.

- it is an equation-based algorithm and hence eliminates packet-level oscillations;
- it uses queuing delay as the primary measure of congestion, which can be more reliably measured by end hosts than loss probability in fast long-distance networks; and
- it has a stable flow dynamics and achieves weighted proportional fairness in equilibrium.

The implementation of FAST TCP involves a number of innovations that are crucial to achieve scalability. As the Internet scales up in speed and size, its stability and performance become harder to control. The emerging theory that allows us to understand the equilibrium and stability properties of large networks under end-to-end control forms the basis of FAST TCP implementation.

Recommendation-3

It is recommended we use Hardware Data encryptors / decryptors

- In the case of point-to-point high speed networks, layer 2 encryption generates much better performance in comparison to layer 3 IPsec encryption on the same link. The encryption of traffic at line speed, addition of constant

minimal latency regardless of frame size, and minimal frame loss make layer 2 encryption a highly desirable solution.

- Enterprises that need to secure a point-to-point link are likely to achieve better encryption performance by shifting the traditional encryption with IPSec at layer 3 to the overhead-free encryption of frame payloads at layer 2.
- In case where we do not have the point-to-point links and have to rely on layer-3 connectivity, it is recommended we use IPSec layer-3 data encryption.

Recommendation-4

The recommended transport protocols suite hold data provider's generated file checksums and should verify these checksums on ingest and re-verify the checksums each time a file is read from the repository. In addition, file checksums are included with the metadata distributed along with each granule order so that users have the option of re-verifying file integrity after transfer of data. Combined with data encryption with certificate authentication for data transfer, data integrity by providing checksum and hardware encryptors/decryptors for link protection

Recommendation-5

The work package development identified many areas where significant efficiency gains could be achieved through adoption of current best practices, better application of existing technology, and research into new technological solutions for data centers. These areas are as follows:

- Activities aimed at understanding the data centre market – The size and growth rate of the market as well as local concentrations of data centers will be of interest to planners and implementers in the production environment
- The benefits of obtaining energy utilization benchmarks – By monitoring and comparing the utilization and consumption of a variety of data centers.
- Identify and promote the best practices – Adopting current best practices in existing or new data centers to provide significant improvement in the short term and plan for long term data centre requirements.
- Improving data center facility systems' efficiency - Facility systems containing resources such as processing power, storage, access grid facility, Uninterruptible Power Supplies are far from optimal.
- The requirement analysis of high availability data center for primary and secondary repositories servers should be combined with resource management software, capacity-on-demand services, maximize Data/Application uptime and to meet the performance goals listed in section-6

Recommendation-6

- Deployment of new grid applications and services, while insuring acceptable quality of service, will require replacement of any bottle necks interfacing the high-speed research network with the LAN connecting the Data Centers with thicker bit pipes.

- Another change that will be needed would be to upgrade the LAN segments to run switched Gigabit Ethernet instead of switched Fast Ethernet. These shall reduce latencies experienced by grid applications and services.
- At this time, we expect all secondary repositories are centrally located and accessible across the backbone. The bandwidth of the links leading from the leaves (users) to the main edge routers interfacing with the backbone to access primary repositories needs to be investigated.
- During the execution of this work package, we discovered that no traffic profiles describing typical grid user behaviors were available. Hence, we recommend the generation of such profiles. These profiles are essential in trend analysis and capacity planning studies.

Finally, we would like to stress the fact that the simulation models constructed in this project provides a correct and accurate documentation about the existing GrangeNet and AARNet network. We recommend that they be properly maintained and updated to reflect any changes made to the future network.

SECTION 11 TERMS OF REFERENCE

11.1 GLOSSARY

Acronym	Definition
802.Q	VLAN tagging and trunking
AARNet	Australia's Research and Education Network
AH	IPSec Authentication header
AIMD	Additive Increase Multiplicative Decrease (AIMD) is the dominant algorithm for congestion avoidance and control in the Internet.
ATM	Asynchronous Transfer Mode
CADPC/PTP	Congestion avoidance with distributed proportional control / performance transparency protocol
CERN	CERN is the European Organization for Nuclear Research, the world's largest particle physics centre.
Cisco OSR	Cisco Optical Service Router
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DART	Dataset Acquisition Accessibility & Annotation e-Research Technologies
DEST	Department of Education, Science and Training
DGRA	Data grid reference architecture
DoS	Denial of Service
DRAM	Dynamic random access memory
DWDM	Dense Wavelength Division Multiplexing
ECC	Error correcting code
ESP	IPSec encryption - Encapsulation Security Payload
FAST TCP	Fast AQM (active queue management) scalable TCP
FTP	File transfer protocol
Gbps	Giga bits per sec
GrangeNet	Grid and Next Generation Network
GSI	Grid Security Infrastructure
HS-TCP	High Speed TCP
IDEA	International Data Encryption Algorithm
IETF	Internet Engineering Task Force
IPSec	IPSec (IP security) is a suite of protocols for securing Internet Protocol (IP) communications by authenticating and/or encrypting each IP packet in a data network
iSCSI	Internet SCSI (Small Computer System Interface)
ISP	Internet service provider

iWRAP	Internet Wide Area RDMA Protocol
LFP	Long Fat Pipe
Mbps	Mega bits per sec
MPLS	Multi Protocol Label Switching
Mpps	Mega packets per sec
MSS	maximum segment size
MTU	Maximum Transmission Unit
OC48	Optical Carrier OC-48 = 2.5 Gbps; OC-192 = 10 Gbps
ONS 16454E	OC48 gigabit interconnect module
PKI	Public Key Infrastructure
POP	point-of-presence
QoS	Quality of Service
RAID	redundant array of independent disks
RDMA/TOE	10 GbE RDMA/TOE NIC chip. can achieve 800-MBytes/s with minimal loading on the processor for large data transfers
RIT	Rochester Institute of Technology
RTT	Round - Trip Time
SABUL	SABUL is a protocol for moving data very efficiently over long haul, high performance networks.
SAN	storage area network
SDH	Synchronous Digital Hierarchy
SLA	Service Level Agreement
SMP	Symmetric multiprocessing, or SMP, is a multiprocessor computer architecture where two or more identical processors are connected to a single shared main memory. Most common multiprocessor systems today use an SMP architecture
SONET	Synchronous Optical Network
SSE	Safe Enterprise SONET encryption
SSL/TLS	Secure Socket Layer - Transport Layer Security
STM	Synchronous Transfer Mode
TCP	Transmission Control Protocol (as in TCP/IP)
TCP/AQM	TCP/AQM Active Queue Management and congestion control algorithm
TDM	Time-division multiplexing
UDP	User Datagram Protocol
UPS	uninterruptible power supply
VLAN	Virtual Local Area Networks
VPN	Virtual Private Networks

11.2 REFERENCES

- [1] www.aarnet.com.au

- [2] www.grangenet.com.au
- [3] Postel, J. B. Transmission Control Protocol. RFC 793, September 1981.
- [4] Postel, J. B. User Datagram Protocol. RFC 768 , September 1980.
- [5] Allman, Paxson, et al. TCP Congestion Control, RFC 2581, April 1999.
- [6] Jacobson, Braden, et al. TCP Extensions for High Performance, RFC 1323, May 1992.
- [7] HighSpeed TCP for Large Congestion Windows, Sally Floyd, Internet draft draft-floyd-tcp-highspeed-02.txt, Work in progress, February 2003.
- [8] Iren, S. and Amer, P. The transport layer: tutorial and survey. ACM Computing survey, vol. 31, N°4, December 1999.
- [9] E. He, J. Leigh, O. Yu, T.A. DeFanti, "Reliable Blast UDP: Predictable High Performance Bulk Data Transfer," Proceedings of IEEE Cluster Computing 2002.
- [10] E. He, J. Alimohideen, J. Eliason, N. Krishnaprasad, J. Leigh, O. Yu, T. A. DeFanti, "QUANTA: A Toolkit for High Performance Data Delivery over Photonic Networks," to appear in Future Generation Computer Systems, Elsevier Science Press.
- [11] README file of tsunami-2002-12-02 release. <http://www.indiana.edu/~anml/anmlresearch.html>
- [12] Yuhong Gu, Xinwei Hong, Marco Mazzucco, and Robert L. Grossman, SABUL: A High Performance Data Transport Protocol, 2002, submitted for publication.
- [13] Yuhong Gu, Xinwei Hong, Marco Mazzucco, and Robert L. Grossman, Rate Based Congestion Control over High Bandwidth/Delay Links, 2002, submitted for publication.
- [14] Tom Kelly, "Scalable TCP: Improving Performance in HighSpeed Wide Area Networks," First International Workshop on Protocols for Fast Long Distance Networks, Geneva, February 2003
- [15] FAST TCP: From Theory to Experiments, C. Jin, D. Wei, S. H. Low, G. Buhrmaster, J. Bunn, D. H. Choe, R. L. A. Cottrell, J. C. Doyle, W. Feng, O. Martin, H. Newman, F. Paganini, S. Ravot, S. Singh; submitted to IEEE Communications Magazine, April 1, 2003
- [16] Congestion Control for High Bandwidth-Delay Product Networks, Dina Katabi, Mark Handley and Charlie Rohrs, Proceedings on ACM Sigcomm 2002.
- [17] Welzl, M.: "Traceable Congestion Control", ICQT 2002 (International Workshop on Internet Charging and QoS Technologies), Zürich, Switzerland, 16-18 October 2002. Springer LNCS 2511, available from the PTP website.
- [18] Data Management and Transfer in High-Performance Computational Grid Environments. W. Allcock, J. Bester, J. Bresnahan, A. Chervenak, I. Foster, C.

- Kesselman, S. Meder, V. Nefedova, D. Quesnel, and S. Tuecke. Parallel Computing, 2001.
- [19] A.Chervenak, I.Foster, C.Kesselman, C.Salisbury, and S.Tuecke, The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Data Sets. J. Network and Computer Applications, 2001.
- [20] EU Data Grid Project, "The Data Grid Architecture", DataGrid-12-D12.4-333671-3-0, 2001.
- [21] Grid Datafarm, <http://datafarm.apgrid.org/>.
- [22] I. Foster and C. Kesselman, "A Data Grid Reference Architecture," GriPhyN 2001-6, 2001.
- [23] Sabul, www.dataspaceweb.net/sabul.htm
- [24] FAST TCP, <http://netlab.caltech.edu/FAST/>
- [25] Reliable Blast UDP: Predictable High Performance Bulk Data Transfer, IEEE Cluster Computing, 2002.
- [26] Reliable File Transfer, www-unix.globus.org/toolkit/reliable_transfer.html
- [27] Souza, E. and D. Agarwal, A HighSpeed TCP Study: Characteristics and Deployment Issues, submitted to IEEE Supercomputing 2003.
- [28] Dunigan, T., Mathis, M. and Tierney, B., A TCP Tuning Daemon. IEEE Supercomputing 2002, Baltimore, Maryland, 2002.
- [29] Mandrichenko, I. GridFTP Protocol Improvements. Global Grid Forum, GWD-E-21, 2003.
- [30] Chien, A., Faber, T., Falk, A., Bannister, J., Grossman, R. and Leigh, J. Transport Protocols for High Performance: Whither TCP? Communications of the ACM, 46 (11). 42-49. 2003.
- [31] Clark, D., Lambert, M. and Zhang, L. NETBLT: A Bulk Data Transfer Protocol. IETF, RFC 998, 1987.
- [32] Gu, Y. and Grossman, R.L., UDT: An Application Level Transport Protocol for Grid Computing. Second International Workshop on Protocols for Fast Long-Distance Networks, 2003.
- [33] Jin, C., Wei, D.X. and Low, S.H., FAST TCP: motivation, architecture, algorithms, performance. IEEE Infocom, 2004.
- [34] This research work was supported by the Director, Office of Science. Office of Advanced Scientific Computing Research. Mathematical, Information, and Computational Sciences Division under U.S. Department of Energy Contract No.DE-AC03-76SF00098. See the reports at <http://www-library.lbl.gov/>. This is report no. LBNL-52590.

- [35] W. Allcock, J. Bresnahan, R. Kettimuthu, M. Link, C. Dumitrescu, I. Raicu, I. Foster, *The Globus Striped GridFTP Framework and Server*, SC'05, ACM Press, 2005.
- [36] Robert L. Grossman, Yunhong Gu, Dave Hanley, Xinwei Hong and Parthasarathy Krishnaswamy, *Experimental Studies of Data Transport and Data Access of Earth Science Data over Networks with High Bandwidth Delay Products*, Computer Networks, 2004.
- [37] David Groep, "GSI and encryption in the Fabric" 52-1998, DataGrid-04-NOT-NIKHEF, March 27, 2002
- [38] ANSI, "Triple Data Encryption Algorithm Modes of Operation", American National Standards Institute X9.52-1998, American Bankers Association, Washington DC, July 29, 1998
- [39] FIPS, "Data Encryption Standard", Federal Information Processing Standards Publication 46-3, October 1999 available at: [http://csrc.nist.gov/ encryption/tkencryption.html](http://csrc.nist.gov/encryption/tkencryption.html)
- [40] Sandia National Labs press release, available at <http://www.sandia.Gov/media/NewsRel/NR1999/encrypt.htm>
- [41] C. Wilcox, et. al. "A DES ASIC suitable for network encryption at 10Gbps and beyond." First International Workshop on Cryptographic Hardware and Embedded Systems, Springer Verlag, 1999.
- [42] FreeIP, <http://www.free-ip.com/DES/index.html>
- [43] Patterson, C., "High Performance DES Encryption in Virtex FPGAs using Jbits", FCCM 2000, IEEE Computer Society, 2000.
- [44] S. Trimberger, et. al. "A 12Gbps DES Encryptor/Decryptor Core in an FPGA." Cryptographic Hardware and Embedded Systems - CHES 2000, Springer Verlag, 2000.
- [45] Schneier, B., *Applied Cryptography*, John Wiley and Sons, 1996.
- [46] NIST Test requirements at: <http://csrc.nist.gov/cryptval/des.htm>.
- [47] www.safenet-inc.com
- [48] Tony Rosati, "A High Speed Data Encryption Processor for Public Key Cryptography" 2004
- [49] <http://www.cipheroptics.com/products/products.html>
- [50] http://www.cipheroptics.com/pdf/SG_datasheet_blue_v3.pdf
- [51] http://www.cisco.com/en/US/products/hw/modules/ps2706/products_data_sheet09186a00800c4fe2.html

- [52] S.Kent, R.Atkinson, "Security Architecture for the Internet Protocol", IETF RFC2401, November 1998.
- [52] S.Kent, R.Atkinson, "IP Authentication Header", IETF RFC2402, November 1998.
- [54] S.Kent, R.Atkinson, "IP Encapsulation Security Payload (ESP)", IETF RFC2403, November 1998.
- [55] S.Deering, R.Hinden, "Internet Protocol version 6 Specification", IETF RFC2460, November 1998.
- [56] Paxson, V. End-to-End Internet Packet Dynamics. IEEE Transactions on Networking 7, 3 (June 1999), 277-292.
- [57] Stone, J., Partridge, C. When the CRC and TCP Checksum Disagree. Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (August 2003), 309-319.
- [58] Caulk, Parris M., "The Design of a Petabyte Archive and Distribution System for the NASA ECS Project," Fourth NASA Goddard Conference on Mass Storage Systems and Technologies, (March 28 - 30, 1995), 7-17.
- [59] "Organizations Must Employ Effective Data Security Strategies" Gartner Research Publication Date: 30 August 2005
- [60] "Data Security Enters the Spotlight" Gartner Research Publication Date: 24 October 2005
- [61] "Snap Enterprise Data Replicator Express". Adaptec Storage Datasheet-2006.

SECTION 12 REPORT SIGNOFF

It is agreed between

Lead Investigator: [Abdul Malik Khan](#)

and

Chief Investigator: [Dr. Asad Khan](#)

and

DART Project Director

That the **Final Report Document** for the [DART SI8 Pilot long-distance high speed and secure data transfer between repositories](#) gives a full account of the work undertaken for the DART Project.

[Abdul Malik Khan](#)

abdul.malik.khan@gmail.com

- has been read and reviewed by all parties,
- shows that the [work package SI 8](#) has been completed satisfactorily,
- Clearly outlines the [functionality that was delivered](#).

Dated this [ddth](#) day of [mmmm](#) 20yy

Signed by [name of CI](#) for and on
behalf of the Chief Investigator

Signed for and on behalf of DART by the
Project Director [Andrew Treloar](#)

APPENDIX

SECTION 13 AARNET & GRANGENET COMPARISON

13.1 OVERVIEW

Tables 5 through 14 provide an overview and comparison of AARNet and GrangeNet

Table 5: Infrastructure

	AARNet	GrangeNet
Global Network	<p>The AARNet Fiber consists of undersea and terrestrial fiber-optic cables, including indefeasible right of use (IRU) connecting several countries across North America, Latin American, Western Europe and South Asia, and east Asian countries. Additionally, it has installed network switching and routing equipment on leased fiber-optic facilities in some countries. This Core Network has several regional points of presence (POPs), and many regional POPs are located on the Fiber Network and other regional POPs are located on the leased facilities. It has deployed Multiprotocol Label Switching (MPLS) as a transport technology on its dense wave division multiplexing (DWDM)-based Internet Protocol (IP) backbone network. SX TransPORT will provide dual 10 gigabit per second capacity circuits connecting Australia's Academic and Research Network (AARNet) to the advanced Research and Educations Networks in North America, as part of a bundle of services, for approved non-commercial scientific, research and educational use.</p>	<p>GrangeNet uses the AARNet for Global reach. But its national network presence is strong. It consists of GrangeNet Lightpath - Layer 1 (Physical Layer), Layer 2 (Data link layer), Layer 3 (Network Layer). GrangeNet Lightpath is a Layer 1 clear channel service that transports data between member endpoints. Available in either a 1Gbit/sec or 50 - 200 Mbit/sec variants. The implementation of the Lightpath service is at a TDM (Time Division Multiplexing) level direct into the GrangeNet DWDM backbone. GrangeNet LAN: An extension to the current offering of MPLS is GrangeNet LAN - a Layer 2 service. For customers that have deployed a switch/router (such as a Cisco Catalyst 6500) at their border the LAN service can easily be extended deep into their own networks. For customers that have a traditional router (such as a Cisco 7304) then the LAN offering permits two member routers to directly peer (regardless of the number of hops within GrangeNet). GrangeNet offers MPLS service. GrangeNet R&E is the traditional layer 3 services that permit members to peer with both local and international Research and Education members. Finally a member's use of GrangeNet R&E doesn't preclude the</p>

		use of GrangeNet LAN down the same physical connection. These are both complimentary service offerings and are fully compatible on the same physical GrangeNet POP connection.
Regional Network	<p>A single fibre pair across every path of the Nextgen network will be used to connect regional University campuses, other educational and research facilities (for example, radio telescopes forming the Australian National Telescope Facility) to each other and the AARNet3 GigaPOPs, providing regional reticulation of AARNet services. This regional aspect of the AARNet3 network will provide educationalists and researchers outside capital cities with the same excellent access to national and international education and research faculties as is available to metropolitan institutions.</p>	<p>The GrangeNet Network Operations Center operates and maintains the core and regional infrastructure and facilitates the connection to clients. The network itself is composed of Cisco GSR 12410 and OSR 7609 routers connected using Cisco ONS 15801 DWDM equipment. Typical connections are at Gigabit speed however other connection types can be accommodated. The Physical Network consists of the Cisco OSR 7609s are use as edge routers and are located in each of Melbourne, Canberra, Sydney and Brisbane. The core routers are Cisco GSR 12410 platforms. Between the edge and core 2.5Gbit Packet-Over-SONET POS links are aggregated. This aggregation results in a capacity of 10Gbps to both Melbourne and Sydney edge routers while Canberra and Brisbane aggregate to 5Gbps.</p> <p>On each of the long haul links (between Sydney and Brisbane; Melbourne and Sydney/Canberra; and Sydney and Canberra) Cisco ONS 15801 DWDM equipment is used. The long haul equipment is maintained by GrangeNet partner PowerTel.</p>
Physical Network type	<p>Ability to run multiple active networks including a full service ATM network, an IP-only gigabit Ethernet / Packet-over-Sonet(POS) network, experimental networks and high capacity point-to-point research links</p>	<p>GrangeNet Lightpath is a Layer 1 clear channel service that transports data between member endpoints. Available in either a 1Gbit/sec or 50 - 200 Mbit/sec variant.</p> <p>The implementation of the</p>

		<p>Lightpath service is at a TDM (Time Division Multiplexing) level direct into the GrangeNet DWDM backbone. Data carried in this service does not travel through any of the GrangeNet routers.</p>									
<p>Packet-over-SONET (POS) is a standardized way for mapping IP packets across a provider's network backbone, into SONET (Synchronous Optical Network) / SDH(Synchronous Digital Hierarchy) frames. Cisco Systems helped pioneer PoS technology and has been at the forefront in delivering high-performance and cost-effective PoS solutions for use in service provider and enterprise networks.</p> <p>IMPORTANT: POS is not generally used by customer access lines because they are rarely on a SONET ring. It is primarily backbone architecture.</p> <p>Without POS, backbone routers or switches also transmit IP packets over SONET rings, but they have to first encapsulate them within Frame Relay frames or ATM cells, and then encapsulate the frames/cells into SONET frames. But Frame Relay has no QoS, and ATM has a large amount of overhead (the dreaded "cell tax"). PoS overhead, which averages about 3 percent, is significantly lower than the 15 percent average for the asynchronous transfer mode (ATM) cell tax.</p> <p>POS is for "pure IP" backbones, and it allows encapsulation with either PPP, HDLC, or both (IP encapsulated into PPP, which is encapsulated into HDLC, which is encapsulated into SONET).</p> <p>IMPORTANT: POS does not directly encapsulate IP packets into SONET frames. As a comparison, with IP encapsulation into FR or ATM, the packets are encapsulated into the frames or cells, which are then encapsulated into SONET. POS also encapsulates the packets. The only difference is that it uses PPP, and/or HDLC frames for the encapsulation. The advantage is simplicity, and lower overhead.</p> <p>SONET Synchronous Optical Network is an ANSI standard used in the United States. SDH Synchronous Digital Hierarchy is an ITU standard used in Europe. Both SONET and SDH are very similar, and both transmit data over fiber optic cables. Both standards are point-to-point synchronous networks that use TDM multiplexing across a ring or mesh physical topology. The ATM standard and the FDDI use SONET as the Physical layer.</p>											
<p>Point-of-Presence (POP)</p>	<table border="1"> <thead> <tr> <th colspan="3" data-bbox="464 1642 1154 1707">AARNet3 Points of Presence</th> </tr> <tr> <th data-bbox="464 1707 654 1745"></th> <th data-bbox="654 1707 906 1745">POP "A"</th> <th data-bbox="906 1707 1154 1745">POP "B"</th> </tr> </thead> <tbody> <tr> <td data-bbox="464 1745 654 1885">Adelaide</td> <td data-bbox="654 1745 906 1885">Nextgen Networks, Adelaide</td> <td data-bbox="906 1745 1154 1885">Pulteney Street Data Centre, The University of Adelaide</td> </tr> </tbody> </table>		AARNet3 Points of Presence				POP "A"	POP "B"	Adelaide	Nextgen Networks, Adelaide	Pulteney Street Data Centre, The University of Adelaide
AARNet3 Points of Presence											
	POP "A"	POP "B"									
Adelaide	Nextgen Networks, Adelaide	Pulteney Street Data Centre, The University of Adelaide									

Brisbane	Prentice Centre, University of Queensland	Queensland University of Technology, Gardens Point	Point-of-Presence POP's in Brisbane, Sydney, Canberra, Melbourne and currently deploying a link to Perth. The access to international POP's is via Sydney POP through the AARNet.
Canberra	TransACT	Leonard Huxley Building, Australian National University	
Darwin	Charles Darwin University	-	
Fiji	Fintel Cable Station, Suva	-	
Hawai'i	University of Hawai'i, Manoa	-	
Hobart	University of Tasmania, Sandy Bay	-	
Los Angeles	Telehouse America	-	
Melbourne	Nextgen Networks, Melbourne	Melbourne Law School	
Palo Alto	Switch and Data (PAIX)	-	
Perth	Amnet IX, St Georges Terrace	CSIRO ARRC	
Seattle	Pacific Wave, Westin Building	-	
Sydney	University of Technology, Sydney	Nextgen Networks, Sydney	

Table 6: Network Architecture

	AARNet	GrangeNet
Network Architecture	The current generation of the AARNet network, AARNet3 provides high speed access across the country based on STM-64c (10Gbps) circuits AARNet provides dual links from Brisbane to Perth all along through Sydney, Canberra, Melbourne, and Adelaide to Perth. Provides dual Points of Presence in each Australian capital city along its path. The dual 10Gbps links connect Brisbane-Sydney-Canberra-Melbourne. While one 10Gbps & one 622 Mbps link support the Melbourne-Adelaide-Perth path.	The network when first commissioned had the life time until 2004 and was further extended and funded till 2006. The hardware and software were upgraded to support layer 1 –Light paths, Layer-2 VLANS and Layer-3 Routing with support for Unicast and Multicast on both IPv4 and IPv6. The original GrangeNet architecture consisted of routed point-to-point links. With advances in both router and switch architecture the upgraded network architecture is now possible to use the existing 2.5G backbone links and provide layer 1,

		layer 2 and layer 3 services. The routing topology (L2/L3) now supports 4 Gigabit/sec port channel trunks. Cisco 7609 handles all layer 3 (IP routing), layer 2 (VLAN) switching and Cisco ONS 15454E / ONS 16454E handles all layer 1 (TDM) switching.
VoIP Gateways	VoIP gateways configured are located at all regional POP's.	No VoIP Gateway.
Local Networks	Provides support for LAN connectivity	Provides support for LAN connectivity

Table 7: Transport and Networking

Transport Services	AARNet	GrangeNet
Ethernet	Ethernet IP service: Scalable point-to-point Fast Ethernet or Gigabit Ethernet connections delivered via Global Crossing's IP MPLS backbone. Ethernet metro access: All metropolitan locations support Ethernet transport. Ethernet private line: Dedicated and secure point-to-point Fast Ethernet or Gigabit Ethernet.	Ethernet IP service: Scalable point-to-point Fast Ethernet or Gigabit Ethernet connections delivered via Global Crossing's IP MPLS backbone. Ethernet metro access: All metropolitan locations support Ethernet transport. Ethernet private line: Dedicated and secure point-to-point Fast Ethernet or Gigabit Ethernet.
Wavelength Services	Unprotected, bidirectional, point-to-point circuits at variable speeds. Example 1.0Gbps to 10 Gbps	Unprotected, bidirectional, point-to-point circuits at variable speeds.

Table 8: Internet Protocol

Internet Protocol Services	AARNet	GrangeNet
Internet Access	Dedicated Internet access — Speeds range from 64 Kbps to 10 Gbps, including Fast Ethernet and Gigabit Ethernet, including managed routers and managed security for dedicated Internet access are also available.	Dedicated Internet access — Speeds range from 64 Kbps to 2.5 Gbps, including Fast Ethernet and Gigabit Ethernet, including managed routers and managed security for dedicated Internet access are also available.

IP Video	<p>The AARNet Video over IP service is an extension of the VoIP Service in that it is based on ITU-T H.323 technology. IETF SIP compliant equipment will be supported in the future. IP video provided over AARNet provides Video conferencing — Calls are delivered over customer's Integrated Services Digital Network (ISDN) connection; international calls originated on ISDN are routed over IP network.</p>	Does not support
-----------------	--	------------------

Table 9: Internet Protocol Service

Internet Protocol Services	AARNet	GrangeNet
IP	<p>It supports basic transport of IP packets with support for large MTU size (Which helps for high data transfer), Supports Unicast, Multicast, IPv4, IPv6, Netflow used for IP accounting and flow analysis (network attacks, protocol usage, etc.), low cost connection to AARNet member users, supports QoS (diffserv), VoIP, Video over IP. AARNet3 supports quality of service for all application to obtain the network service it requires for successful operation. AARNet supports a wide range of quality of service options. QoS provides cost savings, DoS attacks don't affect links with QoS, prevents scavenger service to hog the bandwidth. There are a few quality of service software architectures. AARNet, like most networks, using the IETF's Differentiated Services architecture. QoS will be a key issue for providing high data throughput for data transfer between the repositories, preventing high link utilization. Precedence and type of service bits, Differentiated services code point and</p>	<p>IP services are being offered to support traditional Research and Education(R&E) routed protocols. GrangeNet Services are IPv4/IPv6 Unicast, Multicast, R&E access. And support for MPLS, Quality of service (QoS), Grid services, ISCSI attached storage and all Streaming data, Member BGP peers with GrangeNet routers.</p>

	traffic classes' provision will improve our high speed data transfer.	
Layer-2	Supports large MTUs in Linux clusters to maximize throughput. Implementation of MPLS Fast Failover for link protection. Uses MPLS-TE to minimize latency for Voice/Video over IP (real time applications) and Grid Services (distributed computation, visualization ...).	An extension to the current offering of MPLS is GrangeNet LAN - a Layer 2 service are Traffic is carried across the GrangeNet backbone links in dedicated VLAN's, No capacity or quantity constraints on the LAN service. Client can request many LAN services and have these combined with an R&E. 802.1q service is provided to members to select path. A VLAN's can be extended to any POP.
Layer-1	Physical connectivity to the network is user's responsibility. User should make arrangement / connection to AARNet POP.	GrangeNet Lightpath - Layer 1 (Physical Layer) GrangeNet Lightpath is a Layer 1 clear channel service that transports data between member endpoints. Available in either a 1Gbps or 50 – 200 Mbps variant it provides. The implementation of the Lightpath service is with TDM (Time Division Multiplexing) GrangeNet backbone. User Connectivity:- User should make arrangement / connection to two GrangeNet POPs
Converged IP Services	Provisioning of Internet access, IP VPN, VoIP, and IP video services over the same access circuit. Access circuit options of IP VPN, dedicated Internet, or public Internet. Managed solutions including managed routers and managed security. Support for new converged IP applications proposed for future, including Mobile IP Connect, and audio and videoconferencing services.	Provisioning of Internet access, IP VPN, VoIP, and IP video services over the same access circuit. Access circuit options of IP VPN, dedicated Internet, or public Internet. Managed solutions including managed routers and managed security.

Table 10: Grid Services

	AARNet	GrangeNet
<p>Access Grids nodes</p>	<p>Support for Grid services due to rapid growth in deployment of Access Grids nodes and support for next Generation Network/GRID Structures. The future will have a greater focus on location independence for both teaching and learning, and for research. Provide access the resources that will be the key to Health, Bio-informatics, film& Media and education. Some of the examples are haldron, collider, electron microscopes, astronomy, Interactive multimedia on-line teaching and learning.</p>	<p>GrangeNet provide a layer of infrastructure that runs above the network layer, but below the application layer. It sets up a common set of interfaces and standards to seamlessly access network-based resources such as computers, data repositories, collaboration and visualization facilities and online instruments. Examples include the Globus toolkit (computing), SRB (data), and the Access Grid (collaboration). In effect it is providing web-like access to a global set of infrastructure. Besides the resource-specific services, such as how to provide access to a set of supercomputers across the network, there are also technologies that provide more fundamental services. These are often called 'middleware', and include services such as; Authentication, authorization, accounting, resource discovery, security, and network signaling.</p> <p>Examples include Shibboleth (authentication), PERMIS (authorization), and various portal technologies for resource discovery.</p> <p>And pursuing the objective of fostering the development of Grid and Advanced Communications Services, GrangeNet has provided matching funds to a number of research groups (Participants) to build and demonstrate key grid and advanced communications infrastructure. The current round of Participants is: ac3, APAC, ANU, CSIRO, DSTC - FilmEd, USyd-VisLab, QPSF and VPAC.</p>

Table 11: Managed Services

	AARNet	GrangeNet
Managed Services	<p>Does not Provide Solution design — Presale engineering and customer premises equipment (CPE) design • Service installation — Equipment procurement, provisioning and installation.</p> <p>Provide performance management and reporting; and continuous life cycle management (configuration, management and service upgrades and service evolution planning)</p> <p>Does not provide managed security and consulting services, including disaster recovery and business continuity services, among several others. Application performance management.</p> <p>Does not provide premises-to-premises performance information for live monitoring, as well as historical data for analysis and reporting.</p> <p>Does not provide Proactive alerting capabilities with thresholds.</p>	<p>Does not Provide Solution design — Presale engineering and customer premises equipment (CPE) design • Service installation — Equipment procurement, provisioning and installation.</p> <p>Provide performance management and reporting; and continuous life cycle management (configuration, management and service upgrades and service evolution planning)</p> <p>Does not provide managed security and consulting services, including disaster recovery and business continuity services, among several others. Application performance management.</p> <p>Does not provide premises-to-premises performance information for live monitoring, as well as historical data for analysis and reporting.</p> <p>Does not provide Proactive alerting capabilities with thresholds.</p>
	Managed services are available globally with the converged IP services.	Managed services are available globally with the converged IP services.

Table 12: Customer Network Management / Support

Customer Network Management / Support	AARNet	GrangeNet
Portal	<i>Need's Investigation</i>	<i>Need's Investigation</i>
Network Management Platforms	This suite of standards-based service and network management applications provides point-and-click provisioning of end-to-end connectivity for IP services, as well as fault detection, troubleshooting and recovery. AARNet provides	This suite of standards-based service and network management applications provides point-and-click provisioning of end-to-end connectivity for IP services, as well as fault detection, troubleshooting and recovery. GrangeNet provides proactive

	proactive notification for any network outages.	notification for any network outages.
--	---	---------------------------------------

Table 13: Service-Level Agreements

Service-Level Agreements (SLAs)	AARNet	GrangeNet
Private Line (including Ethernet Private Line)	<p><i>Availability — 99.99 percent POP to POP or 99.9 percent end to end.</i></p> <p><u><i>Installation — POP-to-POP and end-to-end guarantees offering needs investigation.</i></u></p>	<p><i>Availability — 99.99 percent POP to POP or 99.9 percent end to end.</i></p> <p><u><i>Installation — POP-to-POP and end-to-end guarantees offering needs investigation.</i></u></p>
Managed Services	<p>A managed SLA can be in addition or supersede the underlying transport SLA, whichever is most suitable:</p> <p>Site availability: Standard is 99.5 percent enhanced is 99.93 percent premium is 99.97 percent</p> <p><u><i>Need's Investigation</i></u></p>	<p>No service Levels are offered by GrangeNet in its current state.</p>
Dedicated Internet Access	<p>Service availability: 99.9 percent for the port and local loop Network availability — 100 percent Latency (round-trip) — 20 ms</p> <p><u><i>Need's Investigation</i></u></p>	<p>Service availability: 99.9 percent for the port and local loop Network availability — 100 percent Latency (round-trip) — 20 ms</p> <p><u><i>Need's Investigation</i></u></p>
Converged IP Services (IP VPN, VoIP and IP Video)	<p>Jitter — 5 ms or less for premium COS. 15 ms or less for enhanced COS Packet delivery — 99.999 percent for premium 99.99 percent for enhanced 99.9 percent for basic Availability — 99.999 percent all three Latency — Region dependent</p> <p><u><i>Need's Investigation</i></u></p>	<p>Jitter — 5 ms or less for premium COS. 15 ms or less for enhanced COS Packet delivery — 99.999 percent for premium 99.99 percent for enhanced 99.9 percent for basic Availability — 99.999 percent all three Latency — Region dependent</p> <p><u><i>Need's Investigation</i></u></p>

Table 14: Pricing

Service	Pricing Elements	
All Services	<p>General pricing includes a nonrecurring charge for service installation and access circuit installation. ‘Usually there are monthly recurring charges for access circuits, with one-, two- and three-year term options.</p> <p><i><u>Need’s Investigation on pricing</u></i></p>	<p>GrangeNet encourage the development of collaborative research, grid services and to deliver the promise of practically unlimited bandwidth at low cost, GrangeNet does not charge for traffic between connected sites. Physical connection: the Client is responsible for all costs involved in the provision of network tails between their premises and the GrangeNet PoP, interface hardware required to connect to the GrangeNet PoP and fees due to other carriers and suppliers.</p>

13.2 COMPETITORS

AARNet and GrangeNet are non profit organizations catering to the IT needs of the Research and Educational Institutions. It does not have any competitors and both organizations share common objectives and goals.

13.3 STRENGTHS

AARNet has maintained strong geographical coverage. The network is again expanding in the region (Australia) with its partnerships and forgoes the expenditure of building or expanding out its own network. This will enable AARNet to attract R&E customers that focuses solely on research activities. While GrangeNet has a strong hold in the South-East Australia to cater to the needs of R&E. The recent upgrade of its network is again targets the Research and Educational institutions.

Major strength lies in both networks support for Grid technologies and architecture.

13.4 LIMITATIONS

There are no limitations from research point of view. The infrastructure of both AARNet and GrangeNet is well suited for the R&E purposes. From commercial point of view, some services and network infrastructure of AARNet is dedicated to commercial use.

13.5 INSIGHT

Require network Infrastructure support to provide more Point-of Presence to Universities and Research Centers.

13.6 REGIONAL HEADQUARTERS

1. AARNet
AARNet Pty Ltd
GPO Box 1559
Canberra ACT 2601 Australia
Tel :-(02) 9963 3538
Email: - noc@aarnet.edu.au

2. GrangeNet
PO Box 1559
Canberra
ACT 2601
Australia
Tel :-(02) 6222 3530
Email: - greg.wickham@grangenet.net

SECTION 14 FAST TCP INSTALLATION

14.1 FAST TCP FOR LINUX 2.6.7 KERNEL

Installation procedure for Linux 2.6.7 kernel is as follow. All the software source files for FAST TCP are available in the work package accompanying CD.

- copy the Linux 2.6.7 kernel to /usr/src directory
- linux-2.6.7.tar.gz downloaded from
`ftp://ftp.kernel.org/pub/linux/kernel/v2.6/linux-2.6.7.tar.gz`
- Change the directory to /usr/src
 - `cd /usr/src`
- Extract the tar file
 - `tar -zxmf linux-2.6.7.tar.gz`
 - `mv linux-2.6.7 linux-2.6.7-fast`
- Change the directory to /usr/src/linux-2.6.7-fast

Clean the directory

- `make clean`
- `make mrproper`
 - Get the working configuration file from the
<good_config_dir> [Optional]
 - `cp ../<good_config_dir>/config .`

- Download the fast headers, Linux integration and fast module tar files to `/usr/src/linux-2.6.7-fast` directory and extract the tar files
 - `tar -zxmf 267fastheaders.tar.gz`
 - `tar -zxmf 267fastmodule.tar.gz`
 - `tar -zxmf 267linuxintegration.tar.gz`
 - Apply the patch using following commands from the `/usr/src/linux-2.6.7-fast` directory
 - `patch -p4 < os.ind.patch`
 - `patch -p4 < os.dep.patch`
 - Run the following commands to compile and install the kernel
 - `make menuconfig`
 - Enable Following options
Select Loadable Module Support --->
[*] Enable Loadable Module Support
[*] Automatic Kernel Module Loading
 - `make`
 - `make modules_install`
 - `make install`
 - Edit the proper lilo/grub config file to add the new kernel entry
- Grub users: Open `/etc/grub.conf` file and edit the line `default=1` to `default=0`. In this way, after booting the machine, the default kernel version should be the 2.6.7-fast version. Also, check that your `/etc/grub.conf` have the following lines:


```
title Red Hat Linux (2.6.7)
root (hd0,0)
kernel /vmlinuz-2.6.7-fast ro root=/dev/hda2
initrd /initrd-2.6.7-fast.img
```
 - Lilo users:edit `/etc/lilo.conf` and add


```
image=/boot/vmlinuz-2.6.7-fast
label=linux-2.6.7-fast
root=/
```

 Choose the proper linux kernel using following command and reboot the system


```
lilo -R linux-2.6.7-fast
reboot
```
 - Steps to Install the FAST module

- `cd /usr/src/linux-2.6.7-fast/fastmodule`
- `make man_install`

14.1.1 Tuning FAST Kernel

- Tuning TCP buffer sizes
 - Run the `tune.sh` (available in CD)
- Tuning Device buffers
 - Tuning `txqueue` length and ring buffer will help in increasing the performance
 - `/sbin/ifconfig eth1 txqueuelen 10000` [name of the interface has to be changed depending on which you are using. Above example is for `eth1` interface]
 - `modprobe e1000 RxDescriptors=4096,4096` [Please read the device specific driver details for setting up the ring buffers. Above example is for Intel e1000 gigabit Ethernet card]
- Setting FAST specific parameters

14.2 FAST TCP FOR LINUX 2.6.15 KERNEL

Installation procedure for Linux 2.6.15 kernel is as follow. All the software source files for FAST TCP are available in the work package accompanying CD.

- copy the Linux 2.6.15 kernel to `/usr/src` directory
- `linux-2.6.15.tar.gz` downloaded from `ftp://ftp.kernel.org/pub/linux/kernel/v2.6/linux-2.6.15.tar.gz`
- Change the directory to `/usr/src`
 - `cd /usr/src`
- Extract the tar file
 - `tar -zxmf linux-2.6.15.tar.gz`
 - `mv linux-2.6.15 linux-2.6.15-fast`
- Change the directory to `/usr/src/linux-2.6.15-fast`
- Clean the directory
 - `make clean`
 - `make mrproper`

- Get the working configuration file from the <good_config_dir> [Optional]
 - `cp ../<good_config_dir>/config .`
 - Download the fast headers, Linux integration and fast module tar files to to /usr/src/linux-2.6.15-fast directory and extract the tar files
 - `tar -zxmf 2615fastheaders.tar.gz`
 - `tar -zxmf 2615fastmodule.tar.gz`
 - `tar -zxmf 2615linuxintegration.tar.gz`
 - Apply the patch using following commands from the /usr/src/linux-2.6.15-fast directory
 - `patch -p1 < os.ind.patch`
 - `patch -p1 < os.dep.patch`
 - Run the following commands to compile and install the kernel
 - `make menuconfig`
 - Enable Following options
Select Loadable Module Support --->
[*] Enable Loadable Module Support
[*] Automatic Kernel Module Loading
 - `make`
 - `make modules_install`
 - `make install`
 - Edit the proper lilo/grub config file to add the new kernel entry
- Grub users: Open /etc/grub.conf file and edit the line default=1 to default=0. In this way, after booting the machine, the default kernel version should be the 2.6.15-fast version. Also, check that your /etc/grub.conf have the following lines:


```
title Red Hat Linux (2.6.15)
root (hd0,0)
kernel /vmlinuz-2.6.15-fast ro root=/dev/hda2
initrd /initrd-2.6.15-fast.img
```
 - Lilo users:edit /etc/lilo.conf and add


```
image=/boot/vmlinuz-2.6.15-fast
label=linux-2.6.15-fast
root=/
```

 Choose the proper linux kernel using following command and reboot the system

```
lilo -R linux-2.6.15-fast
reboot
```

- Steps to Install the FAST module
 - `cd /usr/src/linux-2.6.15-fast/fastmodule`
 - `make man_install`

14.2.1 Tuning FAST Kernel

- Tuning TCP buffer sizes
 - Run the `tune.sh` (available in CD)
- Tuning Device buffers
 - Tuning `txqueue` length and ring buffer will help in increasing the performance
 - `/sbin/ifconfig eth1 txqueuelen 10000` [name of the interface has to be changed depending on which you are using. Above example is for `eth1` interface]
 - `Modprobe 1000 RxDescriptors=4096, 4096` [Please read the device specific driver details for setting up the ring buffers. Above example is for Intel 1000 gigabit Ethernet card]
- Setting FAST specific parameters

14.3 TUNING FAST TCP FOR LINUX KERNEL –SAMPLE CONFIGURATION

```
# This script is for tuning 1Gig Link and intel fast ethernet
```

```
echo "before tuning, /proc/sys/net/ipv4/tcp_mem"
cat /proc/sys/net/ipv4/tcp_mem
echo "33554432 33554432 33554432" > /proc/sys/net/ipv4/tcp_mem

echo "before tuning, /proc/sys/net/ipv4/tcp_rmem"
cat /proc/sys/net/ipv4/tcp_rmem
echo "33554432 33554432 33554432" > /proc/sys/net/ipv4/tcp_rmem

echo "before tuning, /proc/sys/net/ipv4/tcp_wmem"
cat /proc/sys/net/ipv4/tcp_wmem
echo "33554432 33554432 33554432" > /proc/sys/net/ipv4/tcp_wmem

echo "before tuning, /proc/sys/net/core/wmem_max"
cat /proc/sys/net/core/wmem_max
echo "33554432 33554432 33554432" > /proc/sys/net/core/wmem_max

echo "before tuning, /proc/sys/net/core/rmem_max"
```

```

cat /proc/sys/net/core/rmem_max
echo "33554432 33554432 33554432" > /proc/sys/net/core/rmem_max

echo "before tuning, /proc/sys/net/core/wmem_default"
cat /proc/sys/net/core/wmem_default
echo "33554432" > /proc/sys/net/core/wmem_default

echo "before tuning, /proc/sys/net/core/rmem_default"
cat /proc/sys/net/core/rmem_default
echo "33554432" > /proc/sys/net/core/rmem_default

echo "before tuning, /proc/sys/net/core/optmem_max"
cat /proc/sys/net/core/optmem_max
echo "33554432" > /proc/sys/net/core/optmem_max

# Auto tuning of alpha - Disabled
#echo "0" > /proc/sys/net/ipv4/tcp_fast_at_sec

# Set Alpha/Beta/Gamma - 1Gbps Link
echo "200" > /proc/sys/net/ipv4/tcp_fast_alpha
#echo "215" > /proc/sys/net/ipv4/tcp_fast_beta
#echo "200" > /proc/sys/net/ipv4/tcp_fast_gamma

echo "1" > /proc/sys/net/ipv4/tcp_fast
echo "1" > /proc/sys/net/ipv4/tcp_fast_bc

# tcp_fast_kmon_rtt - >0:Enable the fast debug statement.
# you can see the /var/log/messages
# FAST monitor will log various tcp related kernel variable to
# /var/log/messages for the flow with RTT > tcp_fast_kmon_rtt.
# - 0:Disable
#
echo "0" > /proc/sys/net/ipv4/tcp_fast_kmon_rtt
echo "400" > /proc/sys/net/ipv4/tcp_fast_kmon_t

/sbin/ifconfig eth0 txqueuelen 500000
echo "500000" > /proc/sys/net/core/netdev_max_backlog
modprobe e1000 Rx Descriptors=4096,4096
echo "0" > /proc/sys/net/ipv4/tcp_timestamps = 0
echo "0" > /proc/sys/net/ipv4/tcp_sack = 0
sysctl -w net.ipv4.route.flush=1

```