Um Estudo do Desempenho dos Protocolos iSCSI e Fibre Channel

Antônio José Rodrigues Neto

Dissertação de Mestrado

Instituto de Computação

Universidade Estadual de Campinas

Um Estudo do Desempenho dos Protocolos iSCSI e Fibre Channel

Antônio José Rodrigues Neto

Dezembro de 2004

Banca Examinadora

- **Professor Dr. Nelson Luis Saldanha da Fonseca (Orientador)**
  Instituto de Computação – Universidade Estadual de Campinas
- **Professor Dr. Moacyr Martucci Júnior**
  Departamento Engenharia de Computação e Sistemas Digitais
  Escola Politécnica – Universidade de São Paulo
- **Professora Dr[a]. Islene Calciolari Garcia**
  Instituto de Computação – Universidade Estadual de Campinas
- **Professor Dr. Ricardo de Oliveira Anido (Suplente)**
  Instituto de Computação – Universidade Estadual de Campinas

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP

Rodrigues Neto, Antônio José

R618e Um estudo do desempenho dos protocolos iSCSI e Fibre Channel / Antônio José Rodrigues Neto -- Campinas, [S.P. :s.n.], 2004.

Orientador : Nelson Luis Saldanha da Fonseca

Trabalho final (mestrado profissional) - Universidade Estadual de Campinas, Instituto de Computação.

1. Protocolos de redes de computadores. 2. Armazenamento. 3. Desempenho. I. Fonseca, Nelson Luis Saldanha da. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Um Estudo do Desempenho dos Protocolos iSCSI e Fibre Channel

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Antônio José Rodrigues Neto e aprovada pela Banca Examinadora.

Campinas, 20 de dezembro de 2004

Professor Dr. Nelson Luis Saldanha da Fonseca
(Orientador)

Dissertação apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

TERMO DE APROVAÇÃO

Trabalho Final Escrito defendido e aprovado em 20 de dezembro de 2004, pela Banca Examinadora composta pelos Professores Doutores:

Prof. Dr. Moacyr Martucci Junior
POLI - USP

Prof. Dr. Islene Calciolari Garcia
IC - UNICAMP

Prof. Dr. Nelson Luis Saldanha da Fonseca
IC - UNICAMP

A humildade é o último degrau da sabedoria

(Professores Celso Luis Zacari Faria e Dib Karam Júnior)

Dedicatória

Oração a Nossa Senhora Aparecida

Ó Virgem Maria, abençoada sois vós pelo Senhor Deus Altíssimo entre todas as mulheres da terra. Vós sois a glória de Jerusalém, vós a alegria de Israel, vós a honra de nosso povo.

Salve, ó Virgem, honra de nossa terra, a quem rendemos um culto de piedade e veneração, a quem chamamos com o belo nome de Aparecida. Quem poderia contar, ó doce Mãe, quantas graças, durante tantos anos, vós dispensastes ao povo brasileiro, compadecida de nossos males?

Quisemos cingir vossa cabeça sagrada com uma coroa de ouro, que vos é devida por tantos títulos; continuai a dobrar-vos benignamente às nossas preces. Quando erguemos aos céus nossas mãos suplicantes, ouvi clemente os nossos rogos, ó Virgem; conservai nossas almas afastadas da culpa e, por fim, conduzi-nos ao céu.

Salvação, honra e poder Àquele que, uno e trino, nos fulgores de seu trono celeste, governa e rege todo o universo.

Nossa Senhora da Conceição Aparecida, **rogai por nós.**

Oração a Santo Expedito

Ó Deus, que a intercessão de Santo Expedito nos recomende junto à vossa divina bondade, a fim de que, pelo auxílio, possamos obter aquilo que nossos fracos méritos não podem alcançar.

Nós vos pedimos, Senhor, que orienteis, com a vossa graça, todos os nossos pensamentos, palavras e ações, para que possamos, com coragem, fidelidade e prontidão, em tempo próprio e favorável, levar a bom termo todos os nossos compromissos e alcançarmos a feliz conclusão de nossos planos.

Por Nosso Senhor Jesus Cristo. Assim seja.

SÚPLICA: Ó Santo Expedito! Animados pelo conhecimento de que foram prontamente atendidos todos aqueles que vos invocaram à última hora, para negócios urgentes, nós vos suplicamos que nos obtenhais de bondade misericórdiosa de Deus, por intecessão de Maria Imaculada (HOJE ou em tal dia) a graça … que com toda a humildade solicitamos que nos alcanceis junto à bondade toda-poderosa de Deus.

(Pai-Nosso, Ave-Maria, Glória ao Pai)

Agradecimentos

Sobretudo a Deus, Nossa Senhora Aparecida e Santo Expedito que sempre guiaram meus passos e ouviram minhas preces.

Aos meus pais Lilo, Thereza, minha irmã Emília e toda a minha família, que torceram e sofreram junto comigo até o fim de mais essa etapa em minha vida. Amo vocês!

A minha esposa e meu amor Patrícia e a minha filha Anna Júlia que sempre estiveram ao meu lado em todos momentos e sempre souberam compreender minha ausência. Amo muito vocês!

Ao meu orientador Prof. Dr. Nelson Luis Saldanha da Fonseca, primeiramente pela oportunidade, confiança e orientação. Sem sua ajuda esse trabalho não seria possível!

A todas as pessoas, funcionários e professores do Instituto de Computação – UNICAMP por toda a ajuda e paciência durante todos esses anos. A todos os meus amigos, especialmente Robledo de Andrade e Castro, Professor Celso Luiz Zacari Faria, Professor Aziz Kalaf Filho, Eder Cassiola e muitos outros que de alguma forma me auxiliam nessa importante conquista da minha vida.

Ao meu grande amigo e mentor Ariovaldo Almeida (Ari), não tenho palavras para expressar minha gratidão por toda ajuda e principalmente por sua amizade.

Ao meu eterno "pai" Prof. Dib Karam Júnior, que desde o início de minha carreira esteve ao meu lado me orientando e me guiando. Você mora no meu coração!

Que Deus os abençoe por toda ajuda!

Resumo

A maioria das corporações está enfrentando, nos dias de hoje, uma crescente demanda por métodos eficientes de manipulação da informação. Cada vez mais, as Aplicações de Sistema estão se tornando mais sofisticadas e com conteúdo mais rico e amplo. Assim, a necessidade de mais espaço para armazenamento de dados e um acesso mais rápido à informação aumenta. No passado, sistemas de armazenamento de dados e redes de computadores eram vistos como duas entidades independentes, porém, devido às crescentes exigências de armazenamento, o acesso à informação armazenada passou ser ultimamente interrelacionada com as redes de comunicação de dados.

Em conseqüência, surgem as Redes de Armazenamento de Dados (*Storage Area Network*) que são redes que tem como principal finalidade a transferência dos dados entre Aplicações de Sistema e os Dispositivos de Armazenamento. Uma SAN (*Storage Area Network*) consiste em um infraestrutura de comunicação, a qual fornece conexões físicas, uma camada de gerenciamento de todas as conexões, elementos de armazenamento e Aplicações de Sistema, tornando a transferência de dados mais segura e robusta. Uma Rede de Armazenamento de Dados (*Storage Area Network*) pode ser baseada em uma infraestrutura de componentes Fibre Channel e utilizar o protocolo de transporte Fibre Channel Protocol (FCP). Outra alternativa existente é uma Rede de Armazenamento de Dados que utiliza uma infraestrutura composta pela pilha TCP/IP e pelo protocolo de transporte iSCSI.

O presente trabalho analisa o desempenho dos protocolos iSCSI e Fibre Channel em Redes de Armazenamento de Dados. Experimentos foram realizados para a análise de desempenho de cada protocolo em relação as métricas Vazão e Utilização da CPU. Utilizou-se um modelo de carga sintética baseado em requisições homogêneas e heterogêneas. Os resultados obtidos indicam que o protocolo iSCSI produz resultados expressivos com desempenho similar ao desempenho do protocolo Fibre Channel. A adoção e implementação do protocolo iSCSI oferece também vantagens em relação a custo, utilização de uma infraestrutura TCP/IP e a possibilidade de interoperabilidade com dispositivos conectados a Internet.

Abstract

Nowadays, most of the corporations face a growing demand for efficient methods to manipulate information. Applications are becoming more sophisticated with richer content and consequent there is the need for more storage and faster access to information. In the past, storage systems and computer networks were two independent entities, but due to the huge demand of storage, these two need to be interrelated.

Storage Area Network (SAN) are networks which primary purpose is the transfer of data between computer systems and storage devices. SANs consist of a communication infrastructure, which provides physical connections, a management layer, which organizes the connections, storage devices, and computer systems.

Storage Area Network can be based on a Fibre Channel infrastructure using Fibre Channel Protocol (FCP) as a transport protocol. Another possibility is to be based on an IP infrastructure, using TCP/IP stack and iSCSI protocol.

The present work analyzes the performance of iSCSI and Fibre Channel protocols in the Storage Area Networks. Experiments were conducted to analyze the performance of each protocol. Throughput and CPU utilization were the metrics used for comparison. Results indicate that the iSCSI protocol gives expressive results with a performance close to the Fibre Channel Protocol (FCP).

iSCSI also offers other advantages related to cost, availability of TCP/IP infrastructure and the possibility of interoperability with devices connected to the Internet.

Sumário

Capítulo 5

Capítulo 6

Lista de Acrônimos

Referências Bibliográficas

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

No ano de 2002, um estudo [1] produzido pela *School of Information Management and Systems* (SIMS) da Universidade de Berkeley demonstrou que o mundo produziu cerca de 5 exabytes de novas informações, ou seja, cada indivíduo na Terra gerou cerca de 800 MBytes de novas informações. A Tabela 1 demonstra alguns exemplos de quantidade de informação existentes em objetos distintos. Noventa e dois por cento (92%) destes 5 exabytes de novas informações foram armazenados em algum meio magnético, geralmente discos rígidos, o que evidencia um aumento da demanda de capacidade de armazenamento dos sistemas de Entrada e Saída (E/S), bem como uma maior necessidade de métodos eficientes para o armazenamento e recuperação da informação.

Unidade	Qtde	Descrição
Byte **8 bits**	1 10 100	Simples caracter. Simples palavra. Um telegrama ou 1 cartão perfurado.
KiloByte **1000 bytes** **10^3 bytes**	1 10 100 500	Uma pequena história. Uma página de enciclopédia Uma fotografia de baixa resolução Uma “grande” caixa cartões perfurados.
MegaByte **1,000,000 bytes** **10^6 bytes**	1 5 100 500	Um floppy disk Todos os trabalhos de Shakespeare. Um livro Um CDROM
GigaByte **1,000,000,000 bytes** **10^9 bytes**	1 20 100 500	Uma sinfonia com high-fidelity som. Coleção de trabalhos de Beethoven. Um andar de jornais acadêmicos. Ambiente Small Office.
TeraByte **1,000,000,000,000 bytes** **10^{12} bytes**	1 12 10 100	1 ambiente Mid-Range. Biblioteca de academia de pesquisa do USA. Coleção impressa biblioteca congresso do USA. Ambiente High-End.

Unidade	Qtde.	Descrição
Petabyte **1,000,000,000,000,000 bytes** **10^{15} bytes**	1 2 20 200	3 anos de dados do EOS – NASA – 2001. Todas as bibliotecas de US Academic. Produção de Hard-Disks em 1995. Material impresso ou produção de tape digital em 1995.
Exabyte **1,000,000,000,000,000,000 bytes** **10^{18} bytes**	5	Total de informações geradas no ano de 2002.
Zettabyte **1,000,000,000,000,000,000,000 bytes** **10^{21} bytes**		
Yottabyte **1,000,000,000,000,000,000,000,000 bytes** **10^{24} bytes**		

Tabela 1 – Exemplos da Quantidade de Informação

Assim sendo, a maioria das corporações está enfrentando nos dias de hoje, uma crescente demanda por métodos eficientes de manipulação da informação, onde cada vez mais as aplicações de sistema, tais como sistemas de Correio Eletrônico e sistemas de Banco de Dados estão se tornando mais sofisticados e com conteúdo mais rico e amplo. Houve uma mudança na forma de apresentação dessa informação, por exemplo, no passado, um relatório era apresentado em um formato texto simples e nos dias de hoje, o mesmo relatório é apresentando utilizando gráficos demonstrativos, tabelas comparativas e simulações. Com isso, a necessidade de mais espaço para armazenamento de dados e um acesso mais rápido à informação aumenta cada vez mais.

Por outro lado, tanto as redes corporativas quanto a Internet deparam-se com o desafio de proporcionar acesso a um volume crescente de informação armazenada de forma eficiente.

No passado, sistemas de armazenamento de dados e redes de computadores eram vistos como duas entidades independentes, porém, devido às crescentes exigências de armazenamento, o acesso à informação armazenada passou a ser interrelacionada com as redes de comunicação de dados.

O surgimento da tecnologia Ethernet operando a 1 Gbit/s permitiu a possibilidade do uso da pilha de protocolos TCP/IP para a comunicação com sistemas de armazenamento de dados. No entanto, o processamento da pilha TCP/IP implica em um *overhead* substancial quando comparado ao protocolo Fibre Channel, que é largamente adotado, nos dias de hoje, em implementações de Redes de Armazenamento de Dados.

No entanto, existem alguns questionamentos quanto a utilização do protocolo iSCSI em uma infraestrutura de redes de armazenamento de dados em relação ao desempenho obtido pelo protocolo Fibre Channel, dado que o protocolo iSCSI baseia-se em uma infraestrutura TCP/IP.

Como o protocolo iSCSI utiliza da pilha TCP/IP, este está sujeito ao processamento e aos *overheads* introduzidos por esta pilha, bem como ao enquadramento Ethernet (enlace de dados) gerando, assim, uma grande incerteza e insegurança na adoção e utilização dessa pilha quando comparado a uma arquitetura mais consolidada e amplamente utilizada como a arquitetura Fibre Channel.

O objetivo do presente trabalho é analisar o comportamento geral dos protocolos iSCSI e Fibre Channel quanto ao desempenho. Nesta comparação, utiliza-se as métricas Vazão e Utilização da CPU. Experimentos em plataformas submetidas a carga sintética foram realizados. Comparou-se o desempenho do protocolo iSCSI em relação ao protocolo Fibre Channel baseando em perfis de requisições homogêneo e heterogêneo de operações de E/S e utilizando-se perfis de acesso sequencial e randômico.

Os diversos aspectos envolvidos no Estudo do Desempempenho dos Protocolos iSCSI e Fibre Channel (FCP) foram agrupados e discutidos na seguinte sequência:

No Capítulo 2, são apresentados os Sistemas de Armazenamento, introduz-se as arquiteturas Direct Attached Storage (DAS) e Storage Area Network (SAN). No Capítulo 3, são introduzidos o protocolo Small Computer Systems Interface (SCSI), os componentes e protocolo Fibre Channel, bem como o protocolo iSCSI. No Capítulo 4, discorre-se sobre as características da utilização do protocolo iSCSI em relação ao protocolo Fibre Channel. No Capítulo 5, são apresentados os resultados numéricos das análises de desempenho do protocolos iSCSI e Fibre Channel. No Capítulo 6 são apresentadas as Conclusões.

Capítulo 2

Neste capítulo são descritos as características dos Sistemas de Armazenamento existentes.

2.1 Sistemas de Armazenamento

O relacionamento das corporações com os seus clientes está mudando a cada dia, a fim de que os produtos das empresas estejam cada vez mais presentes nos clientes. Corporações utilizam-se dos meios de comunicação, recursos computacionais e em especial do uso da Internet como um meio de divulgação de sua linha de produtos.

Um exemplo típico dessa interação é o relacionamento existente entre um Banco e seus clientes. Alguns anos atrás, todas as transações bancárias eram feitas diretamente na agência bancária, na qual os clientes possuíam uma conta para efetuar transações. Muitos usuários não realizam mais transações bancárias de forma presencial. A maioria das transações bancárias podem ser efetuadas através de serviços de Home Banking, proporcionando aos usuários uma maior comodidade, flexibilidade e principalmente economia de tempo nas operações bancárias.

Surge, então, a necessidade de uma infraestrutura de armazenamento de dados, que proporcione flexibilidade, desempenho, escalabilidade, segurança e alta disponibilidade para as aplicações de sistema, proporcionando, assim, um melhor nível de serviço para os usuários do sistema. [2]

Os principais componentes (Figura 1) de uma infraestrutura de armazenamento de dados são:

- ✓ **Servidor:** provê recursos de hardware e software para as aplicações de sistema. Um exemplo de um modelo de Servidor é um equipamento Intel Pentium III – 1266 MHz com 1GB de RAM;

- ✓ **Aplicação:** formam ambientes de entrada e saída de dados referentes aos processos das corporações. Um exemplo de uma aplicação de sistema é um servidor de Correio Eletrônico;

- ✓ **Meios de Armazenamento:** provêem a infraestrutura para o armazenamento e recuperação de informações provenientes das aplicações de sistema. Os meios de armazenamento podem ser: discos rígidos, discos ópticos, unidades de fita e CD (*Compact Disk*). O Disco Rígido é o meio de armazenamento mais utilizado em um sistema de armazenamento de dados, sendo largamente utilizado em Servidores de Aplicação, Estações de Trabalho e notebooks;

- ✓ **Dispositivo de Armazenamento:** são equipamentos com a funcionalidade específica de armazenamento de dados, características desejáveis de alto desempenho, flexibilidade, escalabilidade e alta disponibilidade para o armazenamento e recuperação das informações;

- ✓ **Meio de Transporte:** provê conexão física entre os Servidores e Dispositivos de Armazenamento;

- ✓ **Protocolo de Comunicação:** estabelece a comunicação lógica entre os Servidores de Aplicação e os Dispositivos de Armazenamento. Um exemplo de um protocolo de comunicação é o protocolo SCSI ou o protocolo Fibre Channel;

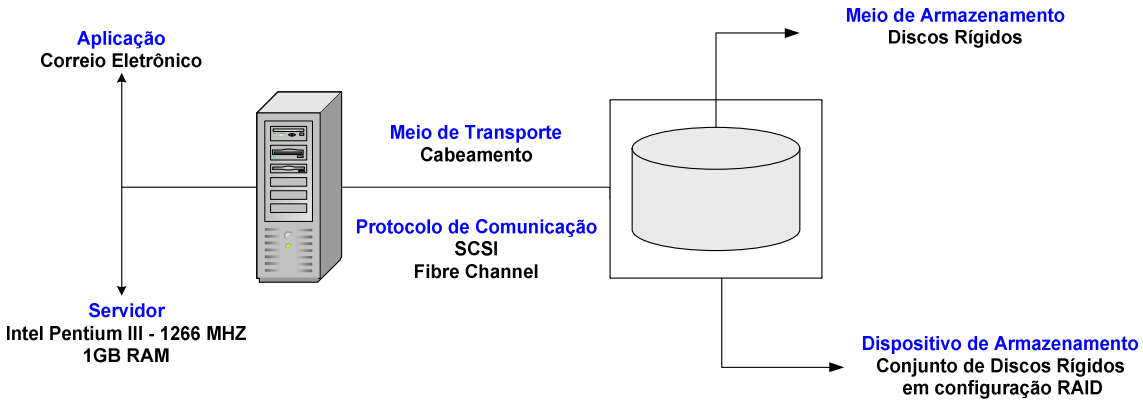


Figura 1 – Componentes Arquitetura de Armazenamento

2.1.1 Arquitetura Direct Attached Storage (DAS)

A arquitetura *Direct Attached Storage* (DAS) é o modelo de interconexão mais simples e comum em sistemas de armazenamento de dados. A grande maioria dos computadores pessoais, estações de trabalho e notebooks utiliza esse modelo. A arquitetura DAS consiste de uma interconexão direta entre o Servidor de Aplicação e o Dispositivo de Armazenamento (Figura 2). O Dispositivo de Armazenamento pode conter um ou múltiplos discos rígidos sem nenhuma configuração de tolerância a falhas, chamada de *Just a Bunch of Disks* (JBOD) ou configurado utilizando um conjunto de discos rígidos com uma configuração tolerante à falhas chamada de *Redundant Arrays of Independent Disks* (RAID). [3] [4]

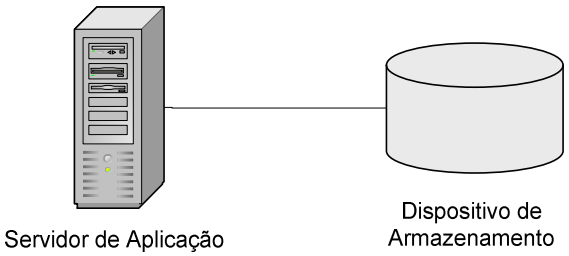


Figura 2 – Arquitetura *Direct Attached Storage* (DAS)

Conforme ilustra a Figura 2, a característica principal da arquitetura DAS é a existência de uma conexão direta entre o Servidor de Aplicação com o Dispositivo de Armazenamento. Uma possível falha em um componente do Servidor de Aplicação ou no meio de transporte põe em risco a disponibilidade da informação para os usuários do sistema, pois existe tipicamente uma única conexão não tolerante à falhas entre as entidades Servidor de Aplicação e o Dispositivo de Armazenamento.

A arquitetura DAS apresenta desvantagens, tais como falta de flexibilidade de compartilhamento de recursos e alto custo de gerenciamento e administração das informações armazenadas e distribuídas. Nestas, quando existe a necessidade de aumento de área para armazenamento, adquire-se novos servidores ou Dispositivos de Armazenamento diretamente conectados ou faz-se a substituição do Dispositivo de Armazenamento implicando em custos adicionais, bem como em uma escalabilidade limitada para as corporações.

A implementação da arquitetura DAS baseia-se na utilização de diversos Servidores de Aplicação com Dispositivos de Armazenamento diretamente conectados, portanto um modelo de utilização de **1-1**, onde para cada Servidor de Aplicação têm-se um Dispositivo de Armazenamento correspondente, será uma grande quantidade de informações não consolidadas e distribuídas. Nesta arquitetura, as tarefas de gerenciamento, administração, *backup* e recuperação tornam-se extremamente complexas e onerosas.

Apesar da arquitetura DAS representar 42% do mercado de armazenamento de dados [5], essa arquitetura tende a não ser mais adotada em virtude do custo operacional, método simples de interconexão e sua inflexibilidade no gerenciamento e no manuseio das informações.

2.1.2 Arquitetura Storage Area Network (SAN)

Nos computadores pessoais, estações de trabalho e notebooks, a arquitetura DAS é a melhor opção de utilização dado a sua simplicidade de interconexão para com o Dispositivo de Armazenamento. Entretanto, nos sistemas corporativos, flexibilidade, escalabilidade, disponibilidade e tolerância à falhas do Dispositivo de Armazenamento, são indispensáveis para a continuidade do negócio e a sobrevivência da corporação. Para atender estas necessidades, surgiu então a arquitetura *Storage Area Network* (SAN).

A arquitetura SAN (Figura 3) é uma infraestrutura de rede dedicada ao compartilhamento de Dispositivos de Armazenamento para Servidores de Aplicação, proporcionando flexibilidade, disponibilidade e escalabilidade para os sistemas corporativos, bem como o armazenamento de todas as informações dos Servidores de Aplicação em um único ponto de armazenamento de dados. Baseia-se em um modelo de armazenamento de ***N-1***, onde para cada ***N*** Servidores de Aplicação têm-se um Dispositivo de Armazenamento, o qual pode estar localizado fisicamente junto aos Servidores de Aplicação ou não.

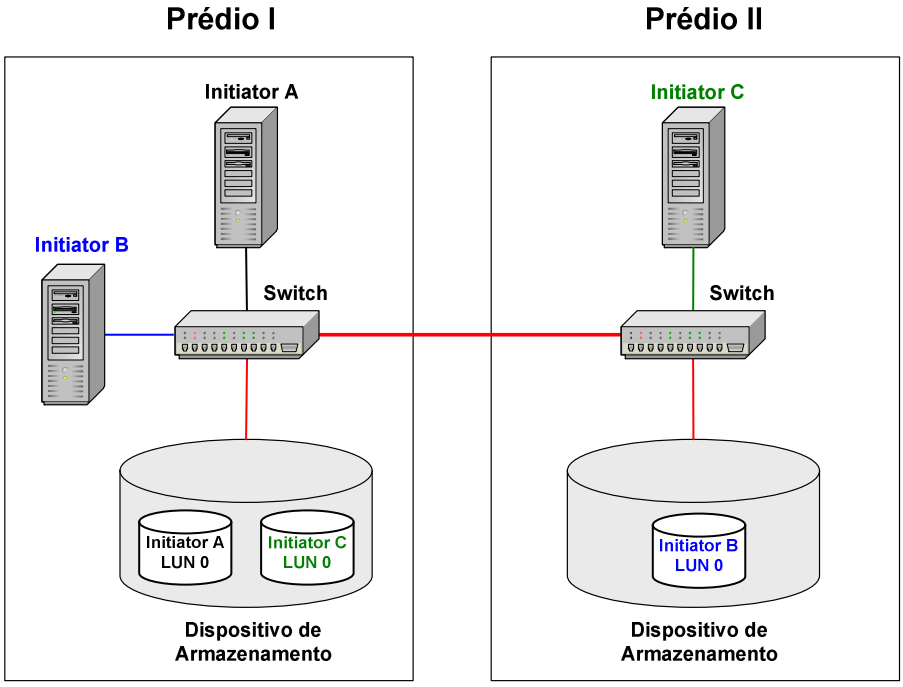


Figura 3 – Arquitetura *Storage Area Network* (SAN)

A arquitetura SAN é, basicamente, uma infraestrutura física Fibre Channel, que utiliza o protocolo Fibre Channel Protocol (FCP) como protocolo de transporte para o protocolo SCSI. O protocolo SCSI é responsável pelo envio de instruções e comandos para o meio de armazenamento, como por exemplo discos rígidos localizados no Dispositivo de Armazenamento.

Conforme ilustra a Figura 4, a possibilidade de uma falha ocorrer em um sistema de armazenento baseado na arquitetura SAN é menor do que na arquitetura DAS, pois a arquitetura SAN é baseada em um modelo de armazenamento em rede (*Storage Area Network*), no qual pode-se implementar caminhos redundantes e alternativos entre os componentes do sistema, proporcionando, assim, uma maior disponibilidade e confiabilidade aos Servidores de Aplicação.

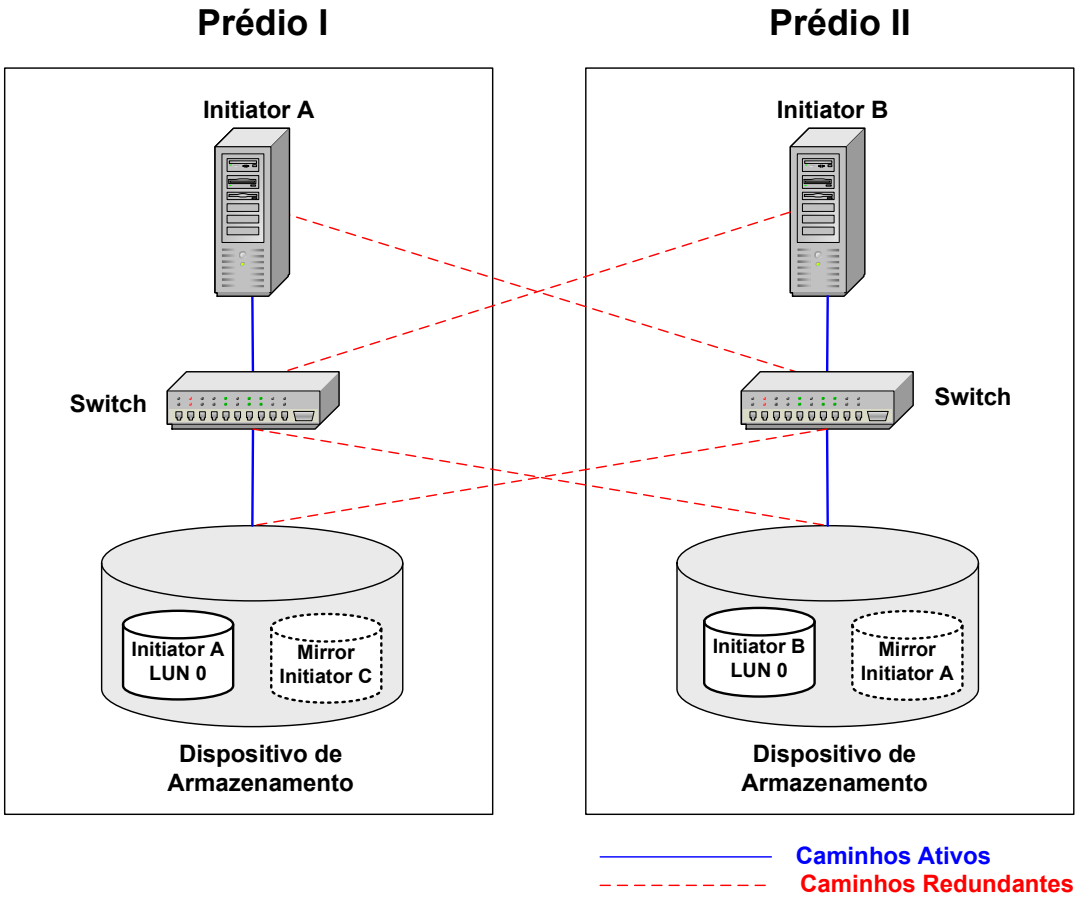


Figura 4 – Alta Disponibilidade – Arquitetura SAN

Capítulo 3

Neste capítulo, são descritos as características e o funcionamento dos protocolos SCSI (Small Computer Systems Interface), Fibre Channel Protocol (FCP) e iSCSI (internet Small Computer Systems Interface).

3.1 Small Computer Systems Interface (SCSI)

SCSI, *Small Computer Systems Interface,* é um padrão utilizado em sistemas de armazenamento com especificações de meios de transporte, métodos de interconexão e protocolos baseados em um modelo de arquitetura SCSI chamada de *SCSI Architecture Model* – SAM3 [6].

Como mencionado anteriormente, a arquitetura SAN utiliza de uma infraestrutura física Fibre Channel, bem como o protocolo Fibre Channel como protocolo de transporte, a fim de proporcionar um transporte confiável e seguro para o protocolo SCSI.

O SCSI originou-se na tecnologia SASI (*Shugart Associates Systems Interface*), desenvolvida pela Shugart Associates em 1979. A tecnologia SASI era bem rudimentar em termos de capacidade e suportava um conjunto de comandos bem limitado. A empresa Shugart Associates tinha como principal objetivo tornar a tecnologia SASI um padrão ANSI e como conseqüência difundi-lo no mercado. Em 1981, a empresa Shugart Associates uniu-se à empresa NCR Corporation e sugeriram a ANSI a organizar um comitê para a análise da tecnologia SASI.

Em 1982, o comitê técnico da ANSI denominado de X3T9.3 foi formado para trabalhar em uma padronização da tecnologia SASI e um grande número de modificações foram realizadas na interface para ampliar o conjunto de comandos e melhorar o desempenho da tecnologia. O nome foi alterado para *Small Computer Systems Interface* (SCSI), pois o nome SASI dava a impressão que a interface seria uma tecnologia proprietária.

O primeiro padrão de interface SCSI foi publicado no ano de 1986, onde o nome da especificação foi definido inicialmente como SCSI. Posteriormente, foi alterado para SCSI-1. O padrão SCSI-1 [7] definia o básico dos primeiros barramentos SCSI em relação ao tamanho do cabeamento, características de sinalização, comandos, modos de transferência e quantidade de dispositivos suportados. O padrão SCSI-1 possuia características de interface paralela e tinha um barramento de velocidade de 5MHz, representando taxas de transferência de dados com velocidades de até 5 MBytes/s, suportando até 7 dispositivos SCSI em um barramento de 8 bits.

O padrão SCSI-1 era bastante limitado. Diversos fabricantes implementaram diferentes superconjuntos de comandos e funcionalidades com a principal finalidade de ampliar a capacidade existente no padrão SCSI-1, diferindo do padrão previamente estabelecido pela ANSI. Atualmente, o padrão SCSI-1 está obsoleto e foi descontinuado pela ANSI.

No ano de 1985, foram iniciados os trabalhos para a especificação do padrão SCSI-2 devido as limitações existentes no padrão SCSI-1. Os principais objetivos do padrão SCSI-2 eram:

- ✓ Aumento do desempenho;
- ✓ Aumento da confiabilidade;
- ✓ Adição de propriedades e capacidades de processamento nas interfaces;
- ✓ Formalização e padronização dos comandos SCSI;

Foi desenvolvido um documento pela ANSI para a definição de um conjunto completo de comandos padronizados para Dispositivos de Armazenamento, o qual foi chamado de *Common Command Set* (CCS). A principal finalidade do CCS era evitar problemas anteriormente ocasionados com as implementações não padronizadas de comandos da tecnologia SCSI original.

O padrão SCSI-2 foi aprovado pela ANSI no ano de 1994 [8] e apresentava melhorias significativas em relação ao padrão SCSI-1. Tinha como uma das principais características a padronização e aperfeiçoamento do conjunto de comandos SCSI utilizado pelos Dispositivos de Armazenamento.

O padrão SCSI-2 define um modelo de interface paralela denominado *FAST SCSI,* o qual implementa um modo de transferência de maior velocidade comparado ao padrão SCSI-1. A velocidade do barramento passou de 5 MHz para 10 MHz representando uma velocidade nominal de 10 MBytes/s com cabeamento regular de 8 bits.

O padrão SCSI-2 também ampliou a largura do barramento original de 8 bits para 16 bits. Esse modelo foi chamado de *WIDE SCSI,* o qual possibilitou o suporte a 16 dispositivos no barramento. O padrão SCSI-2 definiu também uma densidade de conexões mais elevada, ampliando a utilização dos conectores básicos definidos anteriormente no padrão SCSI-1.

Um dos fatores principais introduzidos com o padrão SCSI-2 foi a possibilidade de permitir o enfileiramento de comandos, o qual possibilita o tratamento de solicitações não atendidas entre dispositivos de comunicação em um barramento SCSI.

O padrão SCSI-2 adicionou novos conjuntos de comandos para suporte a uma maior diversidade de dispositivos, tais como, unidades de *Compact Disk* (CD) e *Scanners*, ampliando o padrão SCSI-1, que tinha como foco principal os discos rígidos.

As principais melhorias com o surgimento do padrão SCSI-2 em relação ao padrão anterior foram:

- ✓ Possibilidade de barramentos mais largos com 16 bits;
- ✓ Aumento do clock de 5 MHz para 10 MHz;
- ✓ Novos comandos para permitir a utilização de outros dispositivos, como unidades de CD e *Scanners*;

- ✓ Enfileiramento de comandos para Dispositivos de Armazenamento;
- ✓ Capacidade da controladora do Dispositivo de Armazenamento de reordenar os comandos para obter a máxima eficiência possível na execução dos mesmos;
- ✓ Inclusão de comandos para permitir diagnóstico de dispositivos;

O SCSI-2 tornou-se um padrão "de facto" para as interfaces de comunicação nos Servidores de Aplicação, pois necessitavam de desempenho e confiabilidade para o acesso ao Dispositivo de Armazenamento. Entretanto, comprovou-se que o padrão SCSI-2 possuia algumas limitações relacionadas as suas características físicas e elétricas pois baseava-se na utilização de cabos paralelos de cobre, nos quais sinais SCSI eram trafegados tipicamente atráves de doze pares de condutores em paralelo.

As principais limitações do padrão SCSI-2 são:

- ✓ O SCSI-2 é um padrão baseado em interfaces paralelas e era limitado em até 16 dispositivos no barramento;
- ✓ Interferências elétricas entre os condutores individuais no cabo paralelo;
- ✓ Limitações de distância, com o limite de até 25 metros;

Os padrões SCSI-1 e SCSI-2 apresentam inflexibilidade de componentes e recursos, dado que esses padrões não foram projetados e organizados em camadas. Nestes, uma única camada está composta de nível físico, interfaces e protocolos de comunicação, tornando impossível a utilização de um outro meio físico ou de um outro protocolo de transporte para a implementação do padrão SCSI-2.

A solução para essas limitações foi o desenvolvimento do padrão SCSI-3, que têm como principal característica um modelo em camadas [9]. O desenvolvimento do padrão SCSI-3 possibilitou a utilização de outros meios físicos e outros meios de transporte (Figura 5) para o protocolo SCSI. O padrão SCSI-3 define diferentes padrões para cada conjunto de comandos, protocolos e métodos de sinalização.

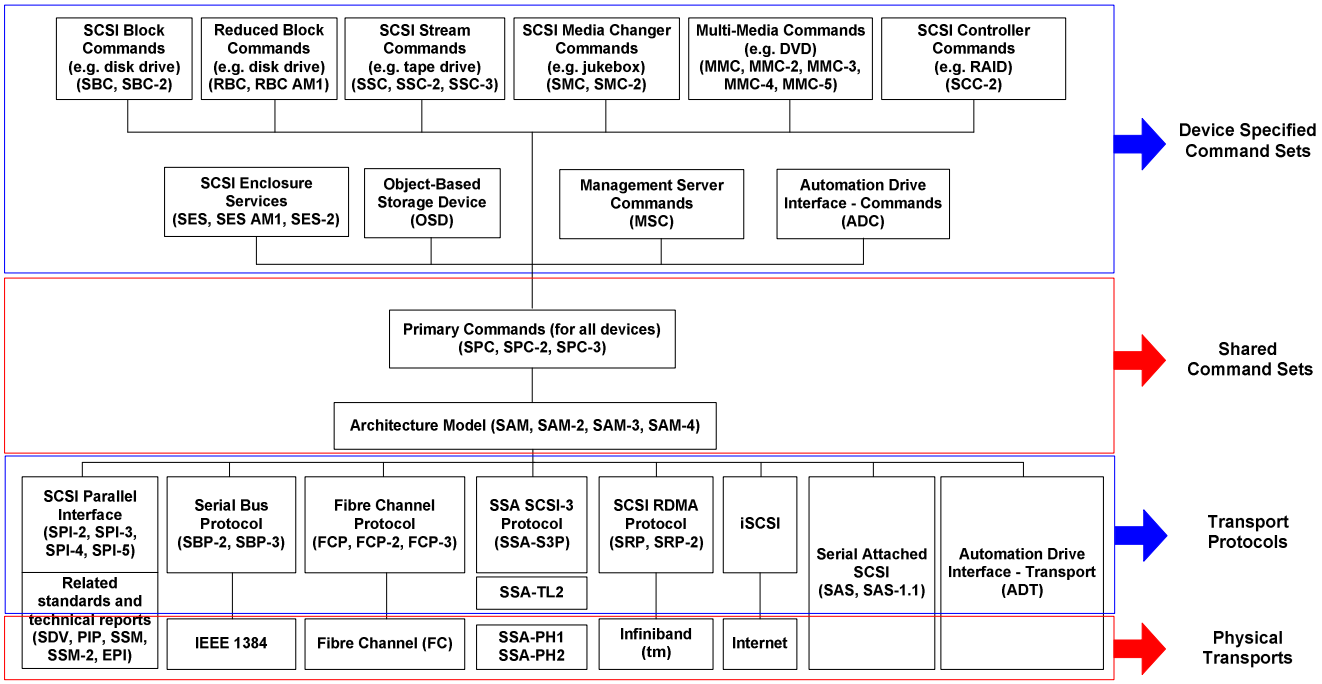


Figura 5 – Padrão SCSI-3 (SAM-3)

A Figura 5 organiza e classifica os diversos padrões existentes no padrão SCSI-3 pertencentes as seguintes categorias:

- ✓ ***Device Specific Command Set*:** define conjuntos de comandos específicos para todos os dispositivos SCSI ou para tipos particulares de dispositivos SCSI, tais como unidades de CD e *Scanners*;

- ✓ ***Shared Command Set*:** define um conjunto comum de comandos permitindo que a camada de transporte acesse dados, comandos e informações de estado entre os dispositivos;
- ✓ ***Transport Protocols*:** formaliza as regras pelas quais vários dispositivos possam se comunicar e trocar informações, permitindo a interoperabilidade de diferentes dispositivos;
- ✓ ***Physical Protocols*:** define detalhes específicos da interface, tais como métodos de sinalização elétrica e modos de transferência, os quais são geralmente chamados de padrão da camada física.

Pode-se observar que no padrão SCSI-3 (Figura 5), a camada de comandos SCSI (*Device Specific Command Set)* está separada da camada de transporte (*Transport Protocols).* Com essa divisão em camadas, permite-se escolher o tipo de transporte mais adequado para os comandos SCSI a serem enviados para o Dispositivo de Armazenamento.

Conforme mencionado anteriormente, a arquitetura SAN utiliza tipicamente o protocolo de transporte Fibre Channel para prover o transporte dos comandos SCSI – mensagem SCSI CDB para o Dispositivo de Armazenamento. (Figura 6).

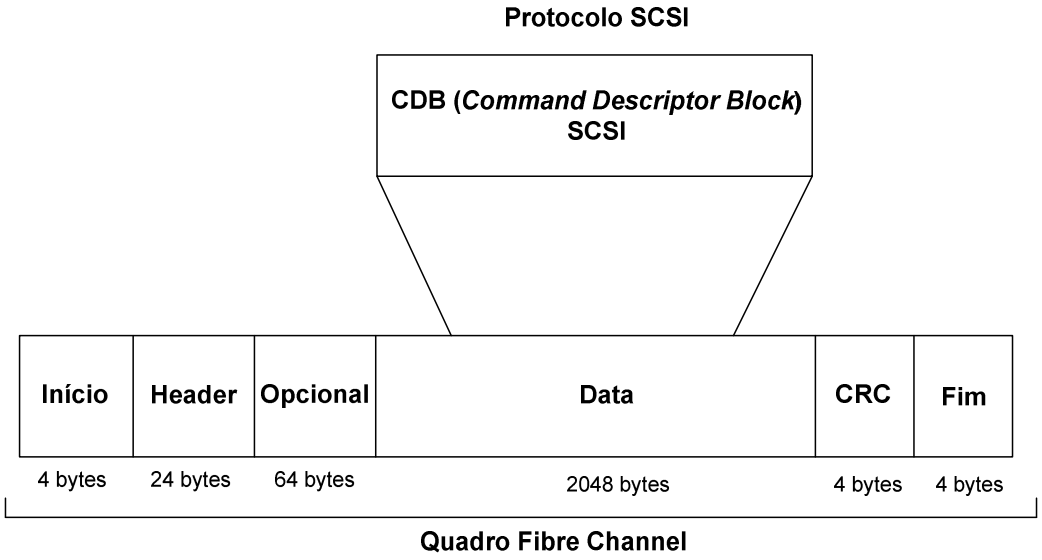


Figura 6 – Mensagem SCSI CDB + Quadro Protocolo Fibre Channel

O modelo de arquitetura SCSI – SAM-3 define um relacionamento entre Servidores de Aplicação os quais são chamados de ***Initiator*s** e Dispositivos de Armazenamento de dados chamados de ***Target*s**. (Figura 7)

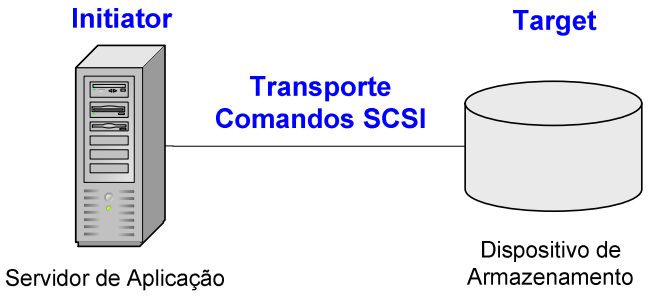


Figura 7 – Modelo de conexão SCSI

Um ***SCSI Class Driver*** é um *device driver* SCSI instalado no *Initiator* responsável por converter requisições das aplicações de sistema em comandos SCSI-3. Como pode ser visto na Figura 6, os comandos SCSI-3 são armazenados em uma estrutura SCSI-3, chamada de SCSI-3 CDB (*Command Descriptor Blocks*). Essa estrutura CDB é repassada para a camada de transporte adotada, a qual têm como principal finalidade a entrega dessa estrutura para o Dispositivo de Armazenamento (*Target*). Este é responsável pela interpretação dos comandos contidos na estrutura CDB, pelo recebimento dos dados e pela entrega dos comandos para a unidade lógica de armazenamento apropriada. Após a execução dos comandos pelo *Target*, o mesmo envia uma mensagem SCSI-3 CDB de confirmação da execução dos comandos para o *Initiator*.

O modelo de conexão SCSI-3 é um modelo cliente-servidor, onde os envios das requisições de E/S e as respostas a essas requisições são transportadas através de um protocolo de transporte em um meio físico de interconexão.

O protocolo SCSI possui um modelo de operação E/S constituíndo três fases: (Figura 8)

- ✓ **Command:** fase em que é enviada o comando requerido e parâmetros através da estrutura CDB;
- ✓ **Data:** fase para transferência dos dados (informações) em concordância com a fase Command;
- ✓ **Status:** fase responsável pelo envio da confirmação da execução do comando;

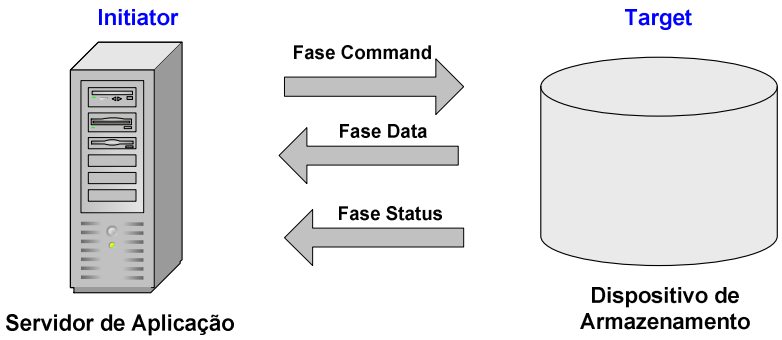


Figura 8 – Modelo de Fases do Protocolo SCSI

Um Dispositivo de Armazenanamento (*Target*) é composto de unidades lógicas de armazenamento chamadas de *Logical Unit* (LU) onde são executados os comandos de E/S. As mensagens de comunicação utilizadas pelo protocolo SCSI são chamados de SCSI CDB e têm como principal finalidade o transporte de todas as operações e comandos SCSI para o Dispositivo de Armazenamento.

A entidade *Initiator* contém tipicamente inúmeras unidades de armazenamento, como, por exemplo, vários discos rígidos. A unidade *Target* contém tipicamente uma ou várias *Logical Unit* (LU), as quais são identificadas através de um número sequencial associado a entidade *Initiator* sendo chamado de *Logical Unit Number* (LUN), (Figura 9).

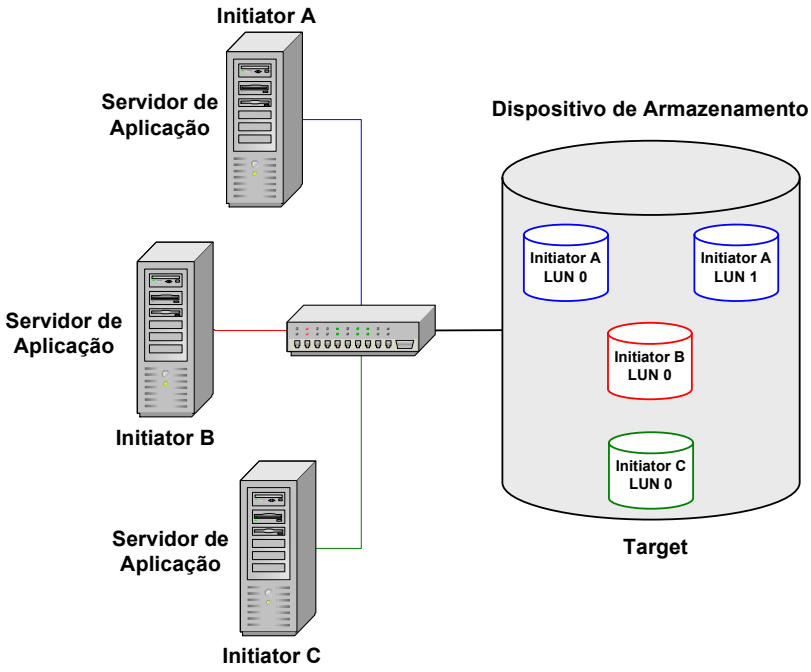


Figura 9 – Componentes *Initiator* e *Target – LUN*

A Figura 9 ilustra um cenário onde cada Servidor de Aplicação (*Initiator*) possui uma ou mais LU configuradas em um Dispositivo de Armazenamento (*Target*). Por exemplo, o Servidor de Aplicação (*Initiator* A) possui duas unidades lógicas de armazenamento (*Logical Unit*) identificadas como (LUN 0) e (LUN 1). Tanto o Servidor de Aplicação (*Initiator* B) quanto o Servidor de Aplicação (*Initiator* C) possuem cada um uma unidade lógica de armazenamento identificadas como Servidor de Aplicação (*Initiator* B) (LUN 0) e Servidor de Aplicação (*Initiator* B) (LUN 0).

A entidade *Target* possui dois componentes essenciais (Figura 10), para o gerenciamento e processamento das requisições e comandos SCSI. Esses componentes são chamados de *Task Manager* e *Device Server*. A principal função do *Task Manager* é proporcionar o gerenciamento das tarefas de sequenciamento das conexões iniciadas no *Initiator* com destino uma unidade lógica de armazenamento (*Logical Unit*) específica.

Já a principal função do *Device Server* é o processamento das operações originadas no Servidor de Aplicação (*Initiator*) e o direcionamento dessas operações para uma unidade lógica de armazenamento (*Logical Unit*) específica.

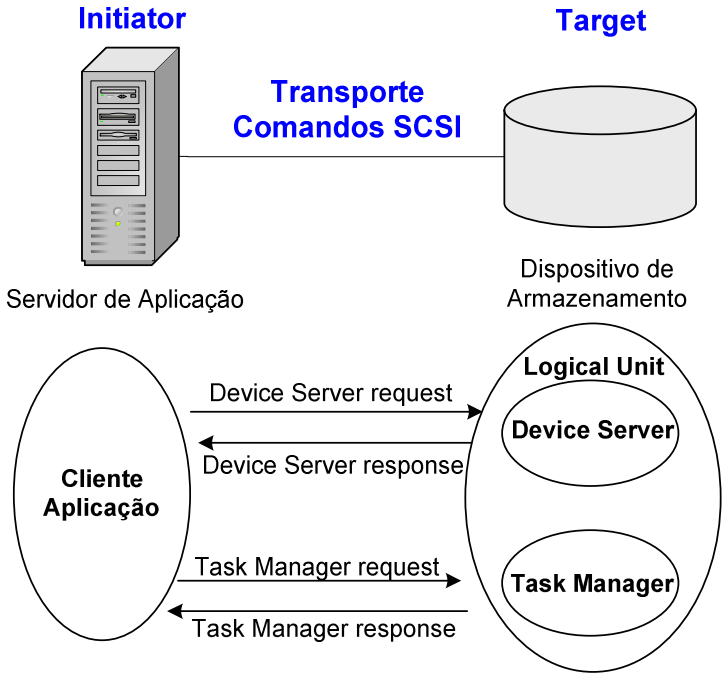


Figura 10 – Componentes *Target* – *Task Manager* e *Device Server*

Uma *Logical Unit Numbe*r (LUN) é a identificação de uma ou mais *Logical Unit* (LU) no Dispositivo de Armazenamento (*Target*). O campo de informação referente ao LUN, nos padrões SCSI-1 e SCSI-2, estava contido na estrutura SCSI CDB e possuia somente 3 bits de tamanho, permitindo a de identificação de 2^3 Logical Unit (LU).

Atualmente com a especificação SCSI-3, o campo de informação referente a LUN foi movido da estrutura SCSI-3 CDB para a camada de transporte, onde essa camada é responsável pelo transporte das mensagens SCSI CDB para o Dispositivo de Armazenamento, possuindo 64 bits para indentificação das *Logical Unit* (LU).

Os campos na estrutura SCSI-3 CDB [11] são, Figura 11:

- ✓ ***Operation Code*****:** contém o tamanho do CDB que pode ser de 6, 10, 12 e 16 bytes e a operação a ser executada no *Target*;
- ✓ ***Logical Block Address:*** contém a informação de onde o dado requerido está localizada no meio físico. Blocos lógicos começam com o número 0 (zero) e são contínuos até a localização do último bloco no meio físico do dispositivo. A menor unidade de medida de um disco rígido é um bloco de dados que é especificado por um quantidade de bytes. O tamanho típico de um bloco de um disco rígido é de 512 bytes ou 520 bytes;
- ✓ ***Transfer Length*****:** indica a quantidade de dados transmitidos associados com o mensagem SCSI CDB na forma de blocos;
- ✓ ***Control Byte:*** contém informações de controle para os Dispositivos de Armazenamento (*Target*);

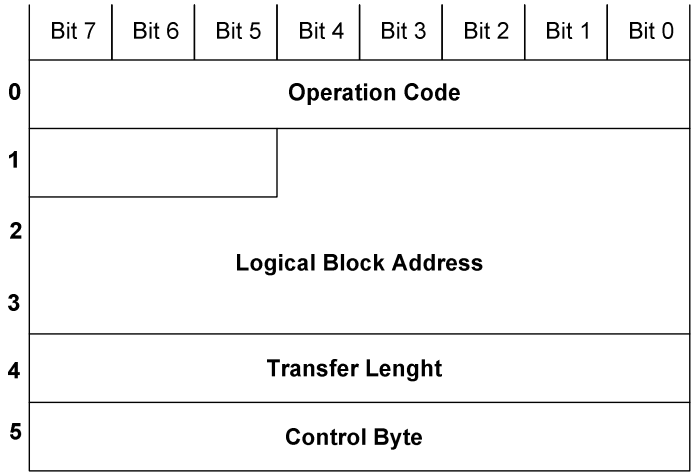


Figura 11 – Command Descriptor Block (CDB)

Portanto, o protocolo SCSI é largamente utilizado em arquitetura *Storage Area Network* (SAN), onde o termo SCSI é utilizado na arquitetura SAN como um **protocolo de comunicação** entre as Aplicações de Sistema e os Dispositivos de Armazenamento (*Target*). Utiliza tipicamente uma infraestrutura de componentes Fibre Channel baseada em um protocolo de transporte Fibre Channel (FCP), implementando, assim, uma arquitetura *Storage Area Network* (SAN) – FCP.

3.2 Infraestrutura e Protocolo Fibre Channel

A arquitetura *Storage Area Network* (SAN) – FCP é atualmente a arquitetura de sistemas de armazenamento mais utilizada pelas corporações, pois proporciona um modelo de armazenamento em rede (*Storage Area Network)* que implementa caminhos redundantes e alternativos entre os componentes do sistema, proporcionando uma maior disponibilidade e flexibilidade aos Servidores de Aplicação. (Figura 4)

Como mencionado anteriormente, uma SAN (*Storage Area Network*) consiste em um infraestrutura de comunicação, no qual fornece conexões físicas e uma camada de gerenciamento de todas as conexões, elementos de armazenamento e Aplicações de Sistema, tornando a transferência de dados mais segura e robusta.

As implementações da arquitetura SAN baseiam-se em três componentes:

- ✓ **Componentes de interconexão:** responsáveis pela conexão física entre os Servidores de Aplicação (*Initiator*) e o Dispositivo de Armazenamento (*Target*) ;

- ✓ **Protocolo de Transporte:** responsável pelo transporte das mensagens SCSI CDB para o Dispositivo de Armazenamento (*Target*). O protocolo de transporte utilizado atualmente pela arquitetura SAN é o protocolo Fibre Channel (FCP);

- ✓ **Protocolo de Comunicação:** responsável pela comunicação entre as entidades *Initiator* e *Target.* O protocolo de comunicação utilizado pela arquitetura SAN é o protocolo SCSI (SCSI CDB).

A tecnologia Fibre Channel emergiu como uma solução rápida e confiável para armazenamento, recuperação e transferências de dados entre os Servidores de Aplicação e os Dispositivos de Armazenamento, tendo sido adotada como uma tecnologia "padrão" para a arquitetura SAN.

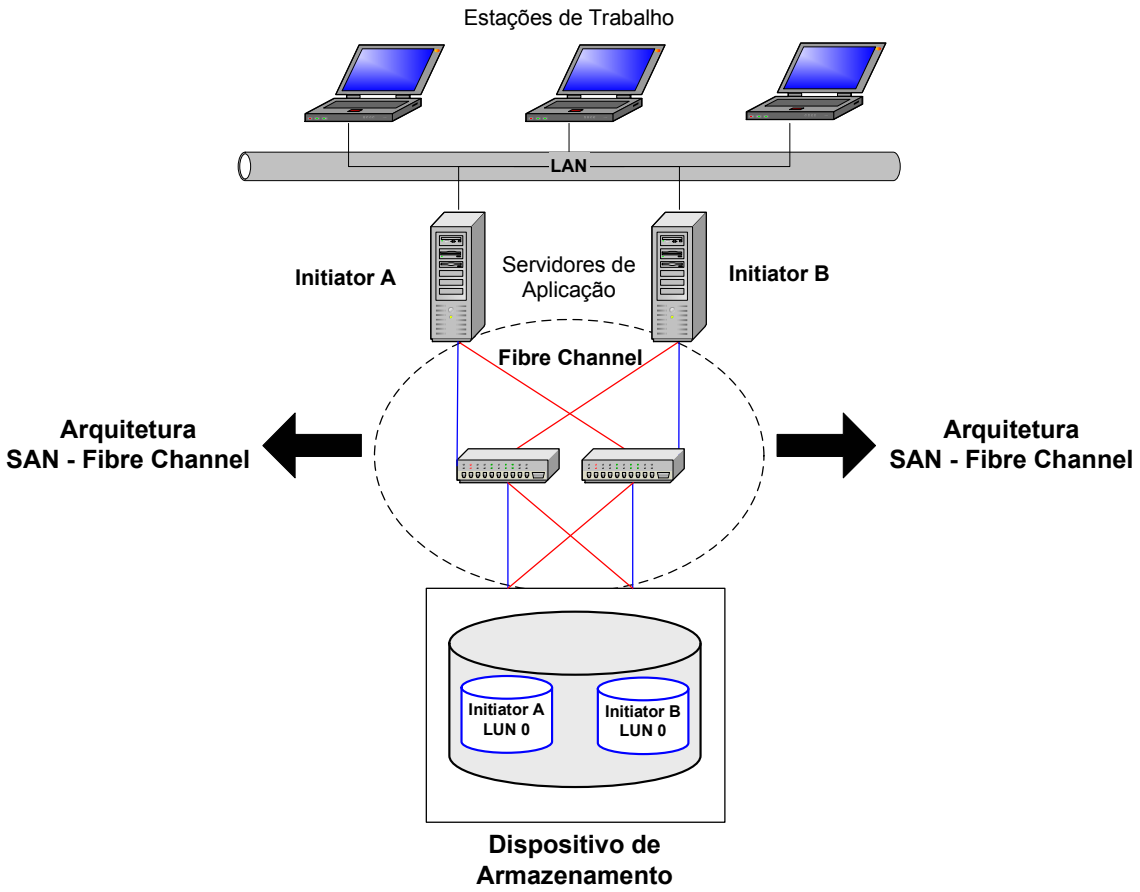


Figura 12 – Arquitetura SAN – Infraestrutura e Protocolo Fibre Channel

Os Servidores de Aplicação estão interconectados, através de caminhos redundantes, com switches Fibre Channel implementando uma rede de armazenamento de dados denominada de Arquitetura SAN – FCP (Figura 12).

Do mesmo modo que a tecnologia SCSI, a tecnologia Fibre Channel possui um conjunto de padrões, desenvolvidos pelo *American National Standards Institute* – ANSI. A tecnologia Fibre Channel tem como principal finalidade superar as limitações da

infraestrutura SCSI, além de proporcionar a possibilidade de implementação de uma infraestrutura de armazenamento de dados baseada em um modelo de rede (*Storage Area Network*). Os principais objetivos da tecnologia Fibre Channel são:

- ✓ Possibilitar a implementação de um modelo de rede (*Storage Area Network*) entre os componentes do sistema de armazenamento;
- ✓ Possibilitar alta velocidade de transferência de dados entre *Initiators* e *Targets*;
- ✓ Permitir a implementação de uma solução de transporte e de infraestrutura com suporte a ambientes heterogêneos;
- ✓ Possibilitar o aumento da distância para 10KM entre as entidades *Initiator* e *Target*;
- ✓ Permitir a possibilidade de reusabilidade de protocolos e infraestruturas existentes.

As vantagens da infraestrutura de interconexão Fibre Channel [12] em relação a infraestrutura de interconexão SCSI estão apresentadas na Tabela 2:

	SCSI	Fibre Channel
Taxa de Transferência	O máximo de taxa de transferência suportado pela arquitetura SCSI é de 80 Mbps.	A taxa de transferência oferecida pela tecnologia Fibre Channel é de 1Gbps e 2Gbps.
Número Máximo Dispositivos	O número máximo de dispositivos SCSI suportados simultaneamente é de 32 dispositivos. Mas o protocolo SCSI somente pode suportar 15 dispositivos de uma vez.	O número máximo de dispositivos Fibre Channel suportados varia de 126 até 2^{24} (16.777.216) dispositivos.
Distância	O comprimento máximo de um cabo SCSI permitido entre um Initiator e um Target é de 25 metros.	A distância máxima de um cabo de cobre Fibre Channel é de 30 metros e utilizando cabos ópticos varia de 500 metros a 10 Km.
Tolerância a Falhas	Suporte a tolerância a falhas na arquitetura SCSI é muito limitado pois utiliza-se tipicamente de uma conexão ponto a ponto entre o Initiator e o Target, com isso reduzindo a possibilidade de implementação de alta disponibilidade entre as entidades envolvidas.	Suporte a implementações tolerantes a falhas, pois com a arquitetura Fibre Channel torna-se possível a implementação de switches e a possibilidade de uma configuração de caminhos redudantes e alternativos entre as entidades Initiator e Target.

Tabela 2 – Vantagens Infraestrutura Fibre Channel

A infraestrutura Fibre Channel utiliza um modelo de interconexão cliente-servidor baseado no modelo SCSI, no qual cada ponto de interconexão é denominado de Porta Fibre Channel. Esta possui a finalidade do estabelecimento da conexão física entre os elementos de um sistema de armazenamento de dados, que podem ser classificados como Servidores de Aplicação, Switches e Dispositivos de Armazenamento.

As portas suportadas pela arquitetura Fibre Channel (Figura 13) são:

- ✓ **Node Ports** (N_Port)**:** existentes nos *Initiators* e nos *Targets*. As portas do tipo N_Port são utilizadas para estabelecerem um conexão física com um elemento de rede, por exemplo um Switch Fibre Channel, chamado de ***Fabric Switch***. Esse tipo de porta pode estar somente conectado a outra porta do mesmo tipo N_Port em uma conexão ponto a ponto ou a portas de um Switch Fibre Channel do tipo F_Ports;

- ✓ **Fabric Ports** (F_Ports)**:** existentes nos Switches Fibre Channel e tem como principal finalidade atuarem como portas intermediárias na comunicação entre as portas do tipo N_Ports. As portas do tipo F_Ports somente podem estar conectadas em portas do tipo N_Ports;

- ✓ **Loop Ports** (L_Ports)**:** são portas utilizadas pela topologia Fibre Channel Arbitration Loop (FC-AL). São geralmente utilizadas para o compartilhamento simples de largura de banda entre Dispositivos de Armazenamento. Um simples componente de uma infraestrutura de sistemas de armazenamento que implementa a tecnologia FC-AL é um Hub Fibre Channel. Portanto, um componente interconectado em um componente FC-AL utiliza a sua porta específica, como por exemplo na Figura 13 o *Initiator* C está interconectado com o Hub Fibre Channel através de uma porta FC-AL do tipo NL_Port (Node Loop Port) e o Switch Fibre Channel está interconectado ao componente FC-AL através de uma porta de interconexão do tipo FL_Port (Fabric Loop Port);

✓ **Expansion ports** (E_Ports)**:** fazem parte da arquitetura do Switch Fibre Channel e são utilizadas como portas de expansão para interconexão entre componente Fabric Switch. Podem somente ser conectadas a outras portas do mesmo tipo E_Ports;

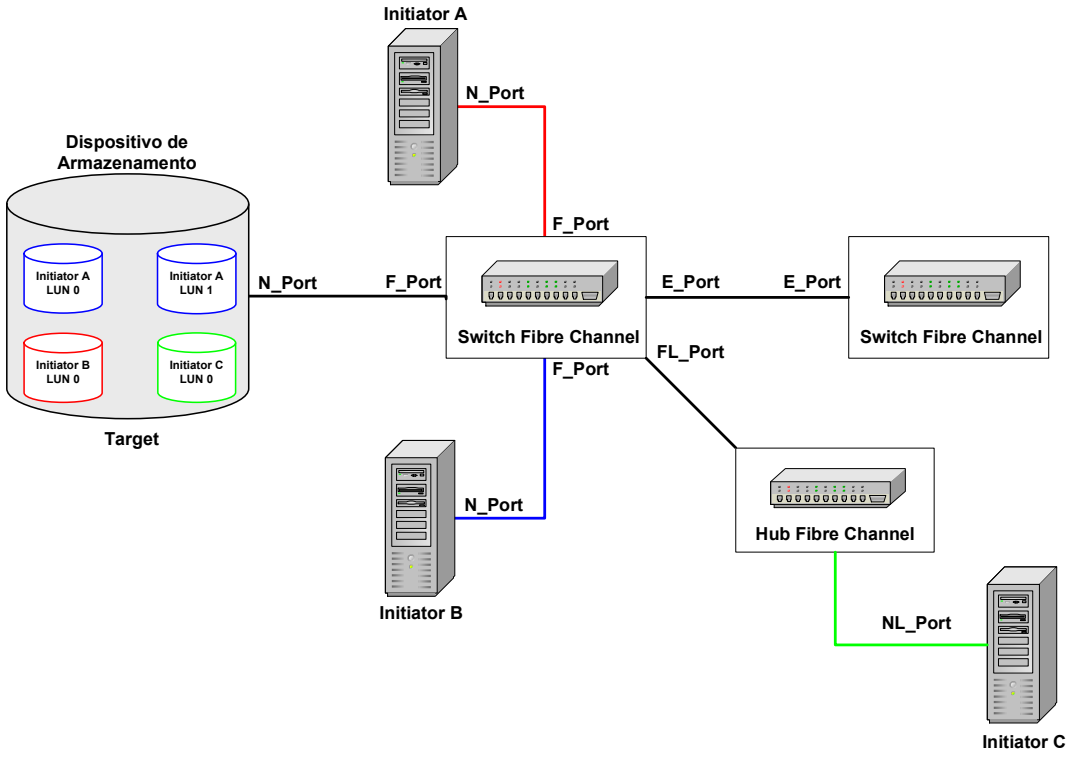


Figura 13 – Portas Fibre Channel

3.2.1 Topologias de Redes Fibre Channel

A tecnologia Fibre Channel é uma arquitetura que possui três topologias de conexão. Cada uma oferece diferentes formas de interconexão na implementação da arquitetura SAN [12] [13].

As topologias estão descritas a seguir:

Topologia *Point-to-Point*

Na topologia *Point-to-Point* (Figura 14) tanto o *Initiator* quanto o *Target* estão conectados diretamente através de portas do tipo N_Ports. No modelo de interconexão SCSI existe tipicamente uma conexão direta entre as entidades *Initiator* e *Target*. A topologia *Point-to-Point* não é uma topologia compartilhada, portanto, somente as duas entidades podem estar conectadas simultaneamente ao canal.

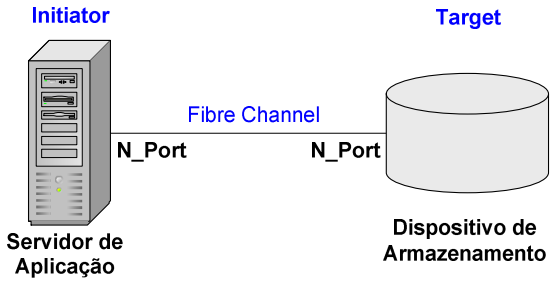


Figura 14 – Topologia *Point-to-Point*

A topologia *Point-to-Point* é a topologia Fibre Channel mais simples e fácil de se implementar e pode ser caracterizada também como um arquitetura DAS, possuindo todas as limitações quanto a inflexibilidade de conexão, falta de tolerância à falhas e não consolidação dos dados.

Topologia FC-AL (Fibre Channel Arbitrated Loop)

A topologia FC-AL é utilizada na interconexão de dispositivos em uma rede de armazenamento no formato de um *loop* (Figura 15). Nesse loop, cada componente é interligado entre si, tendo-se, então, uma topologia escalável possibilitando, a adição de mais componentes na arquitetura de armazenamento.

Essa topologia utiliza um endereçamento físico chamado de *Arbitrated Loop Physical Address* (AL_PA) no qual uma *string* hexadecimal de 8 bits define o endereçamento do dispositivo no loop e cada componente compartilha a largura de banda disponível no loop. Com a utilização da topologia FC-AL, pode-se ter conectado na forma de um *"loop"* até 127 dispositivos. Depois que um dispositivo é adicionado ao controle do loop, ele estabiliza uma conexão ponto a ponto com o próximo componente do loop e inicia a transferência dos dados.

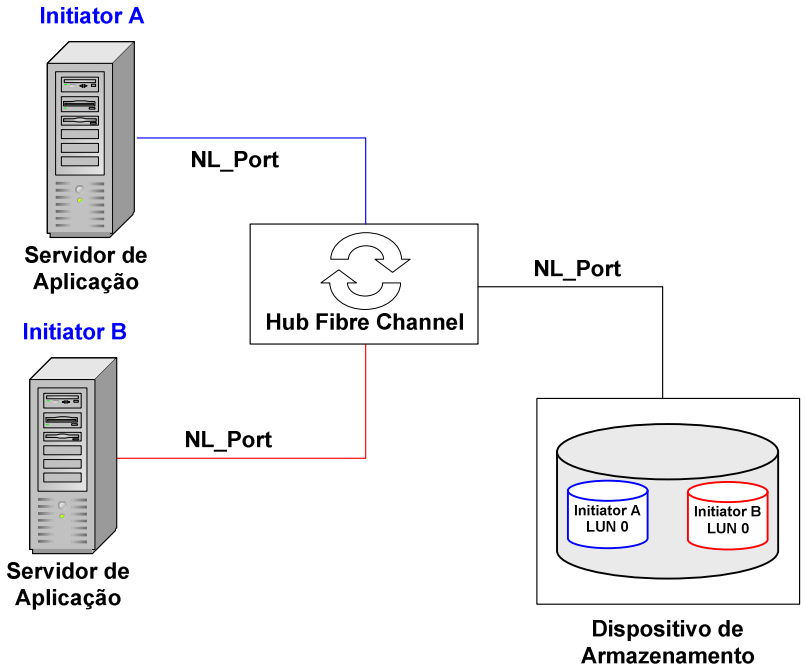


Figura 15 – Topologia FC-AL

Topologia Switched

Na topologia *Switched*, um conjunto de um ou mais elementos de rede, tipicamente Switches Fibre Channel são utilizados para interconectar *Initiators* e *Targets*, de forma a implementar um modelo de armazenamento em rede (*Storage Area Network*) para o compartilhamento de recursos e componentes. A topologia *Switched* é também conhecida como topologia Fabric podendo suportar até 2^{24} nós correspondendo a 16.777.216 de dispositivos.

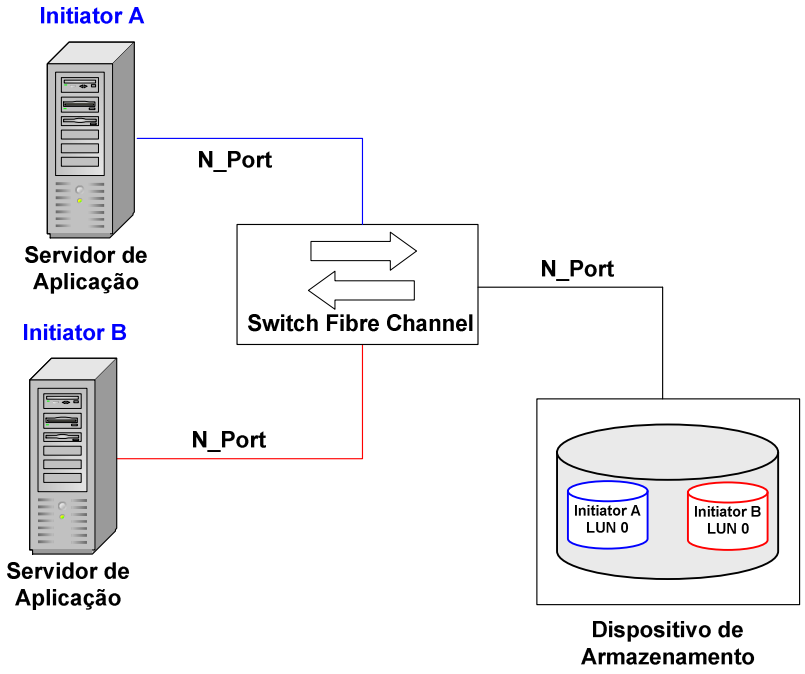


Figura 16 – Topologia Switched

A principal vantagem da topologia Switched em relação a topologia FC-AL é que cada componente possui equivalência na largura de banda disponível no Fabric, diferentemente da topologia FC-AL, na qual todos os componentes do sistema estão interconectados em um *"loop"* cuja a largura de banda é compartilhada entre as entidades do sistema de armazenamento.

As principais características da topologia Switched são:

Topologia Fibre Channel Switched	
Largura de Banda	Cada entidade no sistema de armazenamento possui equivalência na largura de banda disponível no Fabric;
Dispositivos	Suporte a um grande número de dispositivos - 2^{24}, correspondendo a 16.777.216 de dispositivos;
Tolerância a Falhas	Possibilidade de implementação de sistemas de armazenamento de dados com tolerância a falhas, pois com a utilização de uma topologia em rede possibilitando a implementação de caminhos redundantes entre os elementos da rede de armazenamento proporcionando assim um melhor nível de disponibilidade no caso de falhas (Figura 4)
Escalabilidade	Proporciona escabilidade e flexibilidade, pois as entidades envolvidas no sistema de armazenamento podem ser substituídas ***online*** sem afetar outros dispositivos que estão conectados em outras portas do Fabric;

Tabela 3 – Características Topologia Switched (Fabric)

3.2.2 Camadas do Protocolo Fibre Channel

O desenvolvimento do protocolo Fibre Channel [14], começou no ano de 1988 mas foi padronizado pela *American National Standard Institute* (ANSI) somente no ano de 1994. Atualmente, o responsável pelo desenvolvimento e padronização da tecnologia Fibre Channel é o Technical Committee T11.

A tecnologia Fibre Channel é dividida em cinco camadas (Figura 17) :

- ✓ FC-0 – Nível Físico;
- ✓ FC-1 – Codificação e Decodificação;
- ✓ FC-2 – Enquadramento e Controle de Fluxo;
- ✓ FC-3 – Serviços Comuns;
- ✓ FC-4 – Mapeamento de Protocolos ;
- ✓ Upper Layer Protocols (ULP).

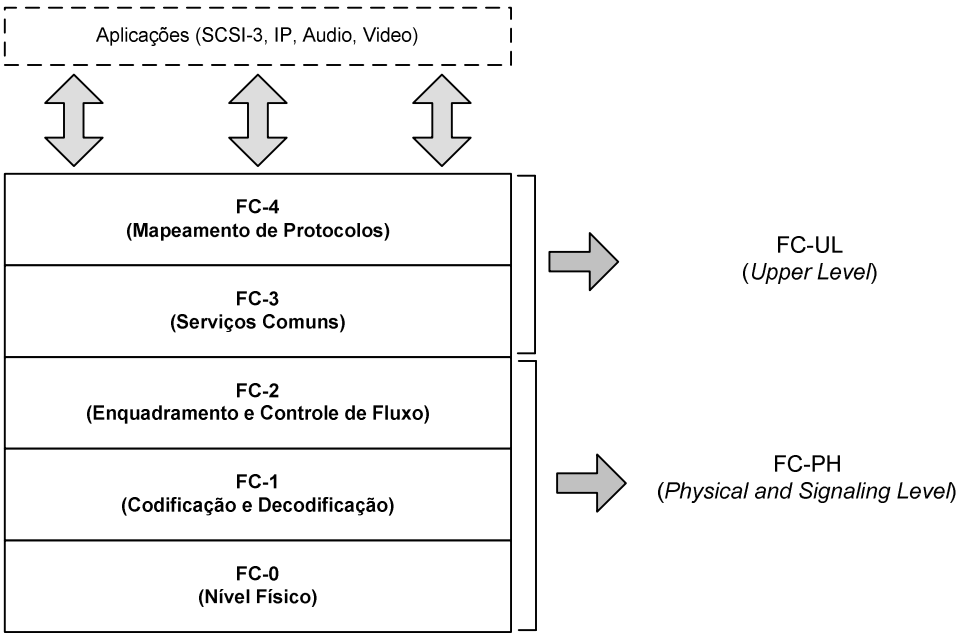


Figura 17 – Camadas Fibre Channel

A camadas do tipo FC–PH faz parte do nível funcional e referencia as camadas FC-0, FC-1 e FC-2 e estão presentes em todos os componentes Fibre Channel em um sistema de armazenamento. Já a camada FC–UL refere-se as camadas FC-3 e FC-4, existentes somente no Servidor de Aplicação e no Dispositivo de Armazenamento. Proporciona serviços e mapeamento para os protocolos de nível superior e para as Aplicações de Sistema (Figura 18).

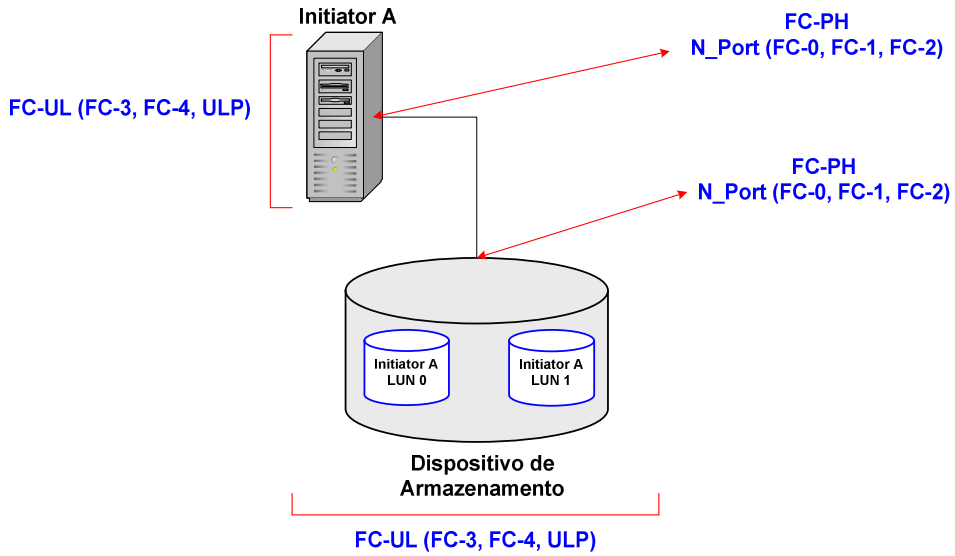


Figura 18 – Camadas FC-UL e FC-PH

A seguir, são descritas as diversas camadas da tecnologia Fibre Channel [15]:

Camada FC-0

A camada FC-0 é a camada mais inferior do modelo de camadas Fibre Channel sendo responsável por:

- ✓ Tipo de Cabeamento utilizado na arquitetura SAN FCP;
- ✓ Parametros ópticos ou elétricos para controle da taxa dos dados;
- ✓ Conectores.

A camada física foi projetada para gerenciar os requisitos físicos relacionados a conectores, parâmetros elétricos e taxa de dados, oferecendo alta flexibilidade e confiabilidade para o protocolo Fibre Channel na utilização da camada física. Por exemplo, o cabeamento de cobre pode ser usado para comunicação em distâncias até 30 metros e o cabeamento óptico utilizado para comunicação em distâncias de 10 Km. O cabeamento óptico é o mais utilizado na arquitetura SAN devido a sua flexibilidade de manuseio e a possibilidade de interligação de componentes a longas distâncias.

Camada FC-1

A camada FC-1 do modelo Fibre Channel têm como principais responsabilidades:

- ✓ Conversão de caracteres e regras de codificação e decodificação;
- ✓ Detecção de erros;
- ✓ Sincronização de bytes e alinhamento de *words;*
- ✓ Manutenção do Link.

A camada FC-1 utiliza o método 8B/10B [12] [15] para codificar e decodificar informações. Esse método de codificação, desenvolvido pela IBM, é considerado um dos melhores métodos, não apenas para codificação de informações, mas também para o controle de erros que possam ocorrer durante uma transmissão. As informações em 8 bits são representados por grupos de código de 10 bits.

Camada FC-2

A camada FC-2 é também conhecida como *Framing* e *Signaling Protocol Layer* sendo responsável pelas seguintes funções:

- ✓ Segmentação e Remontagem de dados dentro dos quadros;
- ✓ Controle de Seqüência na qual os dados estão sendo transferidos;
- ✓ Provisão de vários serviços de classes para garantir transmissões eficientes.

A camada FC-2 suporta as três topologias Fibre Channel e é responsável pelo sucesso e a confiabilidade da transmissão dos dados sem colisão ou perda especificando o mecanismo de transporte dos dados independentemente das camadas superiores.

As transferências de dados de responsabilidade da camada FC-2 são feitas através de protocolos com a principal finalidade de gerenciamento na transferência de dados. Os protocolos podem ser:

- ✓ ***Primitive Sequence Protocol*** – inicia a transferência de uma seqüência de quadros no caso de falha de um link;
- ✓ ***N_Port Login Protocol*** – auxilia as portas do tipo N_Ports utilizadas pelos *Initiators* e *Targets*, no estabelecimento da comunicação e a troca de parâmetros utilizados no suporte de serviços e na comunicação entre as entidades;
- ✓ ***Fabric Login Protocol*** – tem como principal objetivo prover o auxílio ao Switch Fibre Channel na identificação da fase de login de uma porta do tipo N_Port, auxiliando na coleta de informações e na negociação de parâmetros de serviços;

- ✓ ***Data Transfer Protocol*** – tem como principal objetivo auxiliar a camada FC-1 no suporte aos métodos para transferência de dados entre as entidades *Initiators* e *Target*;
- ✓ ***N_Port Logout Protocol*** – tem por finalidade terminar a conexão e liberar os recursos entre as portas envolvidas na comunicação;

A camada FC-2 também é responsável pela multiplexação e demultiplexação das informações em unidades denominadas **quadros**, sendo definidos como unidades básicas de transmissão de dados do protocolo Fibre Channel.

Na arquitetura SAN, por exemplo, o componente Fabric é responsável pela entrega de todos os quadros recebidos para porta de destino requerida [14]. A camada FC-2 é responsável pela multiplexação de todos os quadros que foram originalmente enviados pelo *Initiator* e os entrega para as camadas de nível superior no *Target*. Os quadros podem ser classificados em *Data Frames* e *Link Control Frames.* Quadros do tipo *Data Frames* (Figura 19) transportam somente dados e podem variar de 0 a 2048 bytes de comprimento enquanto os quadros do tipo *Link Control Frames* têm como finalidade prover suporte as informações sobre o estado dos links e os suporte aos reconhecimentos (*ACK*) na entrega dos quadros Fibre Channel.

O comprimento máximo de um quadro Fibre Channel (Figura 19) é de 2148 bytes de comprimento. Esse quadro é responsável por armazenar os dados referente ao protocolo SCSI – mensagem SCSI CDB.

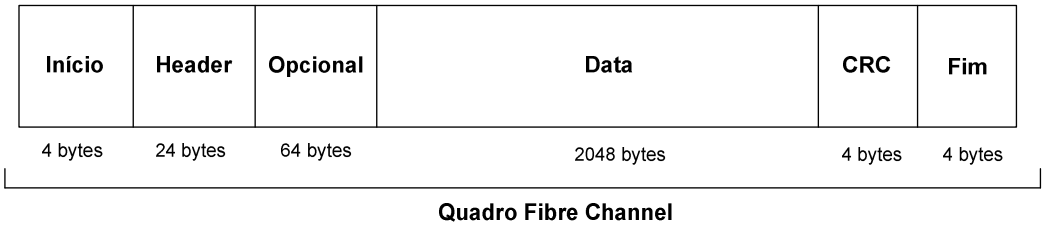


Figura 19 – Quadro Protocolo Fibre Channel

Os campos do quadro Fibre Channel são:

- ✓ **Início:** o comprimento desse campo é de 4 bytes e indica o inicio do quadro;

- ✓ **Header:** o comprimento desse campo é de 24 bytes e contém informações de controle sobre o enlace e a transmissão de dados transportando informações para detecção de perda ou quadros fora de ordem.

O campo Header contém os seguintes registros (Figura 20)

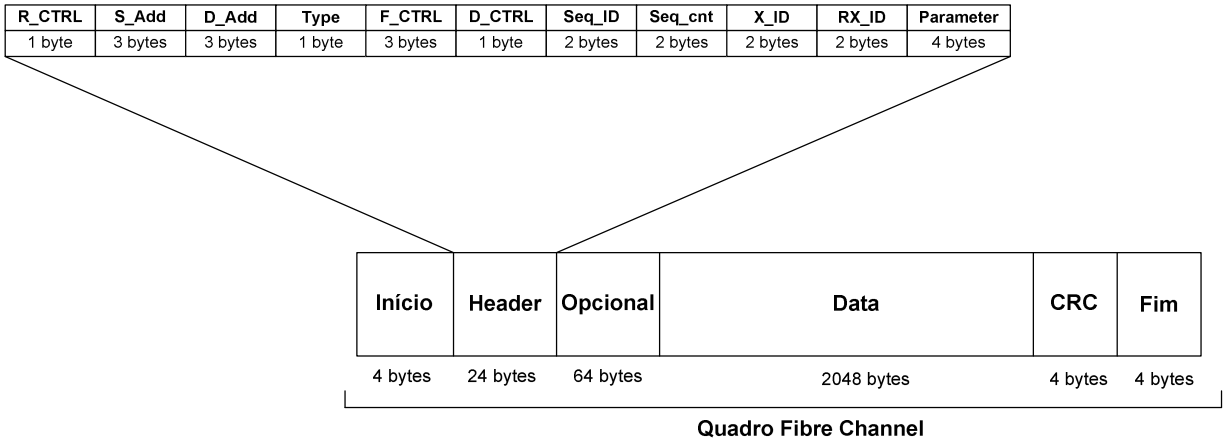


Figura 20 – Campos Header – Quadro Fibre Channel

- ✓ ***Routing Control (R_CTRL)*:** tipo de quadro, ou seja, o tipo de informação que está sendo carregado no campo Data;
- ✓ ***Source Address (S_Add)*:** endereço da porta origem que iniciou a comunicação;
- ✓ ***Destination Address (D_Add)*:** endereço da porta destino;
- ✓ ***Type*:** informações sobre o protocolo que está associado com a informação carregada no campo Data;
- ✓ ***Frame Control (F_CRTL)*:** informações que identificam a seqüência transferida;
- ✓ ***Data Field Control (D_CTRL)*:** opcional, onde esse campo pode conter informações adicionais ao cabeçalho.

- ✓ ***Sequence Identifier (Seq_ID)*:** os dados que são transferidos de uma porta a outra com um conjunto de quadros unidirecionais relacionados, ou seja, o campo Seq_ID identifica a sequência;
- ✓ ***Sequence Count (Seq_Cnt)*:** responsável por identificar a posição do quadro em determinada sequência;
- ✓ ***Exchange Identifier (X_ID)*:** um exchange é um conjunto de um ou mais seqüências não concorrentes que fazem parte de uma transmissão. O exchange pode ser unidirecional ou bidirecional;
- ✓ ***Response Exchange Identifier (RX_ID)*:** resposta a um exchange.
- ✓ ***Parameter*:** contém informações que podem variar dependendo do tipo do frame (*Data Frames* ou *Link Control*) e é usado como uma extensão do campo Data;

- ✓ **Data:** contém as mensagens SCSI CDB que estão sendo transmitidas entre as duas entidades de comunicação. O comprimento desse campo é variável podendo possuir no máximo o tamanho de 2112 bytes;

- ✓ ***Cyclic Redundancy Check (CRC)*:** utilizado para detectação de erros durante a transmissão;

- ✓ **Final –** Esse campo indica o final do quadro e seu comprimento é de 4 bytes.

Camada FC-3

A camada FC-3 é conhecida como uma camada de Serviços Comuns e refere-se a um simples ponto na pilha Fibre Channel, por onde trafegam todos os fluxos entre as entidades durante a comunicação.

Uma característica de uma entidade, por exemplo um *Initiator* ou um *Target* em uma comunicação Fibre Channel é que, essa entidade pode suportar mais do que uma

porta de comunicação envolvida em diferentes transações simultâneas [15]. As camadas FC-0, FC-1, FC-2 e FC-4 podem ser implementadas separadamente para cada porta.

Entretanto, existe somente uma implementação da camada FC-3 na qual todas as portas de comunicação das entidades Fibre Channel compartilham entre si.

A camada FC-3 atua como uma camada intermediária entre as várias portas das entidades Fibre Channel, sendo responsável pela definição dos serviços que são comuns para todas portas, provendo também serviços comuns para múltiplas portas em um switch Fibre Channel.

Os serviços oferecidos pela camada FC-3 são :

- ✓ **Multicasting –** O serviço de multicasting têm a mesma característica do serviço multicasting IP, onde esse serviço permite que a mensagem seja enviada para múltiplas portas de comunicação simultaneamente. Esse mensagem pode ser no formato broadcast, para um conjunto de portas, para um conjuntos de entidades específicas ou para um Fabric Switch completo;

- ✓ **Hunt Groups –** este serviço permite que um conjunto ou todas as portas do nó Fibre Channel possam ser relacionados com um *"alias"* identificador. O grupo de portas que compartilham um alias identificador comum é conhecido como um ***Hunt Group***. Durante a transmissão, se o transmissor origem estiver ocupado, a informação pode ser enviada para qualquer porta que pertence ao ***Hunt Group***. O principal objetivo desse serviço é fazer com que a transmissão seja mais rápida e mais eficiente na redução da latência, gerada pela espera da disponibilidade da porta origem;

- ✓ **Striping –** permite o uso de múltiplas portas de uma entidade ou de múltiplos enlaces para o envio de uma informação de unidade paralela. O uso simultâneo de múltiplas portas e enlaces objetiva aumentar a velocidade de transmissão, explorando, assim, o máximo da banda disponível;

Camada FC-4

A camada FC-4 é a camada mais alta da pilha Fibre Channel e é também conhecida como *Upper Layer Protocol Mapping Layer* (ULP). A principal responsabilidade da camada FC-4 é implementar uma definição de interfaces para permitir que aplicações e *devices drivers* possam utilizar do protocolo Fibre Channel e proporcionar o mapeamento dos protocolos da camada superior nas camadas inferiores (FC-2 e FC-3), melhorando, assim, a interoperabilidade entre aplicações. Tipicamente, os protocolos suportados pela camada FC-4 são *Small Computer Systems Interface* (SCSI), *High Performance Parallel Interface* (HIPPI) e *Intelligent Peripheral Interface* (IPI).

3.2.3 Classes de Serviços

O protocolo Fibre Channel oferece a possibilidade da utilização de três métodos de comunicação para garantir sucesso, eficiência e rapidez na entrega de diferentes tipos de tráfegos, onde esses métodos são conhecidos como Classes de Serviço [12] [14]. Cada classe de serviço relata diferentes aspectos de comunicação, tais como:

- ✓ Estabelecimento de conexão e alocação de banda;
- ✓ Entrega dos quadros entre entidades na comunicação de dados;
- ✓ Reconhecimentos (negativos e positivos) da entrega;
- ✓ Mecanismos de controle de fluxo.

Classe 1

A Classe 1 de serviço provê reconhecimento, confiabilidade e conexão orientada a transação. Suas principais características incluem conexão dedicada *circuit-switching,* através da topologia Switched entre a comunicação de duas entidades, alta vazão e a garantia da entrega ordenada de quadros.

Com a utilização dessa classe de serviço há um entrega confiável de dados sem que exista nenhum *overhead* na transmissão entre as entidades do sistema de armazenamento. Essa classe de serviço é ideal para aplicações com alto consumo de banda que necessitam que os dados sejam transferidos em fluxo contínuo, como por exemplo aplicações de *backup*, *restore* e aplicações multímidia.

Classe 2

A Classe 2 é responsável por fornecer reconhecimentos, multiplexação e serviço sem conexão. A característica principal desse serviço é prover conexões não dedicadas, onde a banda é compartilhada entre transações. Outras importantes características da Classe 2 são conexões multiplexadas, o que permite que uma porta comunique-se com mais de uma porta simultaneamente, garantia de entrega dos quadros com reconhecimento e a entrega de pacotes fora de ordem, já que a banda é compartilhada entre outras transações.

A Classe 2 é ideal para tráfego, nos quais o tempo e a ordem de entrega não são críticos, por exemplo, tipicamente para a transferências de pacotes IP e dados FTP.

Classe 3

As principais características desse serviço é a utilização de conexões não dedicadas onde a banda é compartilhada entre várias transações. A Classe 3 é também conhecida como *Datagram Service*, onde todas as possíveis perdas e corrupções dos quadros durante transmissões são tratadas nas camadas superiores de protocolos e aplicações.

Os serviços da Classe 3 são utilizados quando os Dispositivos de Armazenamento utilizam-se da topologia FC-AL para a transferência de dados.

3.3 Infraestrutura IP e Protocolo iSCSI

Desde o surgimento da tecnologia Fibre Channel no início dos anos 90, o protocolo Fibre Channel tornou-se um padrão "de facto" na implementação da arquitetura SAN (Figura 21). Este tem como principal finalidade proporcionar o transporte para as mensagens SCSI CDB com destino ao Dispositivo de Armazenamento.

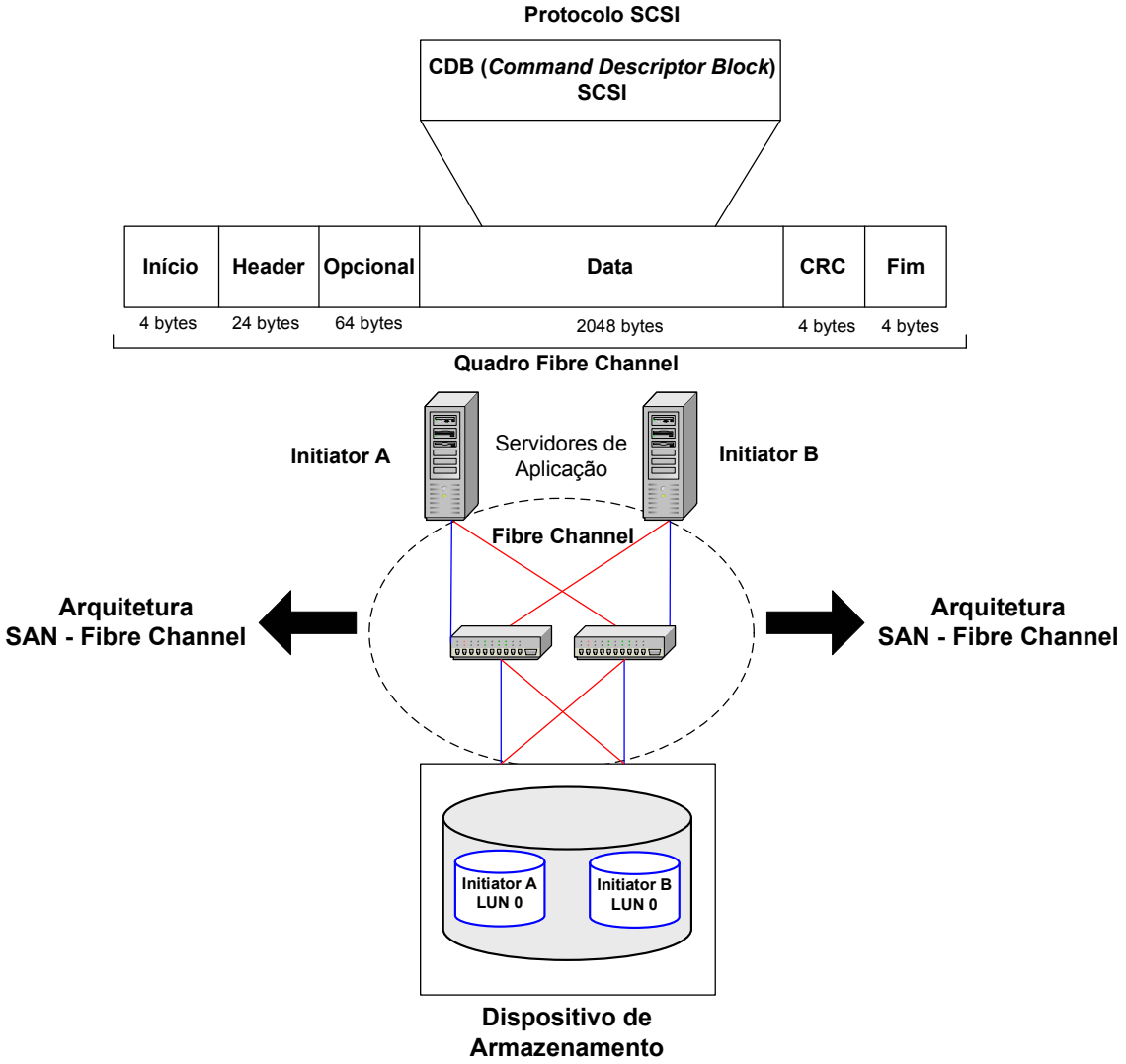


Figura 21 – Arquitetura SAN – Fibre Channel (Padrão "de facto")

Todas as corporações que necessitavam implementar uma arquitetura de armazenamento baseada no modelo SAN até então, somente possuiam como única alternativa a adoção e utilização de componentes e do protocolo Fibre Channel.

No início do ano 2000, vários estudos no IETF por um grupo chamado de IP Storage Group (IPS) foram iniciados. Um dos objetivos era o desenvolvimento e a padronização de um protocolo baseado em uma infraestrutura IP, o qual pudesse ser visto como alternativa para a criação de sistemas de armazenamento baseados no modelo SAN, anteriormente utilizando somente a infraestrutura e protocolo Fibre Channel.

No ano de 2003, foi padronizado pelo IETF [16] o protocolo iSCSI (*internet Small Computer Systems Interface*), o qual têm como principal finalidade prover o transporte ao protocolo SCSI do Servidor de Aplicação para o Dispositivo de Armazenamento utilizando de uma infraestrutura de rede TCP/IP. (Figura 22)

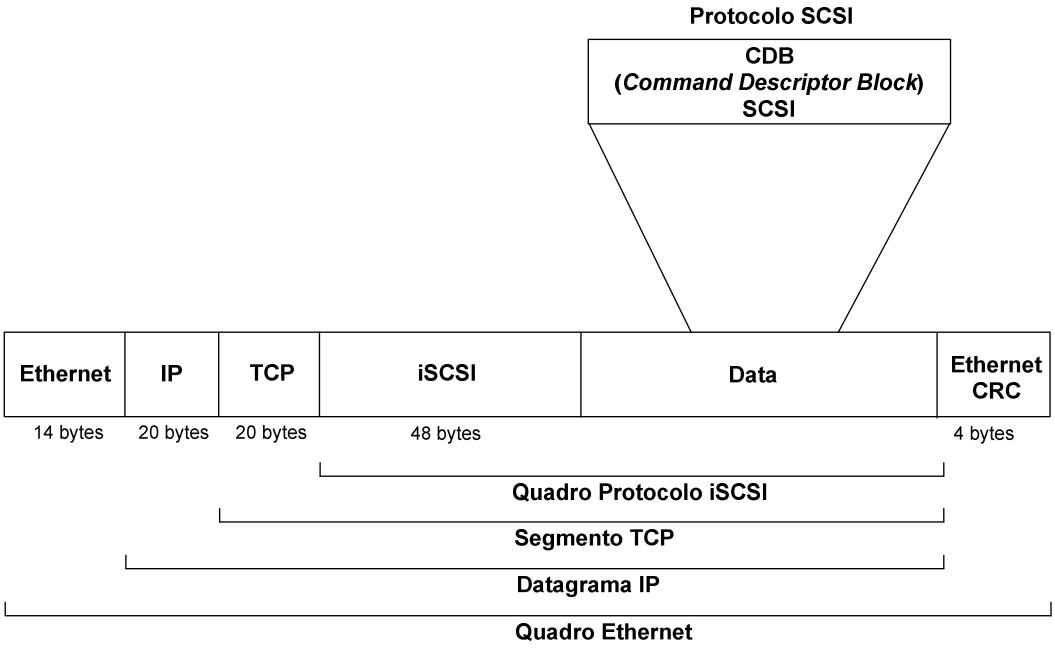


Figura 22 – Mensagem SCSI (CDB) + Quadro Protocolo iSCSI

O protocolo iSCSI é similar ao protocolo Fibre Channel. Ambos possuem o mesmo objetivo que é proporcionar o transporte ao protocolo SCSI (SCSI CDB) para Dispositivos de Armazenamento [17]. Portanto, com o desenvolvimento e a padronização do protocolo iSCSI é possível também implementar uma arquitetura SAN para acesso aos Dispositivos de Armazenamento utilizando-se de uma infraestutura TCP/IP dedicada ou não. (Figura 23)

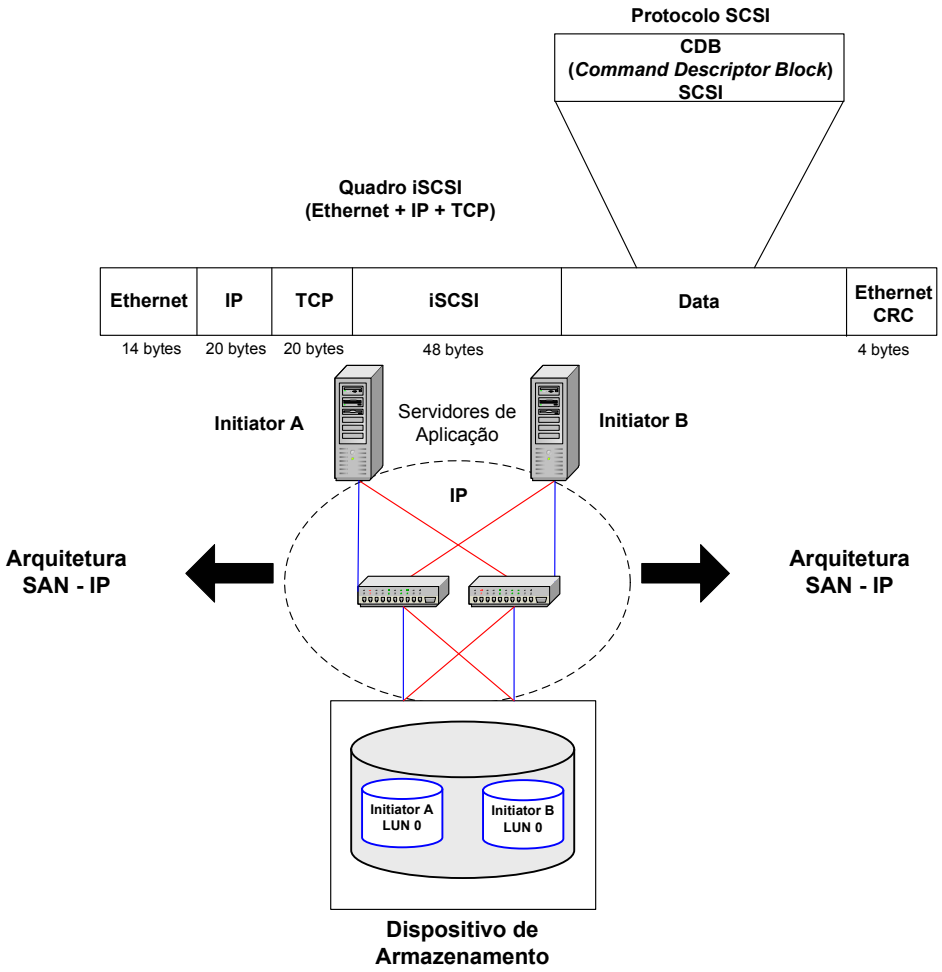


Figura 23 – Arquitetura SAN – IP (iSCSI)

Conforme pode ser visto na Figura 21 e na Figura 23, a única diferença entre uma arquitetura SAN Fibre Channel e uma arquitetura SAN IP é o meio de transporte utilizado em cada uma das arquiteturas.

O processo de comunicação entre as entidades *Initiator* e *Target* na arquitetura SAN IP é o mesmo que na arquitetura SAN Fibre Channel. O *Initiator* é tipicamente composto de aplicações de sistema que possuem a necessidade de se comunicarem com dispositivos SCSI, geralmente discos rígidos, para o armazenamento de dados através de drivers SCSI, instalados no sistema operacional do Servidor de Aplicação. A comunicação através desses drivers baseia-se em um acesso ao Dispositivo de Armazenamento utilizando um sistema de arquivos do sistema operacional. O driver iSCSI é instalado no *Initiator* e sua principal funcionalidade é prover o serviço de transporte para as mensagens SCSI CDB para o Dispositivo de Armazenamento.

Atualmente, uma implementação de uma arquitetura SAN Fibre Channel é composta basicamente de componentes Fibre Channel em uma infraestrutura de rede com velocidades de 1Gbit/s ou 2 Gbits/s, proporcionando um acesso de alta velocidade ao Dispositivo de Armazenamento.

Uma das principais motivações para a adoção do protocolo Fibre Channel para a implementação de uma arquitetura SAN foi a ausência de largura de banda comparável a largura de banda demandada do protocolo Fibre Channel, dado que tecnologias Ethernet de 10 e 100 Mbits/s ora dominante. [17] [18]

O surgimento da tecnologia Gigabit Ethernet [19] permitiu as corporações a possibilidade de utilização de sua infraestrutura para o acesso ao sistema de armazenamento de dados, pois a tecnologia Gigabit Ethernet disponibiliza um largura de banda de 1Gbit/s similar ao do protocolo Fibre Channel, permitindo assim um acesso mais rápido ao Dispositivo de Armazenamento. Outro fator muito importante foi o surgimento e o desenvolvimento de técnicas em redes de computadores que têm como principal objetivo proporcionar uma melhor qualidade no serviço no acesso e na recuperação das informação, tais como Switches IP, VLANs, Link Aggregation, QoS (*Quality of Service*) e outras técnicas [17].

O protocolo iSCSI foi desenvolvido com o objetivo principal de proporcionar implementações eficientes de hardware e software para o acesso a dispositivos SCSI-3 conectados em qualquer rede TCP/IP.

O protocolo iSCSI têm como base duas tecnologias: **iSCSI baseado em software** e **iSCSI baseado em hardware** [17]

- ✓ **iSCSI baseado em software:** (Figura 24) o driver SCSI class é instalado no sistema operacional do *Initiator* e a comunicação entre o *Initiator* e o *Target* utiliza de uma interface de rede tradicional, tipicamente Gigabit Ethernet. Todo o processamento da pilha de pacotes TCP/IP é feita pela CPU do Servidor de Aplicação. Um exemplo de um iSCSI baseado em software é o software Microsoft iSCSI *Initiator*;

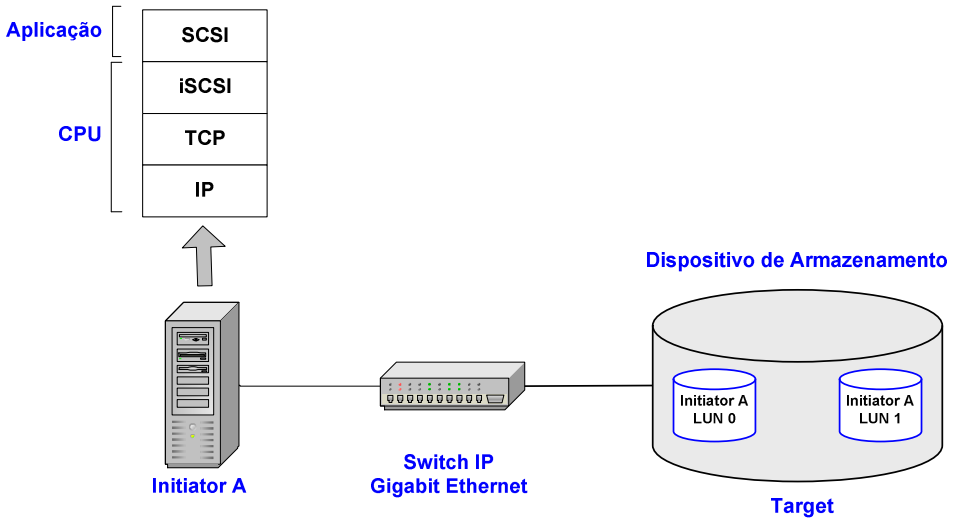


Figura 24 – iSCSI Software

- ✓ **iSCSI baseado em hardware :** Existem dois tipos de implementação da tecnologia iSCSI baseado em hardware: iSCSI HBA e iSCSI TOE

- ✓ **iSCSI HBA** (Figura 25) – o driver SCSI class é instalado no *Initiator* e a comunicação entre o *Initiator* e o *Target* utiliza uma interface especial de rede chamada de **iSCSI HBA**, a qual têm como principal objetivo efetuar o ***offload*** de processamento dos pacotes iSCSI, TCP e IP, em um hardware especial localizado na interface de rede. O funcionamento de uma interface de rede iSCSI HBA está relacionado diretamente com o funcionamento de uma interface Fibre Channel instalada no *Initiator*. Tipicamente essa interface de rede possui implementado a pilha de protocolos necessária para o funcionamento do protocolo iSCSI, não podendo ser utilizada para a comunicação com outros protocolos. Um exemplo de uma interface iSCSI HBA é a interface Intel Pro IP Storage 1000;

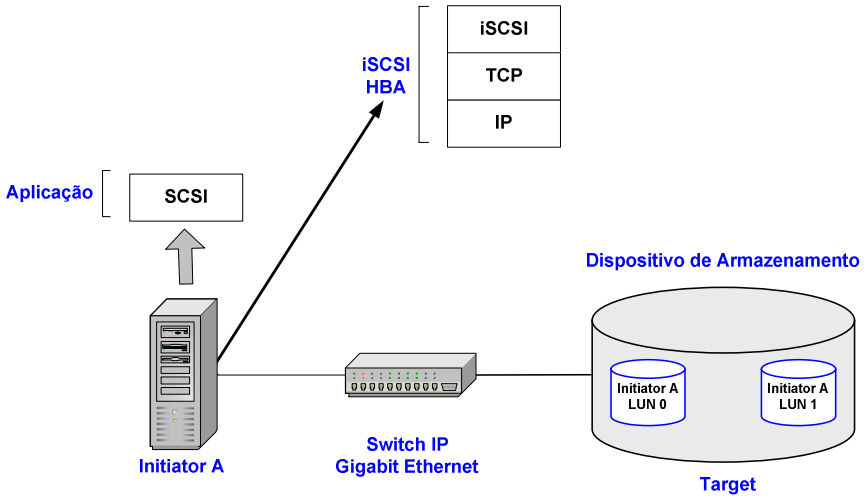


Figura 25 – iSCSI HBA

- ✓ **iSCSI TOE** (Figura 26) – O iSCSI TOE é uma interface de rede com a principal finalidade de proporcionar o ***offload*** de processamento dos pacotes TCP e IP independentemente do protocolo de aplicação utilizado, ficando o processamento deste a cargo da CPU do Servidor de Aplicação. No caso do protocolo iSCSI, o driver SCSI class é instalado no *Initiator* e a comunicação entre o *Initiator* e o *Target* utiliza de uma interface especial de rede chamada de **iSCSI TOE** (*TCP Offload Engine*) que executa o ***offload*** de processamento dos pacotes TCP e IP. Fica a cargo da CPU do Servidor de Aplicação o processamento do pacote iSCSI. Um exemplo de uma interface iSCSI TOE é a interface Alacritech SES1001;

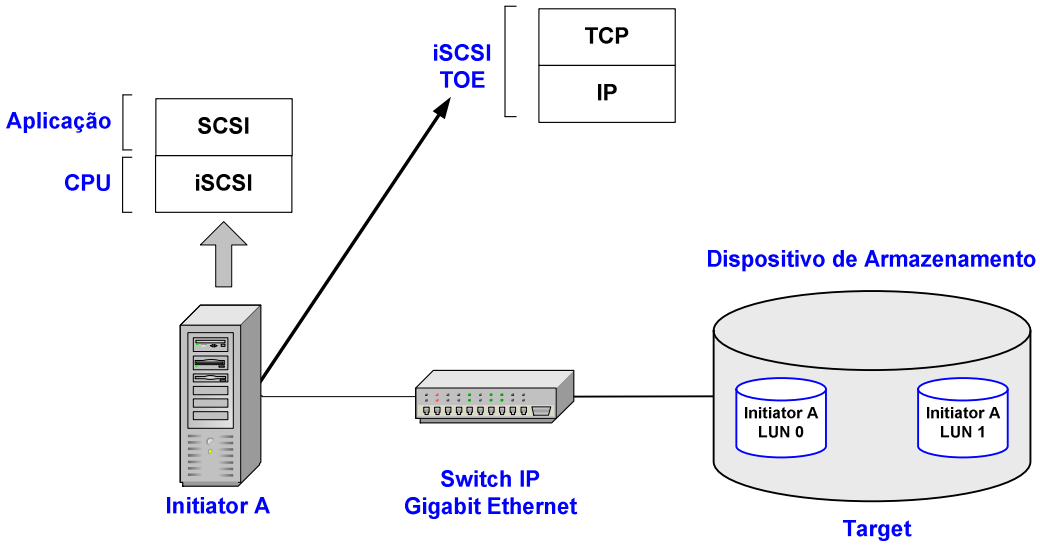


Figura 26 – iSCSI TOE

No início das especificações do protocolo iSCSI pelo grupo de desenvolvimento do IETF – IP Storage Group (IPS) foram feitos inúmeros estudos e tentativas para uma definição de qual protocolo deveria ser utilizado para prover o transporte dos comandos do protocolo iSCSI em uma infraestrutura de rede IP.

Em um modelo de rede IP, existem basicamente dois serviços básicos oferecidos pela camada de transporte:

- ✓ TCP - Serviço Orientado à Conexão ("*connection oriented service"*);
- ✓ UDP - Serviço Sem Conexão ("*conectionless service"*).

O serviço orientado à conexão é um serviço altamente confiável. Neste serviço, a camada de transporte TCP oferece vários serviços adicionais às aplicações de sistema, tais como a transferência confiável de dados, utilizando de números de sequência, reconhecimentos e temporizadores, assegurando, assim, que os dados sejam entregues ao destinatário corretamente e em ordem. A camada de transporte TCP utiliza também do controle de fluxo e do controle de congestionamento. O controle de fluxo evita que a entidade emissora sobrecarregue de dados a entidade receptora e o controle de congestionamento faz com que a entidade transmissora reduza a taxa de transmissão quando a rede está congestionada. [20]

O serviço orientado à conexão possui três fases de operação:

- ✓ Estabelecimento da Conexão;
- ✓ Transferência de Dados;
- ✓ Término da Conexão.

Em uma comunicação entre entidades na arquitetura SAN Fibre Channel ou SAN IP, tipicamente a entidade emissora é o Servidor de Aplicação – *Initiator* e a entidade receptora é o Dispositivo de Armazenamento – *Target*.

Quando uma entidade deseja enviar dados para outra entidade através de um serviço orientado à conexão, a camada de transporte da entidade *Initiator*, faz uma conexão com a camada de transporte da entidade *Target*. Uma vez que esta conexão esteja estabelecida (Estabelecimento da Conexão), as entidades trocam informações e dados (Transferência de Dados). Ao final da transferência, a camada de transporte da entidade *Initiator* pede à camada de transporte da entidade *Target* que finalize a conexão (Término da Conexão).

Apesar de todos os serviços orientados à conexão possuírem estas três fases, eles possuem duas variações, dependendo de como os dados da camada de aplicação serão separados pela camada de transporte antes de iniciar a transmissão. Essas duas variações são: Serviço Orientado a Mensagens (*"message-oriented* services") **e** Serviço de Fluxo de Bytes (*"byte-stream services*").

No serviço orientado a mensagem, a mensagem enviada pela entidade *Initiator* possui um tamanho máximo e os limites da mensagem são preservados. Por exemplo, se uma entidade *Initiator* tem 2 mensagens de 1 KByte de dados cada uma, eles são entregues como sendo dois canais de transmissão distintos; a camada de transporte não irá combiná-los em uma mensagem de 2 KBytes, nem em quatro mensagens de 500 bytes [21]. No serviço orientado a fluxo de bytes, os dados transferidos da entidade *Initiator* são vistos como uma seqüência não estruturada de bytes, os quais são transmitidos na mesma ordem na qual eles chegam. O dado não é tratado como mensagem e qualquer dado enviado à camada de transporte é inserido no final do fluxo de bytes.

No caso de utilização de um serviço sem conexão, por exemplo na utilização do protocolo UDP, existe somente a transferência de dados. Não existe nenhuma fase de estabelecimento e nem de término de conexão. Todos os dados a serem transmitidos são entregues diretamente à camada de transporte, na qual um serviço orientado a mensagem é utilizado para transferir os dados.

A transferência de dados confiável envolve quatro características diferentes e deve ser :

- ✓ **Sem Perdas (*"no loss"*):** O serviço sem perdas garante que todo o dado entregue a entidade *Target* ou a entidade *Initiator* emissor seja notificado caso haja alguma perda de dados durante a transmissão. Isto assegura que a entidade *Initiator* não tenha dúvidas se um dado foi entregue a entidade *Target* ou não. Esta característica está presente no protocolo TCP, mas não no protocolo UDP.

- ✓ **Sem Duplicações (*"no duplicates*"):** O serviço sem duplicações garante que todo o dado que tenha sido entregue à camada de transporte da entidade *Initiator* seja entregue a entidade *Target* no máximo uma vez. Qualquer pacote duplicado que chegar a entidade *Target* será descartado. Esta característica está presente apenas no protocolo TCP, mas não no UDP.

- ✓ **Ordenadas (*"ordered form*"):** O serviço de transporte ordenado garante que a entrega dos pacotes a entidade *Target* será feita na mesma ordem na qual os pacotes saíram do entidade *Initiator*. Caso algum pacote chegue fora de ordem devido a algum retardo existente na rede de transmissão, o protocolo de transporte na entidade *Target* ordena os pacotes primeiro para depois entregá-los às camadas superiores. Esta característica está presente apenas no protocolo TCP, mas não no UDP.

- ✓ **Manter a Integridade dos Dados (*"data integrity"*):** O serviço de integridade dos dados tem como função garantir que os bits dos dados entregues a entidade *Target* sejam idênticos aos bits de dados enviados pela entidade *Initiator*. Esta característica está presente tanto no protocolo TCP, quanto no protocolo UDP.

O grupo de desenvolvimento (IPS) do protocolo iSCSI decidiu que a melhor escolha seria a utilização do protocolo *Transmission Control Protocol* (TCP) como camada de transporte para o protocolo iSCSI, conforme ilustra a Figura 27. Proporcionando, assim, um modelo de conexão fim-a-fim independentemente da camada de rede utilizada. Outro fator muito importante na adoção do protocolo TCP é que esse protocolo é altamente difundido e implementado praticamente em todos os sistemas operacionais.

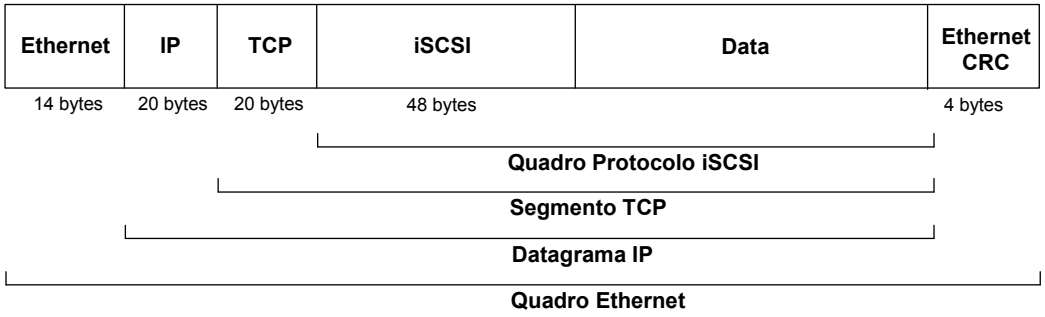


Figura 27 – Quadro Protocolo iSCSI

Se o protocolo iSCSI fosse utilizado no topo de um protocolo com características de "Serviço Sem Conexão", como por exemplo o protocolo UDP, o reconhecimento de pacotes e a sua ordernação seriam de responsabilidade da camada iSCSI, tornando, assim, a especificação, o desenvolvimento e a implementação desse protocolo mais complexa.

Como o protocolo iSCSI utiliza de protocolos desenvolvidos para a pilha TCP/IP, consequentemente, todos os protocolos utilizados na pilha TCP/IP podem ser utilizados pelo protocolo iSCSI; com por exemplo a tecnologia IPSec. O protocolo IPSec é uma das técnicas em que o protocolo iSCSI pode-se beneficiar, visto que, essa técnica proporciona um nível de segurança maior para uma comunicação segura entre *Initiator* e *Target* através de um meio não seguro, como por exemplo a Internet.

O protocolo iSCSI utiliza mensagens que são referenciados como iSCSI PDU (*Protocol Data Units*). As mensagens iSCSI consistem de um cabeçalho e de dados, onde o comprimento é especificado no cabeçalho do iSCSI PDU. A principal finalidade de um iSCSI PDU é proporcionar o transporte das mensagens referentes ao protocolo SCSI para o Dispositivo de Armazenamento (Figura 28).

As entidades *Initiator* e *Target* criam e enviam mensagens iSCSI com a finalidade de enviar requisições e respostas, onde cada mensagem iSCSI é denominada iSCSI PDU.

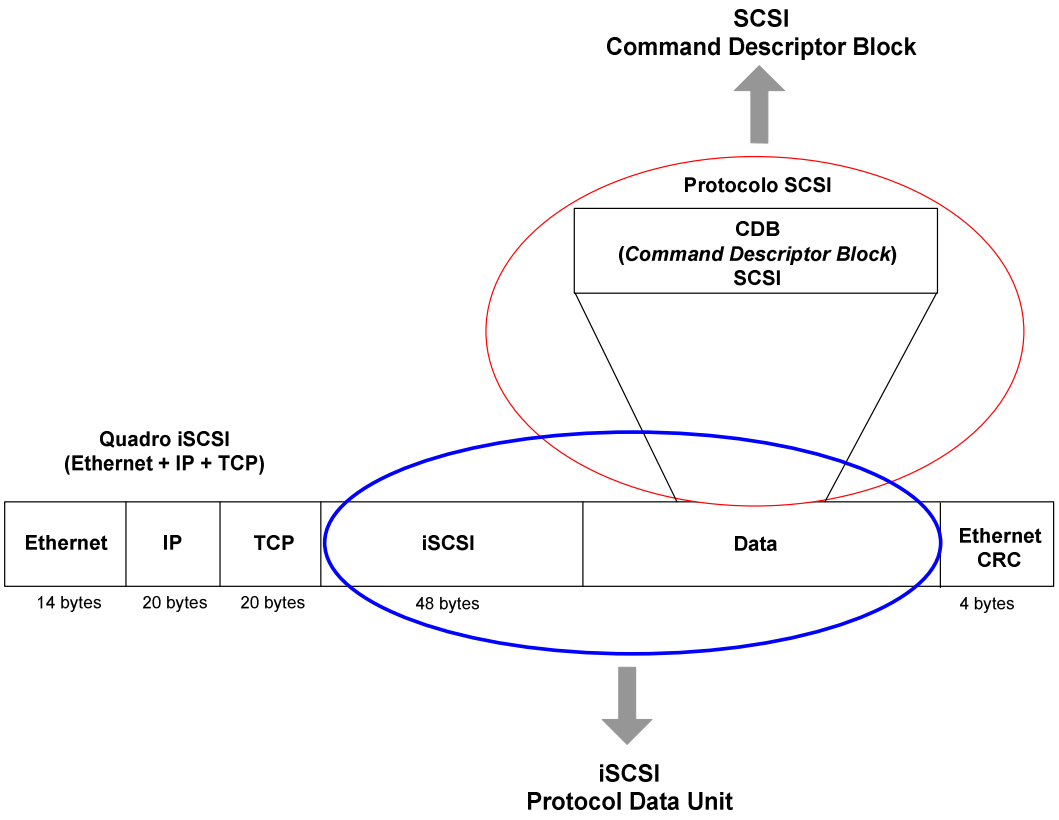


Figura 28 – iSCSI PDU e SCSI CDB

As mensagens iSCSI (Tabela 4) podem ser dos seguintes tipos :

Mensagens iSCSI	
Login Request PDU	Depois de estabelecida a conexão TCP entre o Initiator e o Target, o Initiator deve iniciar um processo de login chamado de fase de login. Portanto, o Initiator inicia a fase de login enviando uma requisição para o Target. No Target, essa requisição é chamada de Login Request PDU;
Login Response PDU	Depois que o Initiator enviou para o Target um Login Request PDU, o Target envia ao Initiator uma resposta indicando o progresso ou o término da fase de login. A resposta enviada para o Initiator é chamada de Login Response PDU;
Logout Request PDU	Essa mensagem é utilizada pelo Initiator para executar um término – logout de uma conexão iSCSI. Um Initiator pode usar o comando de logout para remover uma conexão de uma sessão ou fechar completamente a sessão;
Logout Response PDU	Essa mensagem é utilizada pelo Target com a principal finalidade de informar que as operações de finalização da conexão foram concluídas. Depois da operação de logout, a conexão TCP deve ser finalizada nas duas entidades de comunicação, tanto o Initiator quanto o Target. A identificação da conexão a ser encerrada é informada pelo campo CID (Connection ID);
NOP-Out PDU	A mensagem NOP-Out PDU é usada pelo Initiator como um “ping request” para verificar se a conexão ou sessão continuam ativas e todos os seus componentes continuam operacionais;
NOP-In PDU	Essa mensagem é enviada pela entidade Target em resposta a um NOP-Out PDU, sendo como uma resposta do tipo “ping response” para a entidade Initiator;
Ready To Transfer (R2T) PDU	Em uma operação de WRITE, o Initiator envia comandos SCSI-3 para o processamento no Target. Dependendo da quantidade de dados a ser transferido, o Target informa o estado de ready e a quantidade de dados esperada enviando uma mensagem para o Initiator do tipo Ready To Transfer (R2T);

Mensagens iSCSI	
Reject PDU	A mensagem Reject PDU é utilizada pelo Target quando o mesmo detecta um condição de erro do protocolo iSCSI, tais como protocolo não suportado, versão não suportada, opções não suportadas e outros;
SCSI Data-In PDU	A mensagem SCSI Data_In PDU é enviada pelo Target com destino ao Initiator e é utilizada em operações READ;
SCSI Data-Out PDU	A mensagem SCSI Data_Out PDU é enviada pelo Initiator com destino ao Target e é utilizada em operações WRITE. Dependendo da quantidade de dados a ser transferido do Initiator para o Target, esse utiliza somente uma mensagem do tipo SCSI Command PDU contendo todos os comandos e os dados. No entanto, se a quantidade de dados a ser enviada for maior que o parâmetro MaxRecvDataSegmentLength, o Initiator envia uma mensagem do tipo SCSI Command Request PDU atendendo o parâmetro acima e depois envia as mensagens do tipo SCSI Data-Out PDU contendo os dados;
SNACK Request PDU	A mensagem do tipo SNACK Request PDU (Selective Negative Acknowledgment) é uma requisição de retransmissão de mensagens com status, dados e R2T PDUs. Essa mensagem também é conhecida como Serial Number Acknowledgment para o segmento de dados;

Tabela 4 – Mensagens iSCSI

De um modo geral, as mensagens iSCSI mais utilizadas são ***SCSI Command Request PDU*** que tipicamente é enviado pela entidade *Initiator* com destino ao *Target* e o ***SCSI Command Response PDU*** que executa o processo inverso a mensagem SCSI Command PDU, sendo enviado pela entidade *Target* com destino ao *Initiator*.

Os dados transportados pelo protocolo iSCSI podem ser de três diferentes tipos:

iSCSI – Transferência de Dados	
Immediate Data	Os dados são enviados em um segmento de dados do tipo SCSI_REQUEST PDU juntamente com os comandos SCSI CDB;
Unsolicited Data	Os dados são enviados em um segmento de dados do tipo SCSI DATA_OUT PDU, no qual o Initiator pode enviar um segmento de dados sem a necessidade de espera da requisição de envio da primeira confirmação R2T (Ready To Transfer) pelo Target;
Solicitied Data	Os dados são enviados em um segmento de dados do tipo DATA_OUT PDU em resposta a uma requisição pelo Target na chamada de uma confirmação R2T (Ready To Transfer);

Tabela 5 – Transferência de dados – iSCSI

As entidades *Initiator* e *Target* negociam na fase de login três importantes parâmetros que são: MaxRecvDataSegmentLenght, FirstBurstLenght e InitialR2T, os quais são responsáveis pela quantidade de dados permitida durante a transmissão e recepção entre as entidades.

Por exemplo, em uma operação de WRITE com a quantidade de dados de 96KB entre as entidades *Initiator* e *Target*, os parâmetros na fase de login foram:

- ✓ **MaxRecvDataSegmentLength = 16KB**

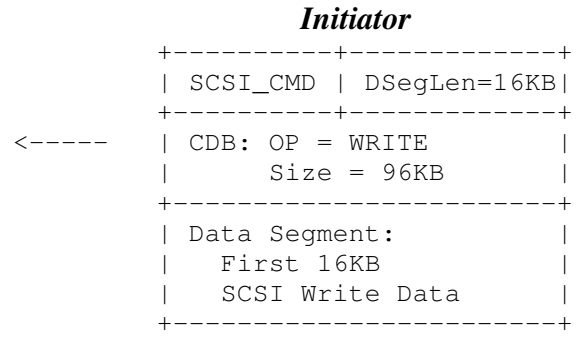
  Esse parâmetro especifica a quantidade máxima a ser transmitida no segmento de dados SCSI;

- ✓ **ImmediateData = yes**

  Esse parâmetro especifica que o *Initiator* pode enviar o segmento de dados como ImmediateData, onde o segmento de dados é enviado na mensagem SCSI DATA_IN PDU que contém um campo chamado ImmediateData para transporte dos dados que contém os comandos SCSI-3, os quais serão executados pelo *Target* na Logical Unit apropriada;
- ✓ **InitialR2T = yes**

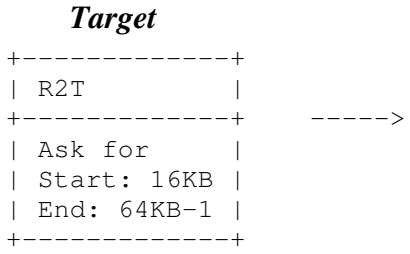
  Esse parâmetro especifica que o *Initiator* depois de enviar o pacote SCSI DATA_IN PDU para o *Target*, o mesmo deve aguardar uma requisição do tipo R2T enviada pelo *Target* em conformidade com o parâmetro FirstBurstLenght;
- ✓ **FirstBurstLenght = 64KB** especifica a quantidade de dados do tipo SCSI DATA_OUT PDU que o *Initiator* pode enviar para o *Target* sem a necessidade de aguardar uma requisição de READY pelo *Target* que é informado através do pacote R2T;

Inicialmente, a entidade *Initiator* irá construir uma mensagem do tipo SCSI DATA_IN PDU a ser enviada para a entidade *Target*, informando-a através do pacote SCSI CDB a operação a ser realizada e a quantidade total de dados é de 96KB. De acordo com o parâmetro MaxRecvSegmentLenght, a quantidade de dados a ser enviada é de 16KB.

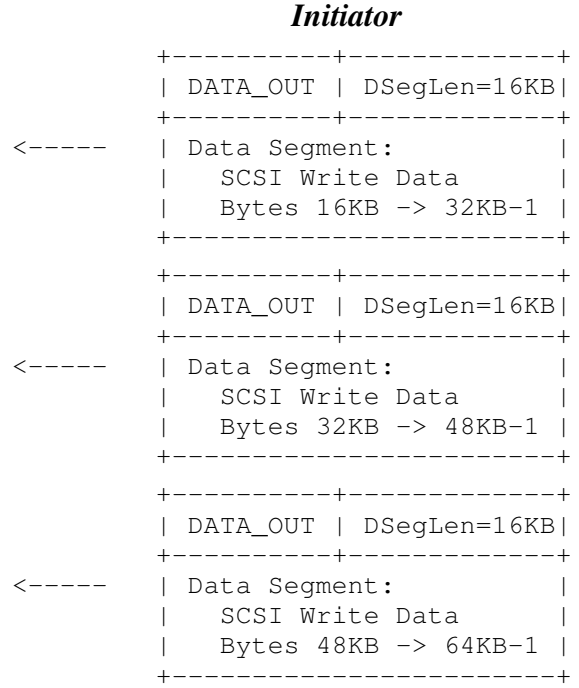


Depois que a entidade *Target* recebeu a mensagem SCSI DATA_IN PDU, a entidade *Target* então de acordo com o parâmetro InitialR2T = yes, envia uma requisição R2T para a entidade *Initiator*, informando-o a nova quantidade de dados a ser transferida. Essa quantidade deve estar de acordo com o parâmetro MaxBurstLenght que

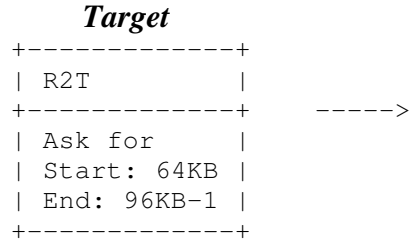
normalmente possui o mesmo valor que o parâmetro FirstBurstLenght = 64KB. A entidade *Target* pode especificar uma quantidade de dados para o R2T menor do que a quantidade especificada no parâmetro MaxBurstLenght, a qual nesse exemplo foi especificado uma quantidade de 48KB dados que é menor que a quantidade de dados do parâmetro MaxBurstLenght.



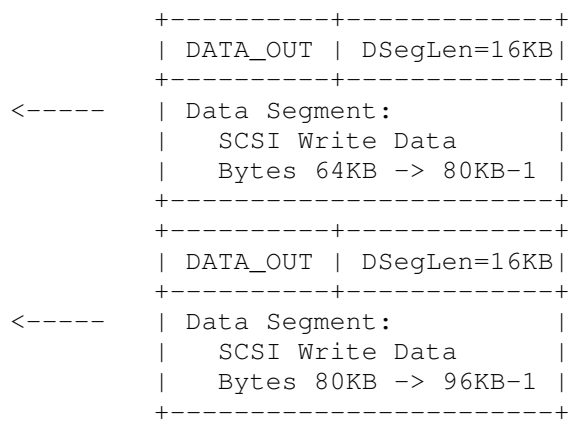
Depois que a entidade *Initiator* recebeu a mensagem R2T, então esse pode enviar a quantidade de dados solicitada pela entidade *Target*, atendendo os mesmos requerimentos do parâmetro MaxRecvDataSegmentLength que especifica a quantidade máxima de dados que pode ser recebida pela entidade *Target* em cada mensagem SCSI DATA_IN ou SCSI DATA_OUT PDUs.



Depois que a entidade *Target* recebeu as três mensagens perfazendo um total de (16KB+16KB+16KB) = 48KB, a entidade *Target* novamente envia para a entidade *Initiator* outra mensagem do tipo READY R2T especificando uma nova quantidade de dados esperada.



Logo após o recebimento da mensagem R2T pela entidade *Initiator*, o mesmo pode enviar a quantidade de dados solicitada pela entidade *Target*, novamente de acordo com o parâmetro MaxRecvDataSegmentLength.

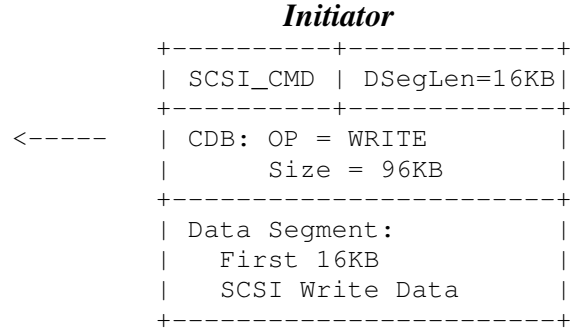


Logo após o recebimento das mensagens enviadas pela entidade *Initiator*, e perfazendo um total de 96KB de dados, a entidade *Target* envia para a entidade *Initiator* uma mensagem informando o status do processamento do pacote SCSI na Logical Unit no Dispositivo de Armazenamento.

Por exemplo, a mesma operação de WRITE com a quantidade de dados de 96KB entre as entidades *Initiator* e *Target*, com o parâmetro InitialR2T = no

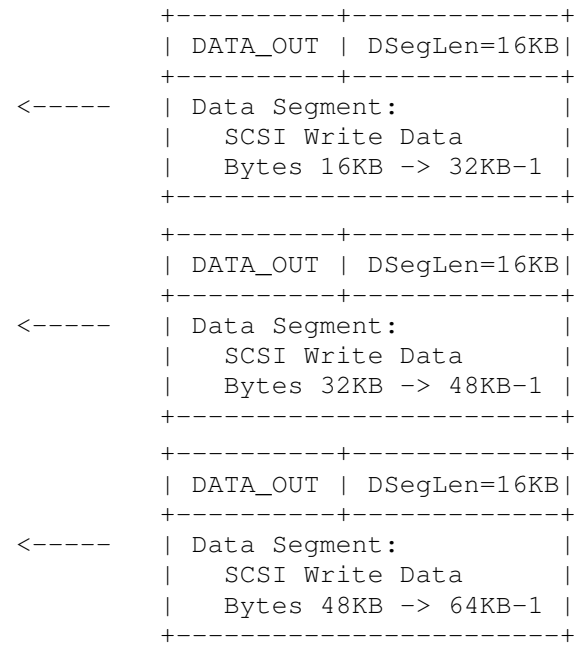
A definição do parâmetro InitialR2T = no, possibilita que a entidade *Initiator* possa enviar um mensagem iSCSI, de acordo com os parâmetros MaxRecvDataSegment Length e FirstBurstLenght sem a necessidade de esperar por uma mensagem iSCSI do tipo R2T enviada pela entidade *Target*.

Da mesma forma que no exemplo anterior, a entidade *Initiator* irá construir uma mensagem do tipo SCSI DATA_IN PDU a ser enviada para a entidade *Target*, informando-a através do pacote SCSI CDB a operação a ser realizada e a quantidade total de dados é de 96KB. De acordo com o parâmetro MaxRecvSegmentLenght, a quantidade de dados a ser enviada é de 16KB.



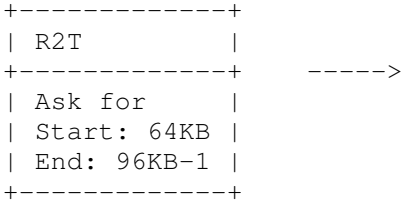
Logo após o envio da mensagem inicial SCSI DATA_IN PDU, a entidade *Initiator* pode enviar mais três mensagens do tipo SCSI DATA_OUT PDU, de acordo com os parâmetro InitialR2T = no e o parâmetro FirstBurstLenght = 64KB. Cada mensagem SCSI DATA_IN PDU está de acordo com o parâmetro MaxRecvDataSegmentLenght que especifica um tamanho de dados em cada mensagem de 16KB.

Initiator

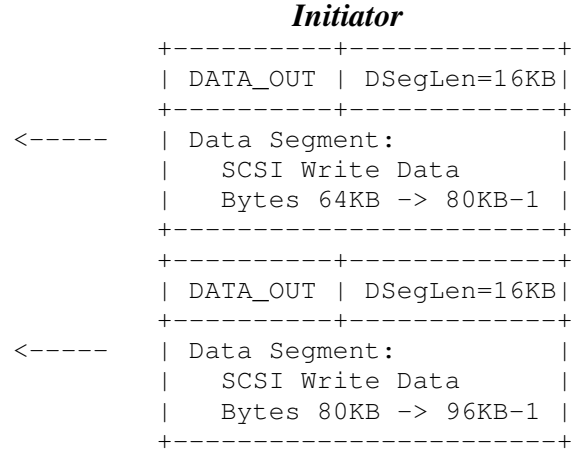


Logo após o recebimento das quatro mensagens iSCSI enviadas pela entidade *Initiator*, a entidade *Target* envia uma mensagem do tipo R2T especificando uma nova quantidade de dados esperada. Essa quantidade deve estar de acordo com o parâmetro MaxBurstLenght que normalmente possui o mesmo valor que o parâmetro FirstBurstLenght = 64KB. A entidade *Target* pode especificar uma quantidade de dados para o R2T menor do que a quantidade especificada no parâmetro MaxBurstLenght, a qual nesse exemplo foi especificado uma quantidade de 48KB dados que é menor que a quantidade de dados do parâmetro MaxBurstLenght.

Target



Logo ápos o recebimento da mensagem iSCSI R2T, a entidade *Initiator* pode enviar mais duas mensagens do tipo SCSI DATA_OUT PDU, de acordo com o parâmetro InitialR2T=no e o parâmetro FirstBurstLenght = 64KB, onde cada mensagem SCSI DATA_IN PDU deve estar de acordo com o parâmetro MaxRecvDataSegmentLenght que especifica um tamanho de dados em cada mensagem de 16KB.



Logo após o recebimento das mensagens enviadas pela entidade *Initiator*, e perfazendo um total de 96KB de dados, a entidade *Target* envia para a entidade *Initiator* uma mensagem informando o status do processamento do pacote SCSI na Logical Unit no Dispositivo de Armazenamento.

3.3.1 Etapas de Comunicação do Protocolo iSCSI

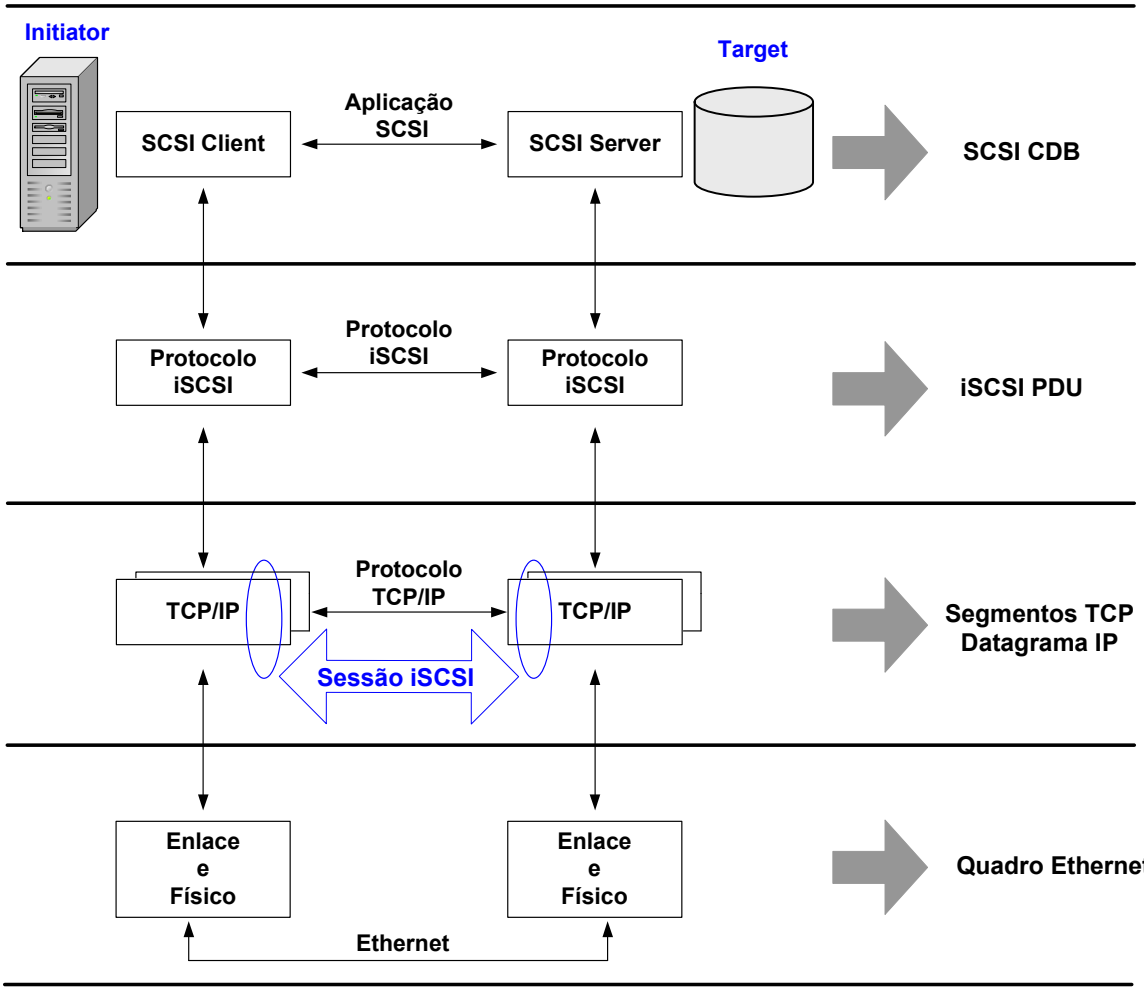


Figura 29 – Camadas de comunicação do protocolo iSCSI

A Figura 29 ilustra a estrutura de camadas na utilização do protocolo iSCSI. A primeira camada pode ser definida como a camada de aplicação, na qual se situam as aplicações e o resultado final da interconexão entre o *Initiator* e o *Target.*

A aplicação localizada no *Initiator* envia uma requisição de E/S para o kernel do sistema operacional, o qual direciona para o *device driver* SCSI class apropriado, onde geralmente essas operações de E/S são feitas indiretamente através do sistema de arquivos utilizado pelo sistema operacional.

O *device driver* SCSI class constrói uma requisição de E/S utilizando do comando SCSI apropriado, que é armazenado em uma mensagem SCSI CDB, enviada para a Logical Unit (LU) requerida.

Após a construção do pacote SCSI CDB, torna-se necessário efetuar o transporte do mesmo. Utiliza-se do protocolo iSCSI para desempenhar tal funcão. O device driver iSCSI localizado na camada iSCSI no *Initiator* armazena o pacote SCSI CDB em uma mensagem iSCSI, juntamente com a informação do LUN requerido e o envia para a entidade iSCSI destino (*Target*) utilizando-se de uma sessão iSCSI feita de uma ou mais conexões TCP/IP.

A camada TCP/IP têm como principal objetivo proporcionar um transporte confiável e seguro ao protocolo iSCSI, onde as mensagens iSCSI podem ser divididas em mais de um segmento TCP e cada segmento TCP é encapsulado em um datagrama IP, que por sua vez, é encapsulado pela camada de enlace tipicamente Ethernet.

A camada iSCSI localizada na entidade *Target* extrai a mensagem SCSI CDB que está sendo transportado pelo protocolo iSCSI e o entrega para a camada SCSI Server indicando o LUN apropriado, e assim, a camada SCSI Server localizada na entidade *Target* processa os comandos contidos no SCSI CDB na *Logical Unit* requerida.

3.3.2 Sessões iSCSI

Uma sessão iSCSI pode ser definida como um enlace lógico para o transporte das mensagens iSCSI entre as entidades *Initiator* e *Target* em um sistema de armazenamento. Uma sessão iSCSI constitui de pelo menos uma conexão TCP/IP da entidade *Initiator* para a entidade *Target*.

Em geral, esse conceito é muito simples, porém não há um método nativo do protocolo SCSI para o gerenciamento e a implementação do envio de comandos e dados através de múltiplos enlaces. Geralmente, cada fabricante desenvolveu um driver para executar tal funcionalidade, o qual é chamado de ***Initiator Wedge Driver*** que tem como finalidade balancear a carga de trabalho entre múltiplos enlaces. Essa funcionalidade é largamente utilizada em implementações da arquitetura SAN baseadas no protocolo Fibre Channel.

O protocolo iSCSI apresenta, também, funcionalidades adicionais e complexas em seu protocolo para a implementação e utilização de múltiplas sessões. Um exemplo disso é o conceito chamado de **MC/S (*Multiple Connections per Session*)** existente no protocolo iSCSI, o qual têm como principal característica a capacidade de fazer com que várias sessões TCP/IP sejam vistas como uma única conexão entre as entidades *Initiator* e *Target*, permitindo, assim, com que comandos e dados fossem transportados, por exemplo, através de vários enlaces diferentes, o que permite balanceamento de carga, alta disponibilidade e maior desempenho entre as entidades envolvidas na comunicação iSCSI (Figura 30).

A técnica MC/S torna o desenvolvimento, a implementação e o gerenciamento do protocolo iSCSI mais difícil, todo o fluxo de dados irá depender de todas as múltiplas conexões TCP/IP entregarem os fluxos de dados em ordem e sem erro a camada iSCSI na entidade *Target*.

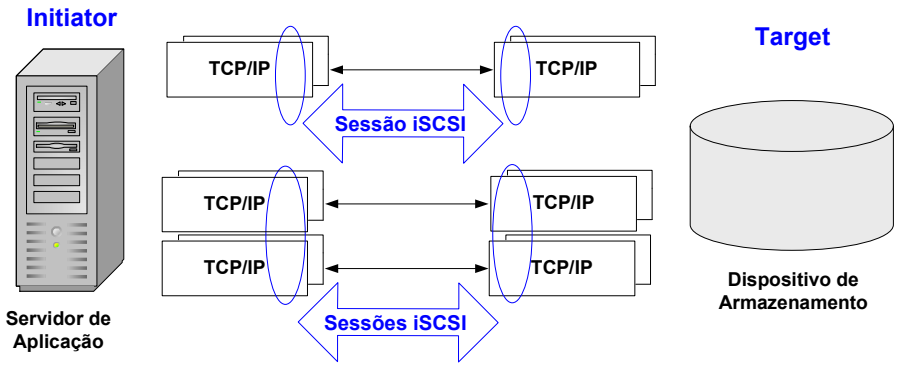


Figura 30 – MC/S (Múltiplas Sessões do protocolo iSCSI)

3.3.3 Header Digest e Data Digest

O protocolo iSCSI também permite a possibilidade de utilização de um campo de checksum para a verificação de erros em suas mensagens. Essa verificação pode ser executada no campo Header e Data das mensagens iSCSI chamados Header Digest e Data Digest respectivamente.

Portanto, toda a implementação do protocolo iSCSI pode suportar um checksum de 32 bits (CRC-32), permitindo assim, detecção de erros na camada iSCSI em transmissões entre entidades *Initiator* e *Target. A* utilização desse recurso é opcional sendo requerido ou não na fase de login pela entidade *Initiator*.

Uma desvantagem da utilização de Header Digests e Data Digests através do protocolo iSCSI é o aumento de processamento adicional na CPU da entidade *Initiator* e/ou *Target* para cálculo de checksum. Esse aumento pode comprometer o desempenho necessário para a aplicação de sistema. Com a utilização do protocolo iSCSI baseado em hardware, todo o processamento adicional anteriormente executado pela CPU da entidade *Initiator* e/ou *Target* passa a ser executado em um hardware especial localizado na interface de rede, implementando, assim, um ***offload*** de processamento dos pacotes iSCSI, TCP e IP.

Outra possibilidade é o fato de permitir que elementos de rede possam dar suporte a camada de aplicação iSCSI. Desta forma, os próprios elementos de rede executam o serviço de análise de checksum do protocolo iSCSI, analisando possíveis mensagens iSCSI corrompidas. Caso seja encontrado qualquer tipo de corrupção de dados em uma mensagem iSCSI, o próprio elemento de rede descartará automaticamente a mensagem iSCSI e informará a entidade *Initiator* a necessidade de uma retransmissão da mensagem corrompida.

3.3.4 Recuperação de Erros

A definição dos mecanismos de recuperação de erros do protocolo iSCSI [22] consideram os possíveis problemas que podem ocorrer em decorrência da não confiabilidade das redes IP. A infraestrutura das redes locais foi significativamente aprimorada nos últimos anos, portanto, não sendo tão necessário o desenvolvimento de um nível especial de recuperação de erros para protocolo iSCSI.

Problemas podem ocorrer em transferência de dados em longas distâncias, como por exemplo na utilização na Internet. Os possíveis problemas são :

- ✓ Conexões TCP ocasionalmente podem falhar e devem ser restabelecidas;

- ✓ Mecanismo de detecção de erros do protocolo TCP possui um checksum de 16 bits e pode não ser suficientemente confiável para aplicações em armazenamento de dados.

O protocolo iSCSI define mecanismos para tratamento desses erros, sem a necessidade de falha para um correspondente comando SCSI. Implementações simples do protocolo iSCSI podem optar por não implementar um nível avançado de recuperação.

- ***Session Recovery –*** é considerado o tipo mais básico de recuperação pelo protocolo iSCSI. No procedimento de ***Session Recovery***, quando um erro é detectado a sessão inteira iSCSI é terminada. Todas as conexões TCP conectando a entidade *Initiator* com a entidade *Target* são fechadas e todos os comandos SCSI pendentes são completados com um status de erro apropriado.

  Uma nova sessão é então estabelecida entre as entidades *Initiator* e *Target* e o nível superior do protocolo SCSI pode então refazer todos os comandos nas quais ocorreram falhas na execução. Estes novos comandos são transportados através de uma nova conexão TCP em uma nova sessão iSCSI.

  Portanto, o procedimento de ***Session Recovery*** passa a recuperação dos comandos com falha para o nível superior do protocolo SCSI.

- ***Digest Failure Recovery*** – O protocolo TCP define um checksum de 16-bits para detecção de pacotes corrompidos. Para aplicações em sistemas de armazenamento de dados, um checksum de 16-bits pode não ser suficiente para uma completa detecção de erros. O checksum do protocolo TCP é utilizado em todo pacote TCP, mas isso não protege o pacote iSCSI no momento da geração do pacote antes de ser entregue a camada TCP, podendo o mesmo ser entregue a camada iSCSI com erros. Portanto, o protocolo iSCSI define um checksum de 32-bits CRC digest em pacotes iSCSI com a finalidade de detectar dados corrompidos em uma conexão fim a fim. Caso o *device driver* iSCSI detecte a chegada de dados com um inválido CRC, então o dado é rejeitado. A entidade *Initiator* é notificada a fim de reenviar o pacote rejeitado.

- ✓ ***Connection Recovery*** – No caso de falha de uma conexão TCP, os comandos SCSI que foram entregues a entidade *Target* não tiverem sido confirmados para a entidade *Initiator*, o comando SCSI fica pendente até que a camada SCSI aborte efetivamente o comando. Depois de detectado a falha na conexão TCP, o *device driver* iSCSI pode imediatamente completar o comando pendente com uma indicação de um erro ou tentar transferir o comando SCSI para uma outra conexão TCP.

Para a maioria das aplicações que executam requisições de E/S para discos rígidos utilizando uma arquitetura local SAN IP, é suficiente a implementação do mecanismo ***Session Recovery***, pois na existência de um erro, a sessão iSCSI é finalizada e todos comandos SCSI pendentes são completados com uma indicação de erro apropriado. Assim a camada SCSI irá atuar nos erros e novos comandos serão transportados sobre conexões TCP em uma nova sessão iSCSI.

Para algumas aplicações que executam requisições de E/S para unidades de fita, torna-se muito oneroso a reinicialização dos comandos. Outro impacto seria executar uma aplicação de backup utilizando um enlace baseado em longas distâncias tornando muito comum a ocorrência de problemas de comunicação. Nesse caso, torna-se mais apropriado a implementação de um nível mais complexo de recuperação podendo ser o mecanismo ***Digest Failure Recovery*** ou o mecanismo ***Connection Recovery***.

Capítulo 4

4.1 Características da utilização do protocolo iSCSI em relação ao protocolo Fibre Channel

Como mencionado anteriormente, antes do desenvolvimento e da padronização do protocolo iSCSI, toda implementação de uma arquitetura SAN era baseada na infraestrutura e no protocolo Fibre Channel, tornando-o um "padrão de facto" nas implementações de um infraestrutura *Storage Area Network*, a qual tipicamente é denominada de SAN FCP. O surgimento do protocolo iSCSI representou uma nova alternativa de transporte para o protocolo SCSI, pois possibilitou a implementação de uma arquitetura SAN, agora baseada em uma infraestrutura TCP/IP e no protocolo iSCSI (Figura 31).

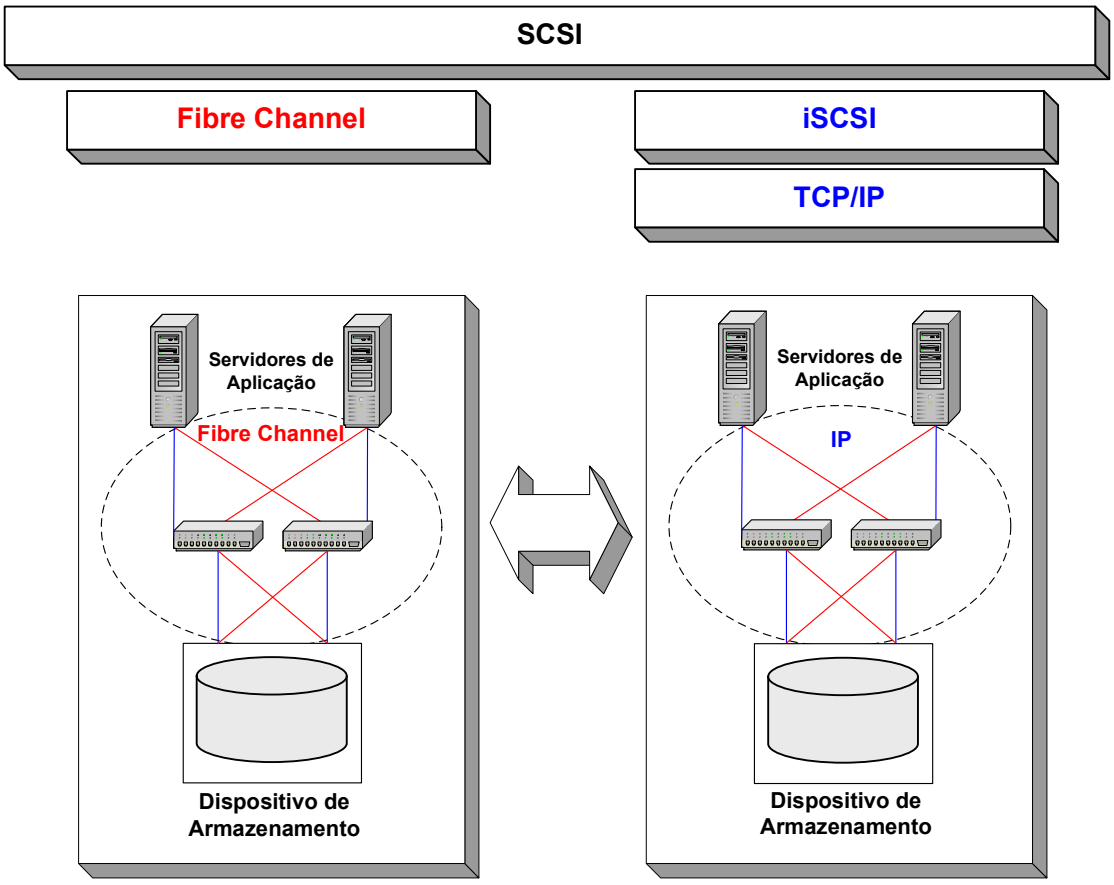


Figura 31 – Arquiteturas SAN FCP e SAN IP

Uma infraestrutura SAN pode ser baseada em componentes e protocolo Fibre Channel (SAN FCP) bem como em componentes IP e no protocolo iSCSI (SAN IP).

Como mencionado anteriormente, a arquitetura DAS representa 42% do mercado de armazenamento de dados. Uma das razões que algumas corporações não fizeram uma transição de implementação baseada na arquitetura DAS para uma arquitetura SAN é o alto custo de implementação e de manutenção dos componentes Fibre Channel. [24]

As corporações que necessitam de uma implementação de uma arquitetura SAN, possuem duas alternativas: uma arquitetura SAN FCP baseada em componentes e no protocolo Fibre Channel ou uma arquitetura SAN IP baseada em componentes TCP/IP e no protocolo iSCSI.

Os componentes básicos de uma SAN FCP são:

- ✓ **Interfaces Fibre Channel:** também conhecidas como adaptadores Fibre Channel HBA (*Host Bus Adapters*). São instaladas e configuradas no Servidor de Aplicação (*Initiator*), sendo responsáveis pela conexão física como o Dispositivo de Armazenamento através de um elemento de rede, como por exemplo um switch Fibre Channel. O adaptador Fibre Channel possui interface óptica com velocidades de 1Gbit/s ou 2Gbits/s;

- ✓ **Switches Fibre Channel:** são elementos de rede responsáveis pelo compartilhamento de recursos entre as entidades *Initiator* e *Target*. Da mesma forma que as interfaces Fibre Channel, os switches Fibre Channel possuem interfaces ópticas com velocidades de 1Gbit/s ou 2Gbits/s.

Os componentes básicos de uma SAN IP são:

- ✓ **Interfaces Gigabit Ethernet:** a implementação da arquitetura SAN IP pode ser baseada em duas tecnologias: iSCSI baseado em Software que utiliza as interfaces tradicionais Gigabit Ethernet e iSCSI baseado em hardware que utiliza de interfaces especiais Gigabit Ethernet – iSCSI HBA ou iSCSI TOE. As Interfaces Gigabit Ethernet são responsáveis pela conexão física como o Dispositivo de Armazenamento através de um elemento de rede, por exemplo, um switch de rede IP. As Interfaces Gigabit Ethernet podem possuir adaptadores ópticos ou adaptadores de cobre com velocidades a 1Gbit/s;

- ✓ **Switches IP:** são elementos de rede responsáveis pelo compartilhamento de recursos entre as entidades *Initiator* e *Target.* Os switches IP são equipamentos de rede tradicionais utilizados em redes TCP/IP e podem possuir interfaces ópticas e/ou interfaces de cobre com velocidades de 1Gbit/s;

Em estudos realizados no ano de 2003 [26] [27], a implementação de uma arquitetura de armazenamento de dados baseada no protocolo iSCSI (SAN IP) possui um grande diferencial em relação a custos de aquisição e manutenção de componentes comparado com a arquitetura SAN FCP.

Componentes	Fibre Channel	iSCSI
HBA	$1000	$400
Switch (porta)	$1000	$70-250
Gigabit Ethernet	0	$50-100
TOTAL	**$2000**	**$520-750**

Tabela 6 – Custo de Componentes Fibre Channel e iSCSI

Conforme pode ser visto na Tabela 6, um componente Fibre Channel HBA possui aproximadamente um valor de US$ 1000,00 por unidade, enquanto um componente tradicional Gigabit Ethernet possui aproximadamente um valor máximo por unidade de US$ 100,00. A adoção de uma infraestrutura SAN IP baseada em componentes tradicionais Gigabit Ethernet reduz o custo de 90% em relação aos componentes HBA Fibre Channel. A utilização de uma arquitetura SAN IP baseada em componentes iSCSI HBA, representa uma redução de 60% em relação aos componentes HBA Fibre Channel.

Na Tabela 6 pode ser verificado também que um elemento de rede Fibre Channel (Switch) possui aproximadamente um custo por porta de US$ 1000,00; enquanto um elemento de rede IP (Switch) possui aproximandamente um custo máximo por porta de US$ 250,00, existindo uma redução de 75% em relação ao elemento de rede Fibre Channel. Nota-se, também, que a adoção de uma arquitetura SAN IP, em relação a uma arquitetura SAN FC, haverá uma redução de aproximadamente de 60% nos custos de aquisição de componentes essenciais para a implementação da arquitetura SAN.

Um dos principais motivadores para o desenvolvimento do protocolo iSCSI é a possibilidade de redução significativa dos custos de aquisição de seus componentes para as corporações, pois o protocolo iSCSI baseia-se na utilização de componentes TCP/IP. Os custos destes componentes em comparação aos componentes Fibre Channel representam uma expressiva redução.

Outro grande motivador da utilização e do desenvolvimento do protocolo iSCSI deve-se ao fato da utilização do conhecimento adquirido pela abrangência do protocolo TCP/IP, que possui uma grande difusão e utilização em universidades e corporações.

Esse conhecimento adquirido é essencial para um melhor entendimento, utilização e implementação do protocolo iSCSI na arquitetura SAN IP e o conhecimento adquirido em relação ao protocolo Fibre Channel é limitado quando comparado ao conhecimento adquirido em relação ao protocolo TCP/IP devido a sua difusão e utilização na Internet.

Outro fator muito importante, ilustrado na Figura 32, é a possibilidade da utilização da Internet como infraestrutura de transporte para o protocolo iSCSI, ou seja com a utilização do protocolo iSCSI um arquitetura SAN IP não possui limitações quanto à distância em relação ao Servidor de Aplicação (*Initiator*) e Dispositivo de Armazenamento (*Target*), diferindo da limitação de distância de uma arquitetura Fibre Channel Switched que é de 10Km.

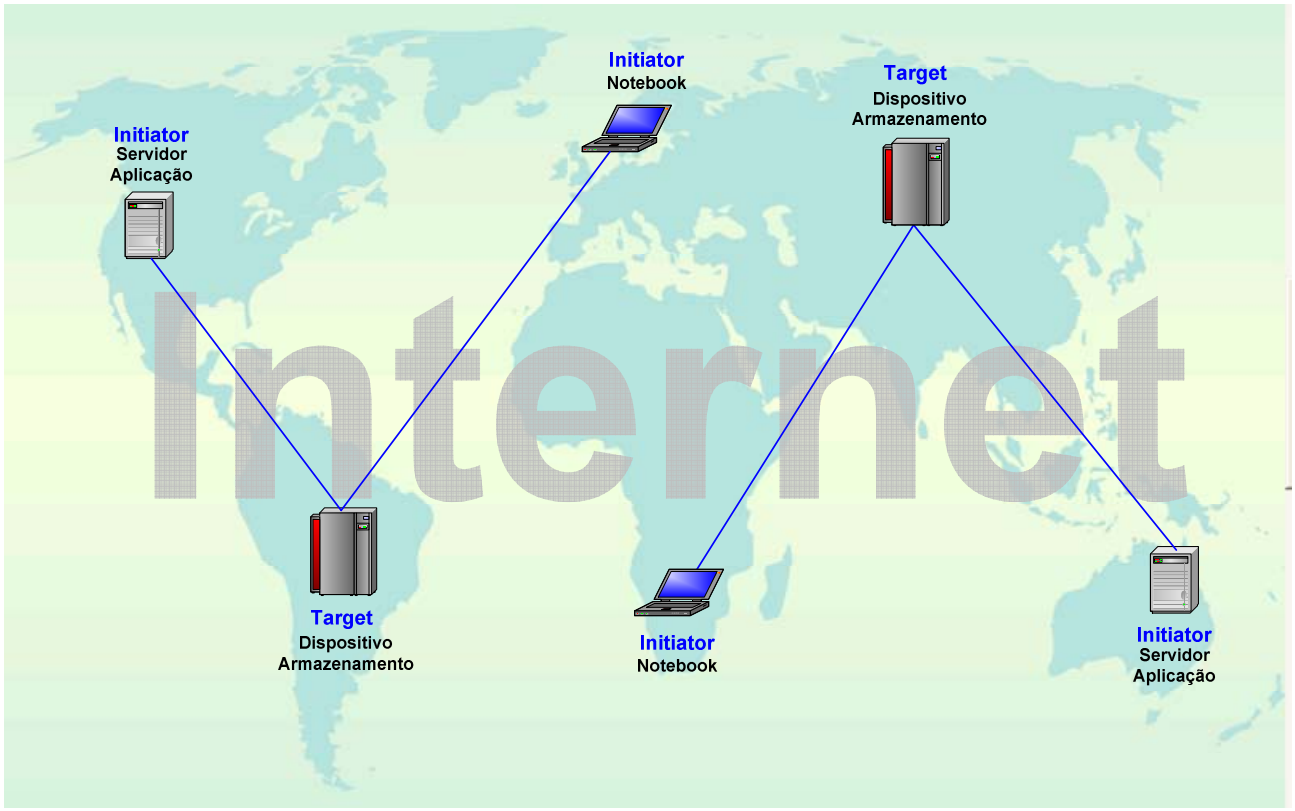


Figura 32 – Protocolo iSCSI – Internet

Capítulo 5

5.1 Análise de Desempenho dos protocolos iSCSI e Fibre Channel

No capítulo 4, foram apresentadas inúmeras características sobre a adoção e implementação do protocolo iSCSI em relação ao protocolo Fibre Channel em redes de armazenamento de dados. No entanto, a utilização do protocolo iSCSI em uma infraestrutura de redes de armazenamento de dados ainda não foi totalmente difundida nas corporações.

O principal questionamento quanto a adoção e a utilização desse protocolo está relacionada quanto ao seu desempenho, em relação ao desempenho obtido pelo protocolo Fibre Channel, pois o protocolo iSCSI baseia-se em uma infraestrutura TCP/IP, estando sujeito ao processamento e aos *overheads* introduzidos pela pilha TCP/IP, bem como ao enquadramento Ethernet.

Como pode ser ilustrado na Figura 33, para a transmissão dos comandos SCSI pelo protocolo iSCSI torna-se necessário, a adição de *overheads* adicionais a cada camada da pilha de protocolos TCP/IP para gerenciamento e controle de informações entre as camadas respectivas do modelo TCP/IP.

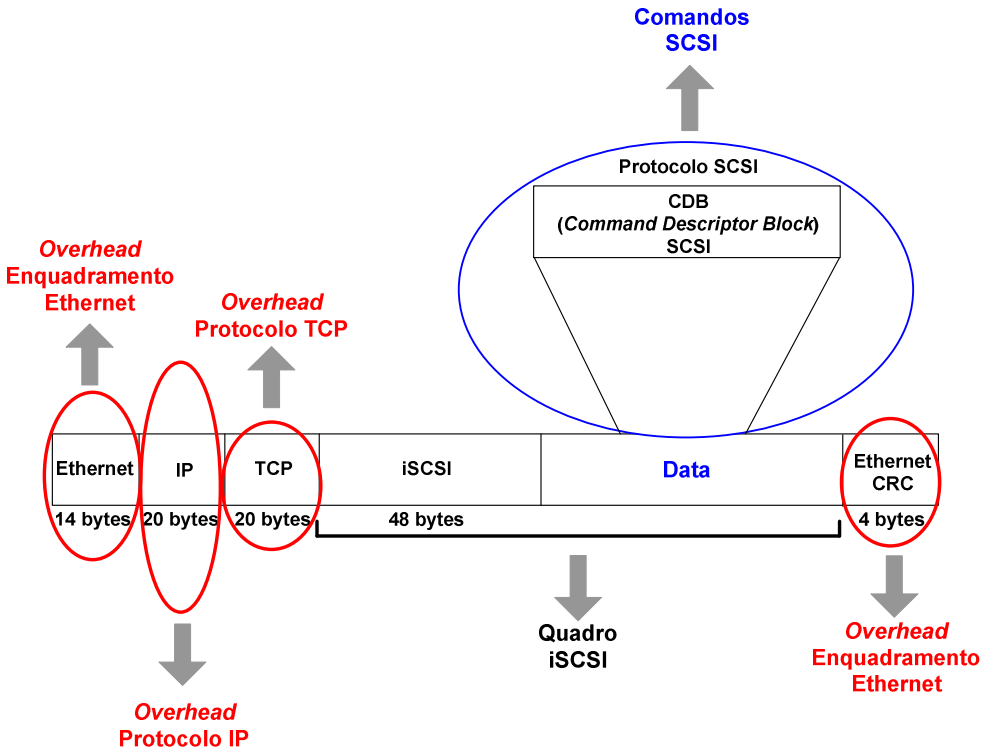


Figura 33 – *Overhead* Pilha TCP/IP + protocolo iSCSI

Como pode ser visto também na Figura 34, o protocolo Fibre Channel (FCP) utilizado na implementação de uma arquitetura SAN FCP adota um quadro que não possui *overheads* adicionais ao seu enquadramento padrão, responsável pelo transporte do protocolo SCSI. O protocolo Fibre Channel (FCP) não possui nenhuma herança relacionada a camadas de transporte, camada de rede e camada de enlace como no modelo TCP/IP.

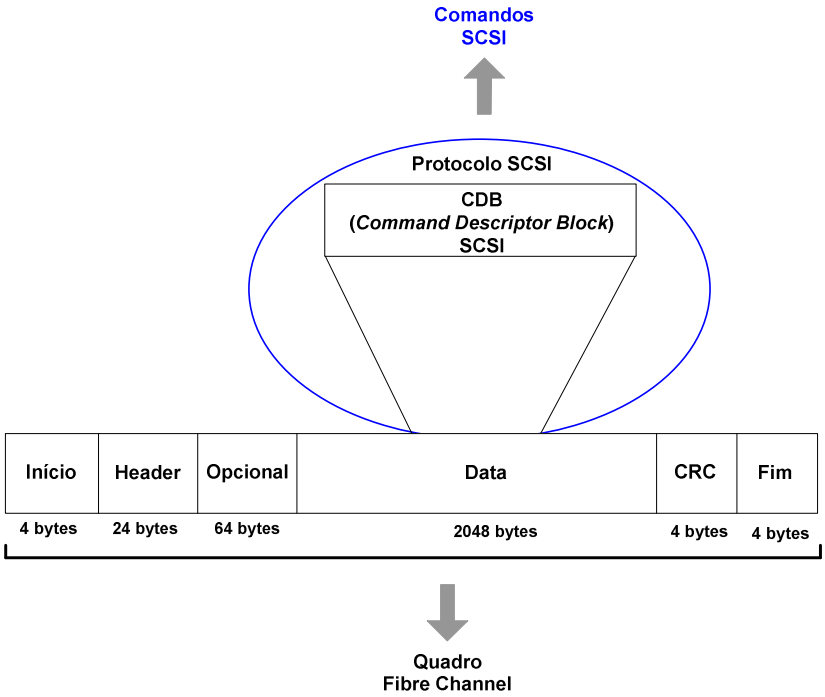


Figura 34 – Formato Quadro Fibre Channel (sem *overheads* adicionais)

O objetivo do presente trabalho é comparar o desempenho dos protocolos iSCSI e Fibre Channel dado que o protocolo iSCSI oferece alto *overhead* quando comparado ao protocolo Fibre Channel. As métricas utilizadas na comparação são Vazão e Utilização da CPU. Experimentos reais em plataformas submetidas a carga sintética foram realizados. Comparou-se o desempenho do protocolo iSCSI com e sem o uso de adaptadores iSCSI HBA cujo objetivo é aliviar a carga da CPU referente ao processamento do protocolo TCP, bem como estas duas variações com o protocolo Fibre Channel.

5.1.1 Descrição dos Experimentos realizados para Avaliação de Desempenho

Os experimentos utilizados para a Análise de Desempenho do protocolo iSCSI foram divididos em três cenários:

- ✓ **Cenário 1 –** Comparação dos protocolos iSCSI (Software) e Fibre Channel sob carga de requisições homogêneas;
- ✓ **Cenário 2 –** O Impacto da utilização de adaptadores iSCSI HBA;
- ✓ **Cenário 3 –** Comparação dos protocolos iSCSI (HBA) e Fibre Channel sujeitos à carga heterogênea.

O propósito dos experimentos realizados foi de analisar o comportamento geral de desempenho dos protocolos iSCSI e Fibre Channel, submetendo-os a padrões sintéticos de solicitações de operações de E/S e as seguintes métricas de interesse foram analisadas:

- ✓ **Vazão –** taxa de transferência efetiva expressa em MBytes/s em cada protocolo;
- ✓ **Utilização da CPU –** porcentagem do tempo em que CPU esteve ocupada com o processamento de pacotes em cada protocolo.

5.1.2 Descrição Componentes utilizados na avaliação dos Experimentos

Como pode ser visto na Figura 35, os componentes de hardware e software utilizados foram:

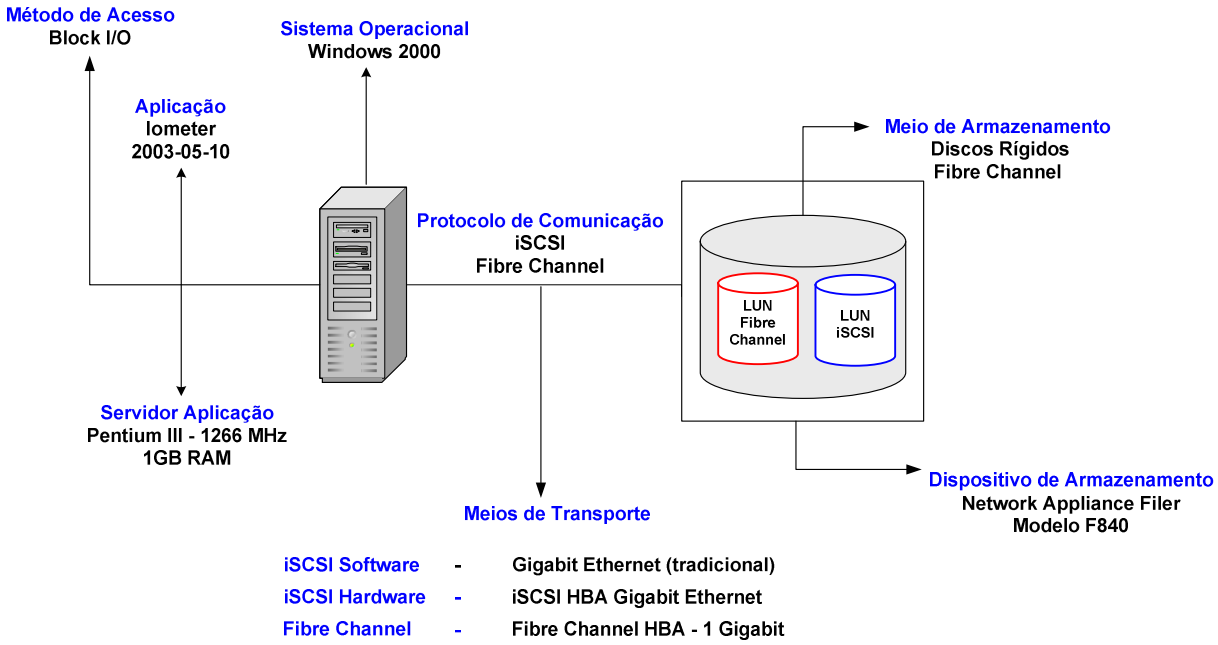


Figura 35 – Componentes utilizados na avaliação dos Experimentos

- ✓ **Servidor de Aplicação –** um servidor Dell modelo Power Edge 2550 – Pentium III com velocidade de 1266 MHz de CPU e 1GB de RAM;

- ✓ **Sistema Operacional –** o sistema operacional "Windows 2000 Advanced Server" utilizando Service Pack 3;

- ✓ **Protocolo de Comunicação –** foi utilizado em cada fase de cada experimento os protocolos iSCSI e Fibre Channel.

- **Meios de Transporte –** os meios de transporte utilizados foram:

  - **Fibre Channel –** uma interface HBA Fibre Channel "Emulex LightPulse LP9002L-F2" em um barramento PCI de 64 bits, configurada a uma velocidade nomimal de 1Gbit/s. O device driver instalado no Servidor de Aplicação foi o *driver* Emulex LightPulse versão 5-4.82a4. A instalação desse driver é essencial para o reconhecimento do dispositivo no sistema operacional Windows 2000 e também para o funcionamento do dispositivo como uma interface *Initiator* Fibre Channel;

  - **iSCSI Software –** uma interface Gigabit Ethernet tradicional "Intel Gigabit Ethernet 1000/MX Server Adapter" em um barramento PCI de 64 bits, configurada a uma velocidade nominal de 1Gbit/s. O device driver instalado no Servidor de Aplicação foi o driver padrão do sistema operacional Windows 2000 Advanced Server e também foi instalado o device driver "Microsoft iSCSI *Initiator*" versão 1.01, onde a instalação do primeiro device driver é essencial para o reconhecimento do dispositivo no sistema operacional Windows 2000 e a instalação do segundo device driver é fundamental para o funcionamento do dispositivo como uma interface *Initiator* iSCSI;

  - **iSCSI Hardware –** uma interface iSCSI HBA Gigabit Ethernet "Intel Pro IP Storage 1000" em um barramento PCI de 64 bits, configurada a uma velocidade nominal de 1Gbit/s. O device driver instalado no Servidor de Aplicação foi o driver Intel Pro IP versão 1.0. A instalação desse driver é essencial para o reconhecimento do dispositivo no sistema operacional Windows 2000 e também para o funcionamento do dispositivo como uma interface *Initiator* iSCSI HBA;

- ✓ **Dispositivo de Armazenamento –** o Dispositivo de Armazenamento [28] equipamento "Network Appliance" modelo F840 com suporte aos protocolos iSCSI e Fibre Channel.

  A configuração utilizada no Dispositivo de Armazenamento foi:

  - Sistema Operacional Data ONTAP 6.4.1;
  - Meio de Armazenamento configurado com suporte a tolerância a falhas (RAID 4);
  - Os discos rígidos utilizados no Dispositivo de Armazenamento são do tipo Fibre Channel com tamanho de 72GB e velocidade de 10,000 RPM (rotações por minuto);
  - Volume /vol/data contendo um configuração de 10 discos + 1 disco, tendo 10 discos de dados e 1 disco de paridade implementando o modelo RAID 4;
  - Logical Unit para os protocolos iSCSI e Fibre Channel, configuradas no volume /vol/data , cada uma com tamanho de 50GB (Figura 34);
  - Interfaces de rede Gigabit Ethernet para o tráfego iSCSI e interfaces Fibre Channel HBA Fibre Channel para o tráfego Fibre Channel.

- ✓ **Meio de Armazenamento –** os discos rígidos utilizados no Dispositivo de Armazenamento são do tipo Fibre Channel, cada um com tamanho de 72GB com velocidade de 10,000 RPM (rotações por minuto);

- ✓ **Aplicação –** o software Iometer [29] [34] [35] em todos os experimentos para geração de carga sintética, simulação de tráfego, captura e análise dos dados. Foi adotado essa ferramenta devido ao fato de ser muito utilizado em diversos experimentos executados por diversos fabricantes para geração de carga sintética e para medições de desempenho de Dispositivos de Armazenamento [30] [31] [32]. Em todos os experimentos a versão utilizada da ferramenta Iometer foi a versão 2003-05-10.

As principais características do perfil de uma aplicação utilizadas são:

- ✓ **Quantidade de Threads:** especifica a quantidade de threads (workers) responsáveis pelo execução em paralelo de requisições de E/S para o Dispositivo de Armazenamento;
- ✓ **Porcentagem de Requisições READ e/ou WRITE:** porcentagem de operações de leitura e ou escrita utilizadas pela aplicação ao Dispositivo de Armazenamento;
- ✓ **Tamanho do Bloco Requisitado:** tamanho do bloco de dados utilizado pela aplicação;
- ✓ **Porcentagem de Acesso Sequencial e/ou Randômico:** especifica a porcentagem do perfil de acesso pela aplicação às informações armazenadas no Dispositivo de Armazenamento. As aplicações podem possuir um perfil de acesso Sequencial ou Randômico aos dados;

Como pode ser ilustrado nas Tabelas 7, 8 e 9, as configurações utilizadas pela ferramenta Iometer nos três cenários foram:

Cenário 1
Comparação dos protocolos iSCSI (Software) e Fibre Channel sob carga de requisições homogêneas

Iometer				
Protocolo	Threads	Requisições	Acesso	Bloco
iSCSI	4	100% READ	Sequencial	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
FCP	4	100% READ	Sequencial	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
iSCSI	4	100% READ	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
FCP	4	100% READ	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
Protocolo	Threads	Requisições	Acesso	Bloco
iSCSI	4	100% WRITE	Sequencial	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
FCP	4	100% WRITE	Sequencial	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
iSCSI	4	100% WRITE	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
FCP	4	100% WRITE	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes

Tabela 7 – Configurações Iometer – Cenário 1

Cenário 2

Impacto da utilização de adaptadores iSCSI HBA

Iometer				
Protocolo	Threads	Requisições	Acesso	Bloco
iSCSI	4	100% READ	Sequencial	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
iSCSI HBA	4	100% READ	Sequencial	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
iSCSI	4	100% READ	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
iSCSI HBA	4	100% READ	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
Protocolo	Threads	Requisições	Acesso	Bloco
iSCSI	4	100% WRITE	Sequencial	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
iSCSI HBA	4	100% WRITE	Sequencial	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
iSCSI	4	100% WRITE	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
iSCSI HBA	4	100% WRITE	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes

Tabela 8 – Configurações Iometer – Cenário 2

Cenário 3

Comparação dos protocolos iSCSI (HBA) e Fibre Channel sujeitos à carga heterogênea

Iometer				
Protocolo	Threads	Requisições	Acesso	Bloco
iSCSI HBA	4	67% READ 33% WRITE	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
FCP	4	67% READ 33% WRITE	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
iSCSI HBA	4	67% WRITE 33% READ	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
FCP	4	67% WRITE 33% READ	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
Protocolo	Threads	Requisições	Acesso	Bloco
iSCSI HBA	4	67% READ 33% WRITE	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
FCP	4	67% READ 33% WRITE	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
iSCSI HBA	4	67% WRITE 33% READ	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes
FCP	4	67% WRITE 33% READ	Randômico	512 bytes, 1024 bytes, 2048 bytes, 4096 bytes, 8192 bytes, 16384 bytes, 32768 bytes e 65536 bytes

Tabela 9 – Configurações Iometer – Cenário 3

O principal objetivo do cenário 1 e 2 foi analisar o comportamento dos protocolos iSCSI e Fibre Channel sob uma carga sintética com um perfil homogêneo de requisições de READ e WRITE.

O cenário 1 baseou-se em uma análise das métricas Vazão e Utilização da CPU utilizando o protocolo iSCSI implementado em nível de software em comparação com o protocolo Fibre Channel. O cenário 2 baseou-se em uma análise das métricas Vazão e Utilização da CPU utilizando o protocolo iSCSI implementado em nível de software em comparação com o protocolo iSCSI implementado em nível de hardware através de um adaptador iSCSI HBA.

O principal objetivo do cenário 3 foi analisar o comportamento dos protocolos iSCSI HBA e Fibre Channel HBA sob uma carga sintética heterogênea baseado [33] em um acesso randômico às informações armazenadas no Dispositivo de Armazenamento.

Em todos os experimentos dos cenários 1, 2 e 3 foi adotado [33] também a quantidade de 4 threads (workers) responsáveis pelo execução de requisições em paralelo de E/S para o Dispositivo de Armazenamento.

Durante os experimentos de todos os cenários, a aplicação Iometer variou o tamanho do bloco a ser transferido e a cada experimento utilizou-se um tamanho de bloco específico com uma duração de tempo de 5 minutos; duração esta suficiente para se coletar valores das métricas de interesse em regime estacionário. Os cenários utilizados nos experimentos foram mencionados nas Tabelas 7, 8 e 9 e especificados na Seção 5.1.2.

Foi adotado também, em todos os cenários, a utilização do Dispositivo de Armazenamento em um formato ***Block I/O,*** o qual pode ser definido como um formato de acesso a uma unidade de armazenamento pela Aplicação de Sistema sem a necessidade de utilização e montagem de um sistema de arquivos pelo sistema operacional do Servidor de Aplicação.

A principal finalidade de acesso ao Dispositivo de Armazenamento em um formato ***Block I/O*** é minimizar a possibilidade de utilização de cacheamento das informações pelo sistema de arquivos no Servidor de Aplicação, onde a possibilidade do cacheamento das informações poderia influenciar a coleta e análise do resultados nos experimentos.

- ✓ **Método de Acesso –** foi adotado em todos os cenários, a utilização do Dispositivo de Armazenamento em um formato ***Block I/O***, com a principal finalidade de minimizar a utilização de cacheamento dos dados pelo sistema de arquivos.

5.2 Resultados Numéricos

Nesta seção, os resultados obtidos nos três cenários de experimentos realizados são descritos. No cenário 1, o desempenho dos protocolos iSCSI (Software) e Fibre Channel são analisados quando submetidos à uma carga de requisições homogêneas de READ e WRITE baseado em acesso sequencial e randômico. No cenário 2, compara-se o desempenho do protocolo iSCSI com e sem o uso de adaptadores iSCSI HBA, submetidos também a mesma carga máxima de requisições homogêneas de READ e WRITE tanto com acesso sequencial quanto com randômico. No cenário 3, compara-se o desempenho do protocolo iSCSI, com uso de adaptadores iSCSI HBA, e do protocolo Fibre Channel submetendo-os a carga genérica de requisições heterogêneas de READ e WRITE baseado com acesso randômico das informações.

O principal objetivo de todos os experimentos não foi analisar pontualmente o comportamento de desempenho dos protocolos iSCSI e Fibre Channel em relação ao tamanho do bloco transferido pela Aplicação. O principal objeto de estudo desses experimentos foi uma análise do comportamento geral de desempenho dos protocolos iSCSI e Fibre Channel tomando como base de medida e análise uma Mediana dos valores obtidos em cada tipo de requisição de E/S em um perfil de acesso correspondente.

5.2.1 Comparação dos protocolos iSCSI (Software) e Fibre Channel sob carga de requisições homogêneas

O propósito inicial desse primeiro experimento foi de avaliar as métricas Vazão e Utilização da CPU utilizando os protocolos iSCSI (Software) e Fibre Channel submetidos a uma carga de requisições homogêneas de READ e WRITE com acesso sequencial e randômico das informações.

5.2.1.1 Métrica Vazão – Acesso Sequencial e Randômico

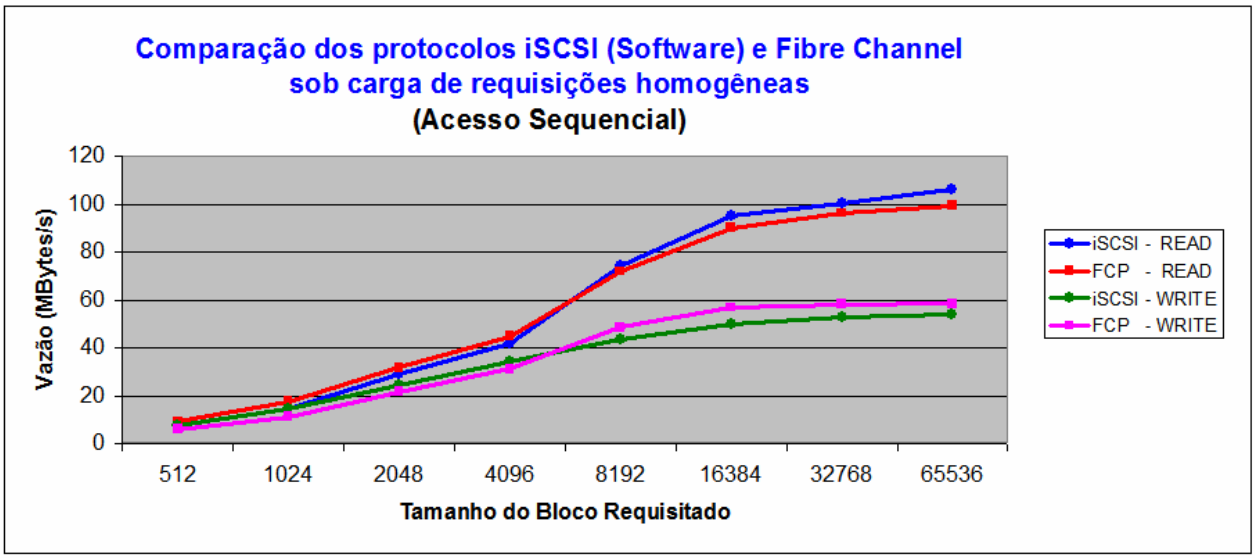


Figura 36 – Cenário 1 – Métrica Vazão – Acesso Sequencial

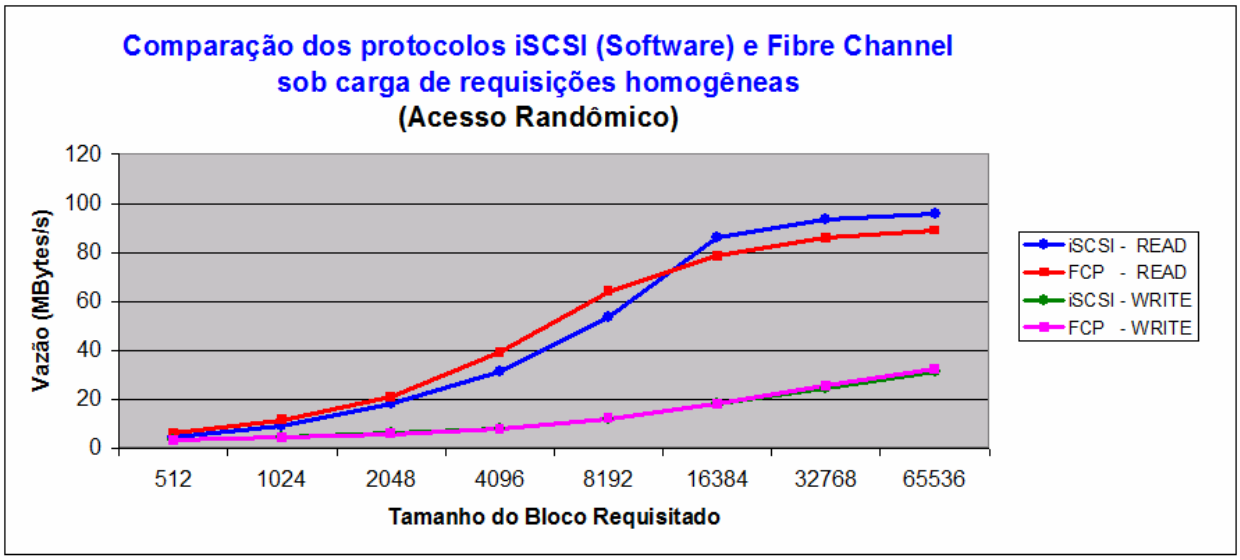


Figura 37 – Cenário 1 – Métrica Vazão – Acesso Randômico

Como pode ser visto nas Figuras 36 e 37, os resultados obtidos nos experimentos do Cenário 1 (Métrica Vazão) para acesso sequencial e randômico dos protocolos iSCSI (Software) e Fibre Channel (FCP).

A Tabela 10 ilustra os resultados dos experimentos do Cenário 1 (Métrica Vazão), onde os resultados obtidos estão classificados em seções (experimentos) baseados no tipo de requisição de E/S com perfil de acesso correspondente. Cada coluna possui oito (8) valores em função do tamanho de bloco transferido pela Aplicação. O resultado final de cada experimento corresponde ao resultado da operação mediana [33] dos valores obtidos de cada seção correspondente.

Cenário 1 (Resultados)								
Comparação dos protocolos iSCSI (Software) e Fibre Channel sob carga de requisições homogêneas								
Métrica Vazão (MBytes/s)								
Tamanho Bloco Requisitado	Protocolo iSCSI (Software)				Protocolo Fibre Channel			
	Sequencial READ	Sequencial WRITE	Randômico READ	Randômico WRITE	Sequencial READ	Sequencial WRITE	Randômico READ	Randômico WRITE
512	7.57279	7.63207	4.688299	3.253081	9.260872	6.011765	6.071385	3.100401
1024	14.670763	14.644231	8.998376	4.237077	17.507276	11.247188	11.552453	4.310764
2048	29.065081	24.618791	18.304336	5.982990	31.706334	21.722187	20.712085	5.691934
4096	41.760071	34.298947	31.479853	7.872376	44.783676	31.177573	38.973830	7.695974
8192	74.558639	43.481217	53.419443	12.047335	71.890050	48.446999	63.951435	11.996745
16384	95.287198	49.994882	86.136293	17.956190	90.111187	56.751974	78.457017	17.987869
32768	100.392091	52.807353	93.668774	24.308214	96.250536	58.181757	85.761666	25.403182
65536	105.977945	53.761480	95.489813	30.948927	99.402723	58.425738	88.898060	32.105307
	MEDIANA							
	58.159355	38.890082	42.449648	9.9598555	58.336863	39.812286	51.4626325	9.8463595

Tabela 10 – Cenário 1 (Vazão) – Resultados

Com base na Tabela 10, o protocolo iSCSI (Software) obteve um perfil geral de de desempenho de 58.159355 MBytes/s em operações de READ com acesso sequencial, 38.890082 MBytes/s em operações de WRITE com acesso sequencial, 42.449648 MBytes/s em operações de READ com acesso randômico e 9.9598555 MBytes/s em operações de WRITE com acesso randômico.

O protocolo Fibre Channel (FCP) obteve um perfil geral de desempenho de 58.336863 MBytes/s em operações de READ com acesso sequencial, 39.812286 MBytes/s em operações de WRITE com acesso sequencial, 51.4626325 MBytes/s em operações de READ com acesso randômico e 9.8463595 MBytes/s em operações de WRITE com acesso randômico.

Observando os dados obtidos nas Figuras 36 a 38 e na Tabela 10, pode-se concluir que o resultados obtidos pelo protocolo iSCSI (Software) demonstram uma diferença pouco expressiva em relação ao desempenho obtido pelo protocolo Fibre Channel, sendo de aproximadamente 0.18 MBytes/s em operações de READ com acesso sequencial, 0.92 MBytes/s em operações de WRITE com acesso sequencial, 9.01 MBytes/s em operações de READ com acesso randômico e -0.11 MBytes/s em operações de WRITE com acesso randômico do protocolo iSCSI (Software) em relação ao protocolo Fibre Channel (FCP) na métrica Vazão.

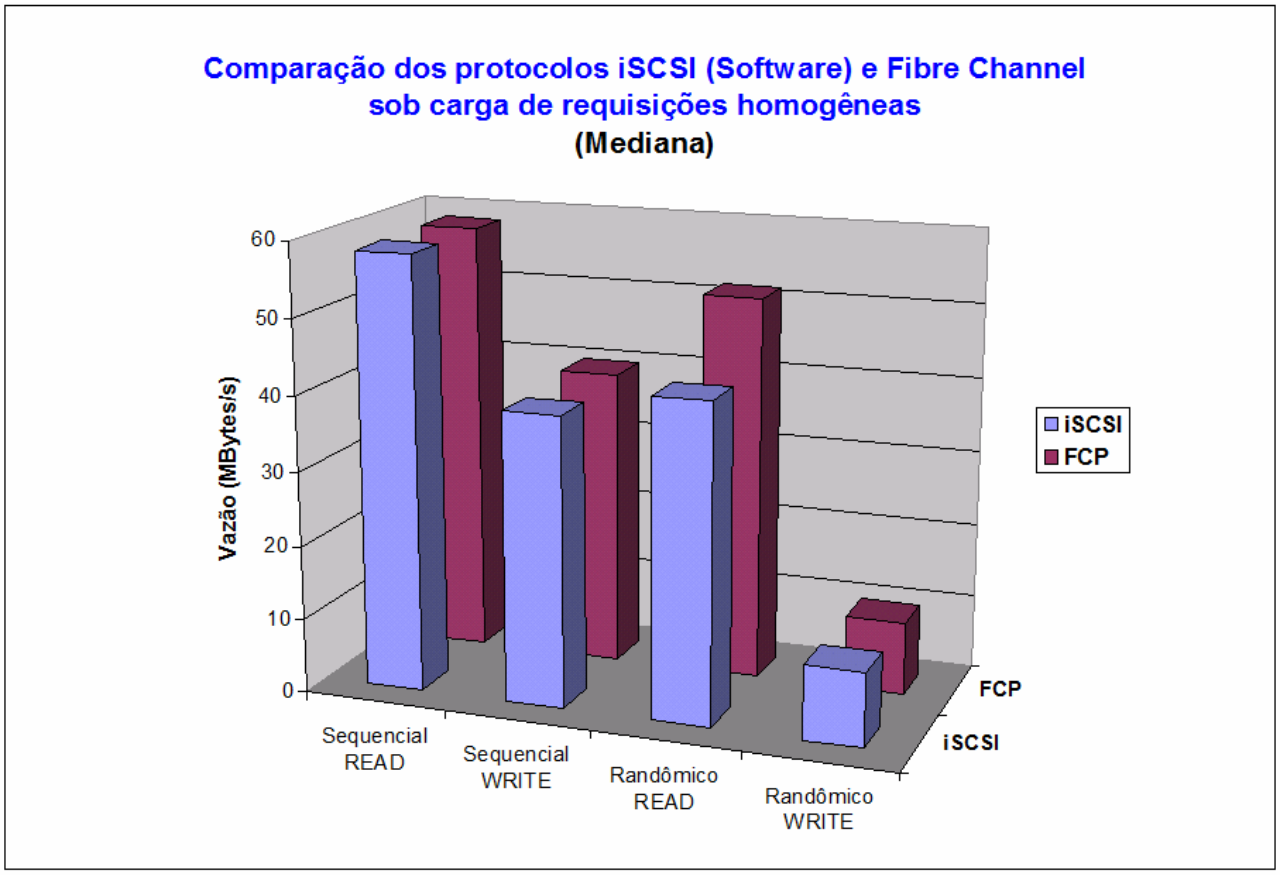


Figura 38 – Cenário 1 (Métrica Vazão) – Resultados Mediana

5.2.1.2 Métrica Utilização da CPU – Acesso Sequencial e Randômico

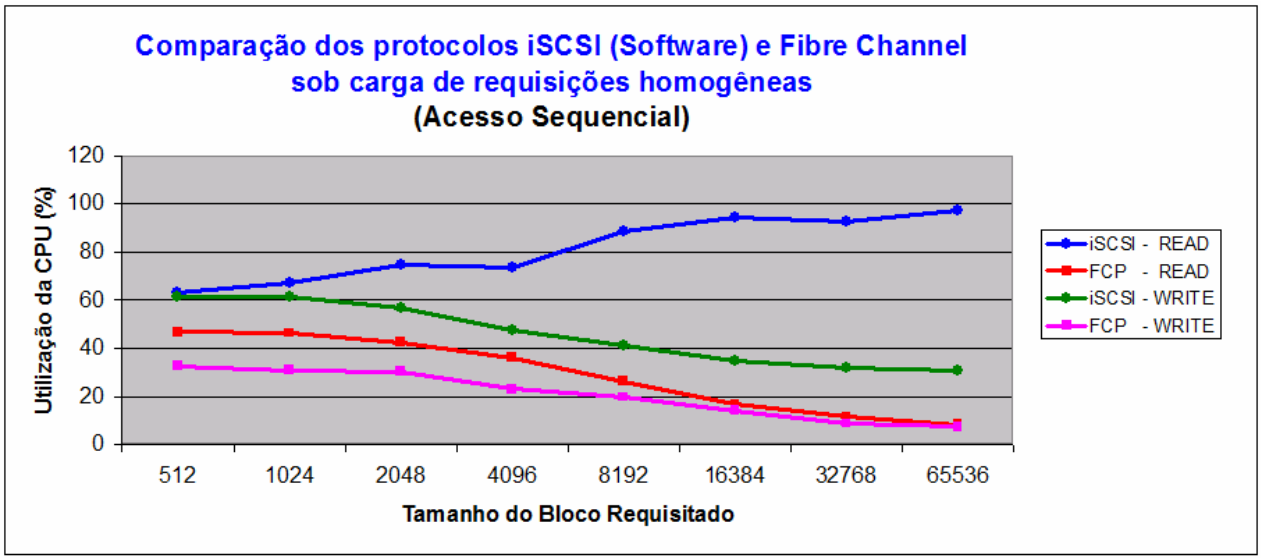


Figura 39 – Cenário 1 – Métrica Utilização da CPU – Acesso Sequencial

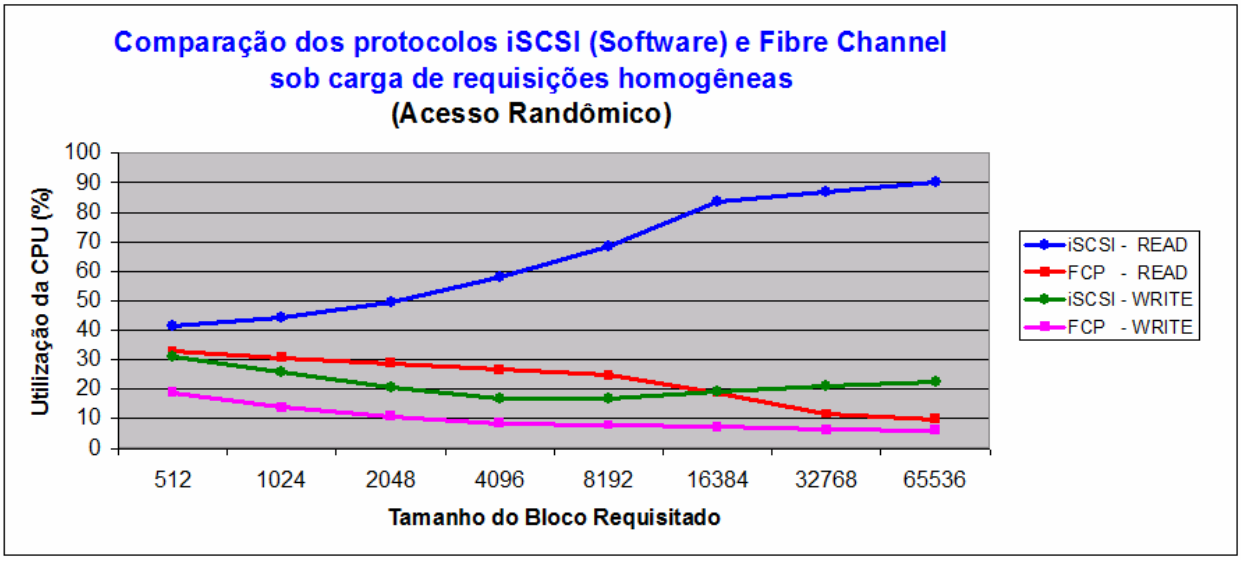


Figura 40 – Cenário 1 – Métrica Utilização da CPU – Acesso Randômico

Como pode ser visto nas Figuras 39 e 40, os resultados obtidos nos experimentos do Cenário 1 (Métrica Utilização da CPU) para acesso sequencial e randômico dos protocolos iSCSI (Software) e Fibre Channel (FCP).

A Tabela 11 ilustra os resultados dos experimentos do Cenário 1 (Métrica Utilização da CPU), onde os resultados obtidos estão classificados em seções (experimentos) baseados no tipo de requisição de E/S com perfil de acesso correspondente. Cada coluna possui oito (8) valores em função do tamanho de bloco transferido pela Aplicação. O resultado final de cada experimento corresponde ao resultado da operação mediana [33] dos valores obtidos de cada seção correspondente.

Cenário 1 (Resultados)								
Comparação dos protocolos iSCSI (Software) e Fibre Channel sob carga de requisições homogêneas								
Métrica Utilização da CPU (% CPU)								
Tamanho Bloco Requisitado	Protocolo iSCSI (Software)				Protocolo Fibre Channel			
	Sequencial READ	Sequencial WRITE	Randômico READ	Randômico WRITE	Sequencial READ	Sequencial WRITE	Randômico READ	Randômico WRITE
512	62.865665	61.525007	41.662019	31.081730	46.757838	32.601017	32.762110	18.773507
1024	67.229053	61.194446	44.418904	25.735051	46.298495	31.049276	30.789615	13.776736
2048	74.520236	56.710700	49.448646	20.485362	42.374758	30.346201	28.834915	10.797099
4096	73.518164	47.347277	57.82632	16.867756	36.239401	23.055753	26.639369	8.512193
8192	88.315293	41.303010	68.421222	16.857806	26.224502	19.858125	24.736191	7.833110
16384	94.261697	34.980084	83.683296	19.238506	16.852087	14.049661	18.957460	7.139059
32768	92.830573	31.724098	87.151165	21.100504	11.581052	8.795209	11.699461	6.305113
65536	97.124376	30.982236	90.098462	22.662323	8.430176	7.325389	9.925467	6.116336
	MEDIANA							
	81.4177645	44.3251435	63.123771	20.792933	31.2319515	21.456939	25.68778	8.1726515

Tabela 11 – Cenário 1 (Utilização da CPU) – Resultados

Com base na Tabela 11, o protocolo iSCSI (Software) obteve um perfil geral de Utilização da CPU de 81.4177645% em operações de READ com acesso sequencial, 44.3251435% em operações de WRITE com acesso sequencial, 63.123771% em operações de READ com acesso randômico e 20.792933% em operações de WRITE com acesso randômico.

O protocolo Fibre Channel (FCP) obteve um perfil geral de Utilização da CPU de 31.2319515% em operações de READ com acesso sequencial, 21.456939% em operações de WRITE com acesso sequencial, 25.68778% em operações de READ com acesso randômico e 8.1726515% em operações de WRITE com acesso randômico.

Observando os dados obtidos nas Figuras 39 a 41 e na Tabela 11, pode-se concluir que a Utilização da CPU obtida pelo protocolo iSCSI (Software) difere significativamente em relação a utilização da CPU detida pelo protocolo Fibre Channel (FCP). Essa diferença de utilização de CPU é de aproximadamente 61.64% em operações de READ com acesso sequencial, 51.59% em operações de WRITE com acesso sequencial, 59.31% em operações de READ com acesso randômico e 60.70% em operações de WRITE com acesso randômico do protocolo iSCSI (Software) em relação ao protocolo Fibre Channel (FCP).

Essa diferença significativa na utilização da CPU deve-se ao fato do protocolo Fibre Channel possuir um quadro sem nenhum *overhead* adicional ao seu enquadramento padrão (Figura 34). Todo o processamento dos quadros Fibre Channel é executado pela interface Fibre Channel (HBA) instalada no Servidor de Aplicação, ficando portanto somente sob responsabilidade da CPU do Servidor de Aplicação o processamento da mensagem SCSI CDB.

Com a utilização do protocolo iSCSI (Software) é necessário executar todo o processamento dos pacotes iSCSI e TCP/IP pela CPU do Servidor de Aplicação, gerando assim um acréscimo significativo da Utilização da CPU no Servidor de Aplicação. Essa alta de utilização da CPU também está relacionado diretamente com a quantidade de tráfego de dados expressiva produzida pelo protocolo iSCSI durante os experimentos do Cenário 1 (Métrica Vazão).

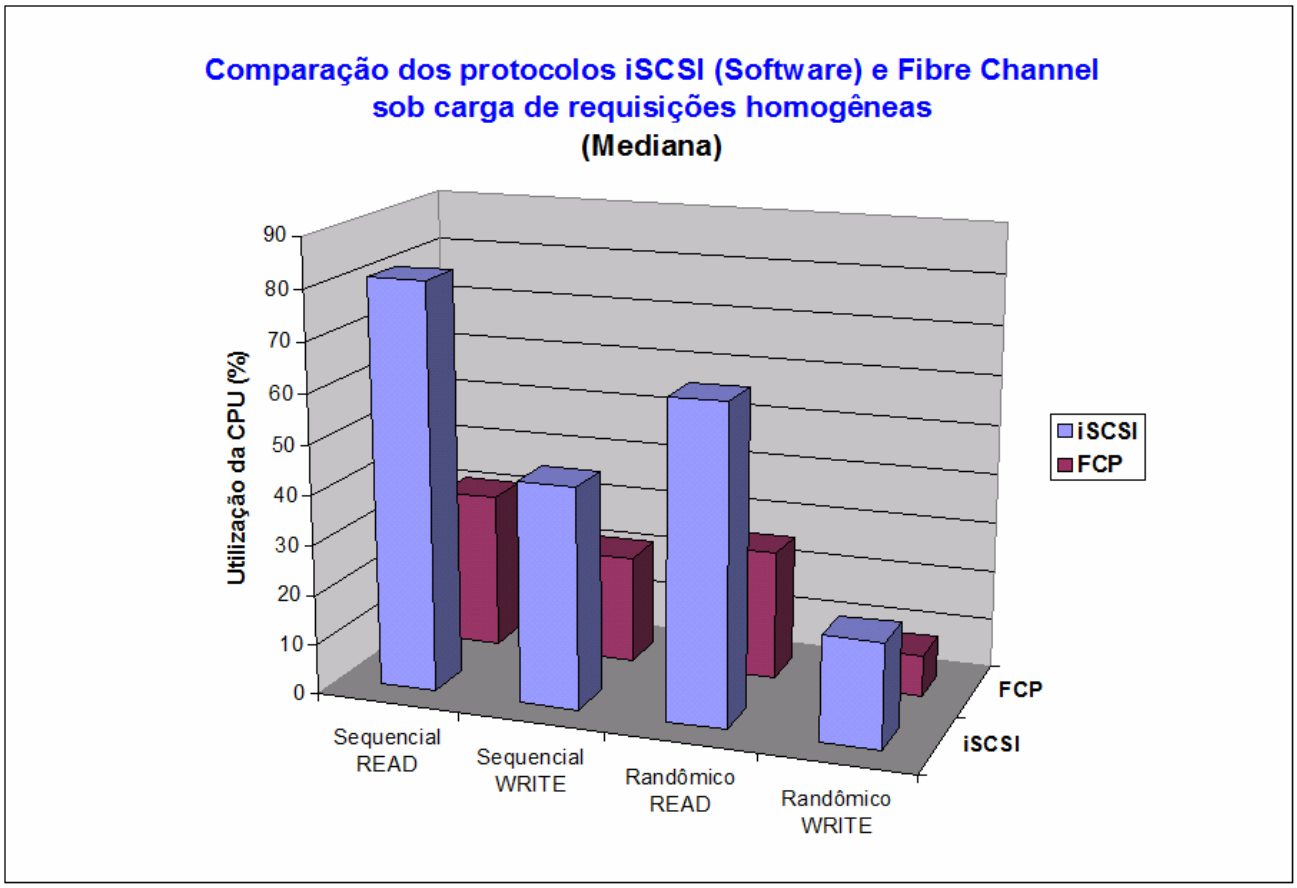


Figura 41 – Cenário 1 (Métrica Utilização da CPU) – Resultados Mediana

5.2.2 O Impacto da utilização de adaptadores iSCSI HBA

O propósito inicial desse segundo experimento foi de avaliar as métricas Vazão e Utilização da CPU na utilização dos protocolos iSCSI (Software) e iSCSI (HBA) em mesmo perfil máximo de requisições homogêneas de READ e WRITE baseados em um acesso sequencial e randômico no armazenamento e na recuperação das informações localizadas no Dispositivo de Armazenamento.

5.2.2.1 Métrica Vazão – Acesso Sequencial e Randômico

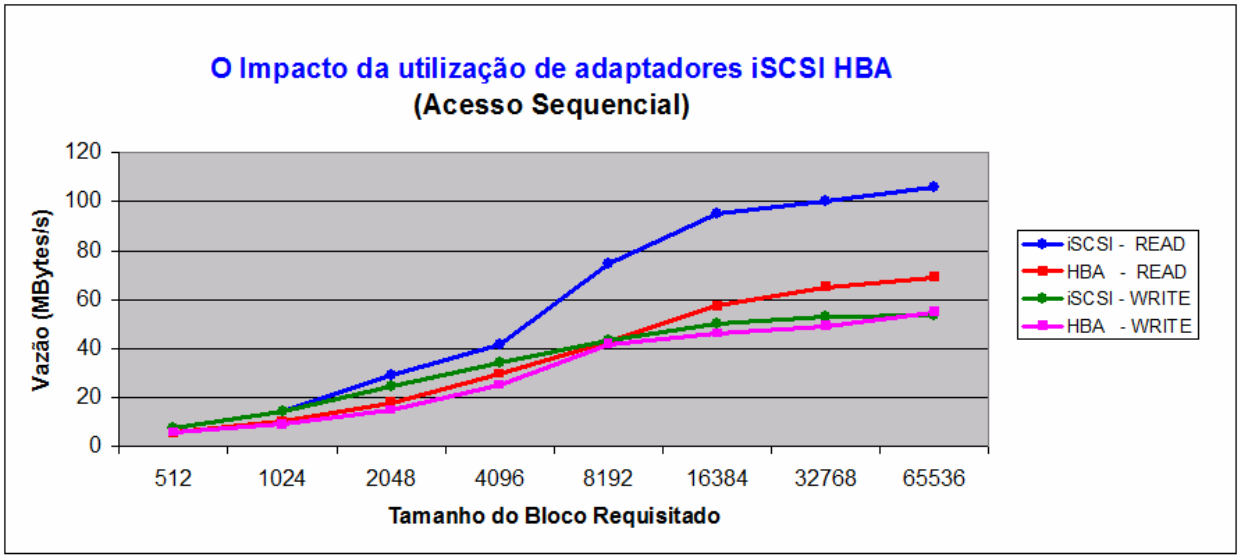


Figura 42 – Cenário 2 – Métrica Vazão – Acesso Sequencial

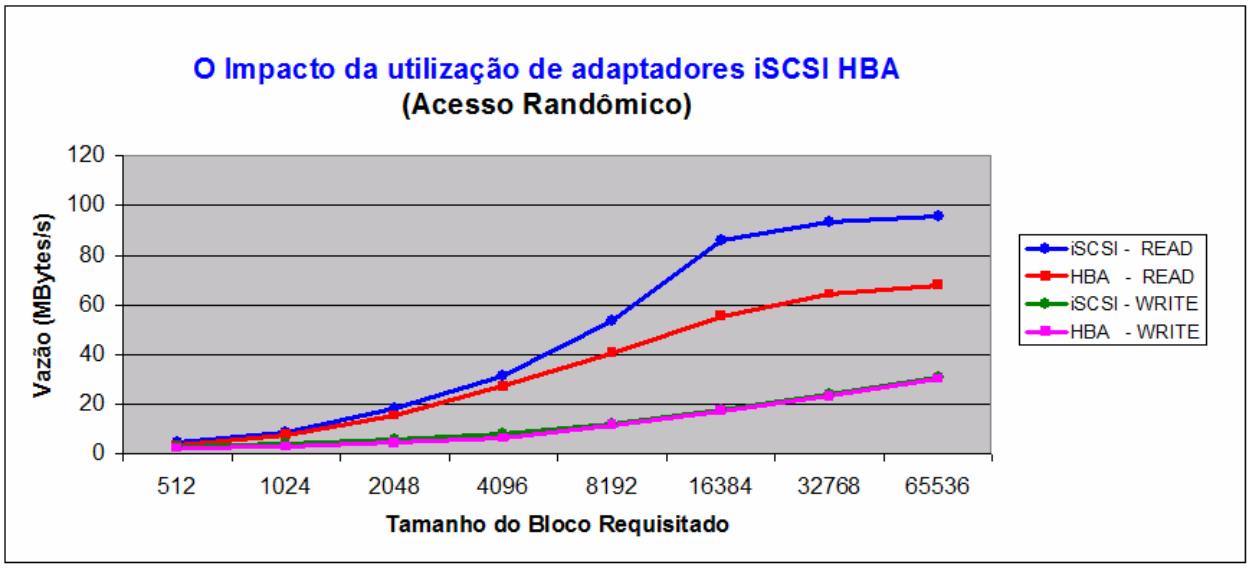


Figura 43 – Cenário 2 – Métrica Vazão – Acesso Randômico

Como pode ser visto nas Figuras 42 e 43, os resultados obtidos nos experimentos do Cenário 2 (Métrica Vazão) para acesso sequencial e randômico dos protocolos iSCSI (Software) e iSCSI (HBA).

A Tabela 12 ilustra os resultados dos experimentos do Cenário 2 (Métrica Vazão), onde os resultados obtidos estão classificados em seções (experimentos) baseados no tipo de requisição de E/S com perfil de acesso correspondente. Cada coluna possui oito (8) valores em função do tamanho de bloco transferido pela Aplicação. O resultado final de cada experimento corresponde a mediana [33] dos valores obtidos de cada seção correspondente.

Cenário 2 (Resultados)								
Impacto da utilização de adaptadores iSCSI HBA								
Métrica Vazão (MBytes/s)								
Tamanho Bloco Requisitado	Protocolo iSCSI (Software)				Protocolo iSCSI (HBA)			
	Sequencial READ	Sequencial WRITE	Randômico READ	Randômico WRITE	Sequencial READ	Sequencial WRITE	Randômico READ	Randômico WRITE
512	7.57279	7.63207	4.688299	3.253081	5.623284	5.894033	3.440867	2.421975
1024	14.670763	14.644231	8.998376	4.237077	10.457136	9.108746	7.762757	3.089170
2048	29.065081	24.618791	18.304336	5.982990	18.130401	14.963502	15.622894	4.551045
4096	41.760071	34.298947	31.479853	7.872376	29.793915	25.218832	27.382917	6.539538
8192	74.558639	43.481217	53.419443	12.047335	43.011669	41.937166	40.744482	11.792425
16384	95.287198	49.994882	86.136293	17.956190	57.445801	46.189846	55.552554	17.517482
32768	100.392091	52.807353	93.668774	24.308214	65.218132	49.310083	64.487698	23.407002
65536	105.977945	53.761480	95.489813	30.948927	69.097392	54.999125	68.028824	30.364047
	MEDIANA							
	58.159355	38.890082	42.449648	9.9598555	36.402792	33.577999	34.0636995	9.1659815

Tabela 12 – Cenário 2 (Vazão) – Resultados

Com base na Tabela 12, o protocolo iSCSI (Software) obteve um perfil geral de desempenho de 58.159355 MBytes/s em operações de READ com acesso sequencial, 38.890082 MBytes/s em operações de WRITE com acesso sequencial, 42.449648 MBytes/s em operações de READ com acesso randômico e 9.9598555 MBytes/s em operações de WRITE com acesso randômico.

O protocolo iSCSI (HBA) obteve um perfil geral de desempenho de 36.402792 MBytes/s em operações de READ com acesso sequencial, 33.577999 MBytes/s em operações de WRITE com acesso sequencial, 34.0636995 MBytes/s em operações de READ com acesso randômico e 9.1659815 MBytes/s em operações de WRITE com acesso randômico.

Observando os dados obtidos nas Figuras 42 a 44 e na Tabela 12, pode-se concluir que o resultados obtidos pelo protocolo iSCSI (HBA) produzem vazão inferior quando comparado a vazão obtida pelo protocolo iSCSI (Software). A diferença de vazão é de aproximadamente, 21.76 MBytes/s em operações de READ com acesso sequencial, 5.31 MBytes/s em operações de WRITE com acesso sequencial, 8.39 MBytes/s em operações de READ com acesso randômico e 0.79 MBytes/s em operações de WRITE com acesso randômico. A maior vazão obtida pelo protocolo iSCSI (Software) deve-se ao fato do adaptador iSCSI (HBA) possuir somente uma CPU com velocidade de 600MHz em comparação com a velocidade da CPU do Servidor de Aplicação que é de 1266MHz, representando aproximadamente uma superioridade de mais de 50% de poder computacional em relação a CPU do adaptador iSCSI (HBA). Apesar dos clocks serem diferentes entres os processadores, a diferença de desempenho obtida pelo protocolo iSCSI (Software) é de aproximadamente, 37.41% em operações de READ com acesso sequencial, 13.66% em operações de WRITE com acesso sequencial, 19.76% em operações de READ com acesso randômico e 7.97% em operações de WRITE com acesso randômico.

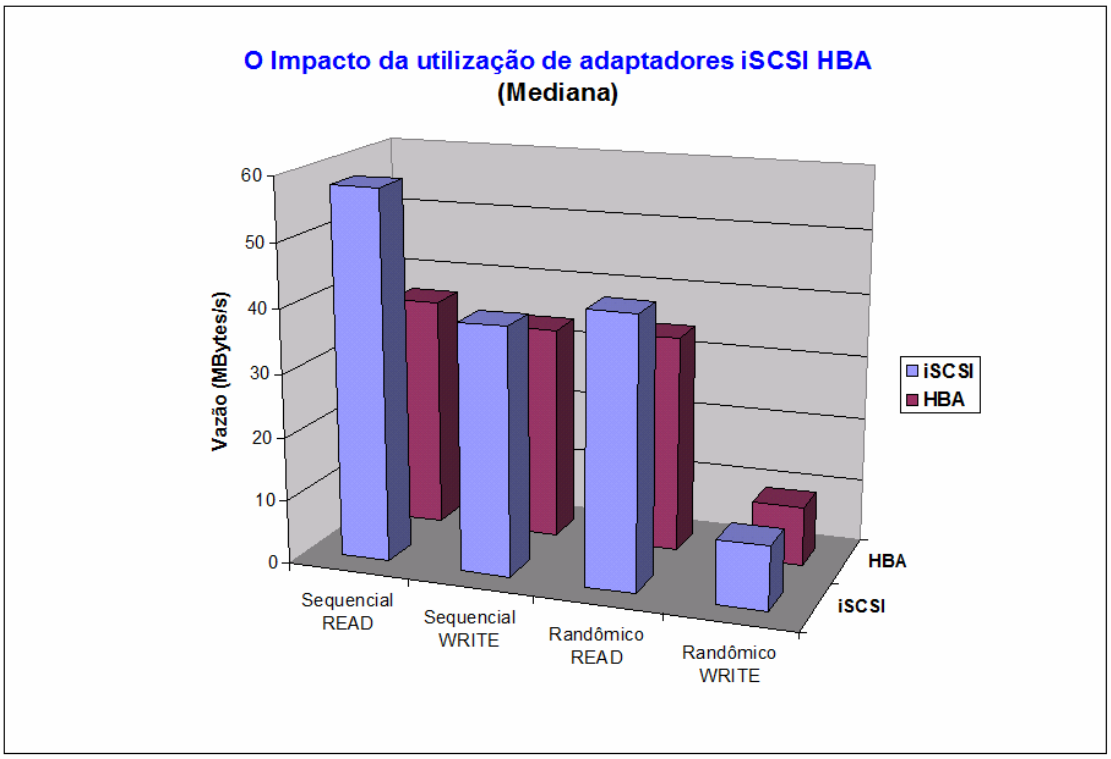


Figura 44 – Cenário 2 (Métrica Vazão) – Resultados Mediana

5.2.2.2 Métrica Utilização da CPU – Acesso Sequencial e Randômico

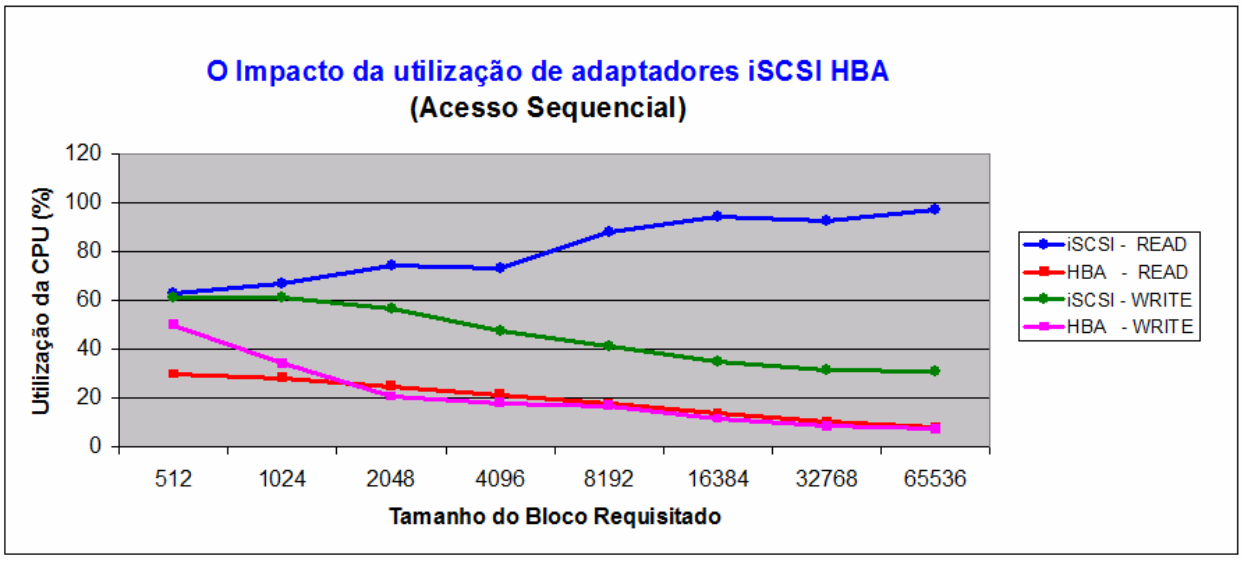


Figura 45 – Cenário 2 – Métrica Utilização da CPU – Acesso Sequencial

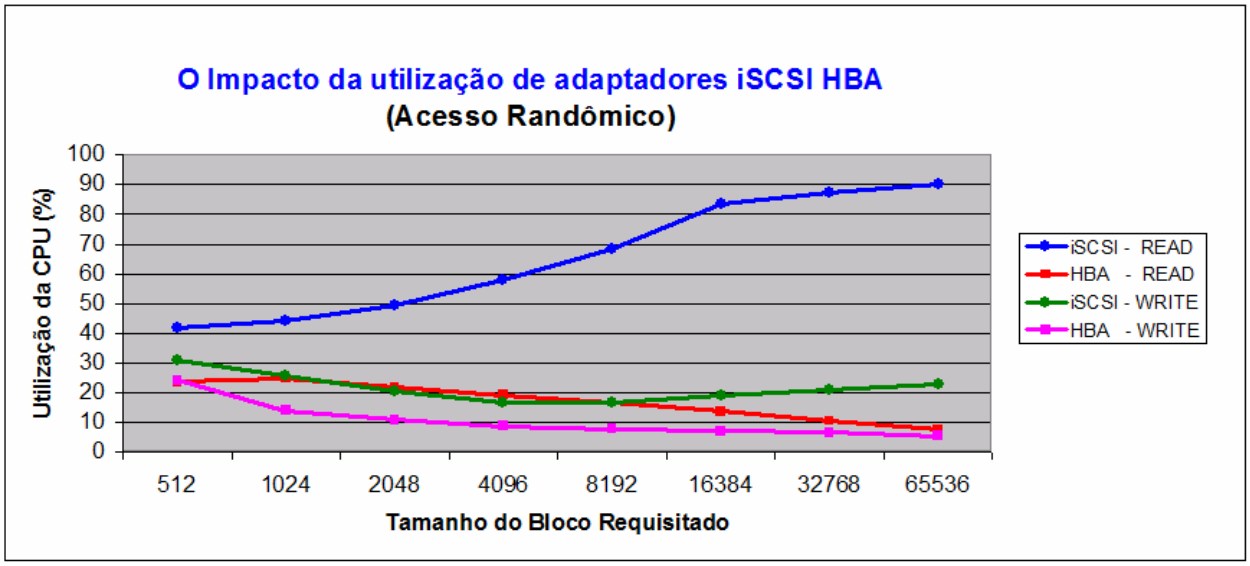


Figura 46 – Cenário 2 – Métrica Utilização da CPU – Acesso Randômico

Como pode ser visto nas Figuras 45 e 46, os resultados obtidos nos experimentos do Cenário 2 (Métrica Utilização da CPU) para acesso sequencial e randômico dos protocolos iSCSI (Software) e iSCSI (HBA).

A Tabela 13 ilustra os resultados dos experimentos do Cenário 2 (Métrica Utilização da CPU), onde os resultados obtidos estão classificados em seções (experimentos) baseados no tipo de requisição de E/S com perfil de acesso correspondente. Cada coluna possui oito (8) valores em função do tamanho de bloco transferido pela Aplicação. O resultado final de cada experimento corresponde ao resultado da operação mediana [33] dos valores obtidos de cada seção correspondente.

Cenário 2 (Resultados)								
Impacto da utilização de adaptadores iSCSI HBA								
Métrica Utilização da CPU (% CPU)								
Tamanho Bloco Requisitado	Protocolo iSCSI (Software)				Protocolo iSCSI (HBA)			
	Sequencial READ	Sequencial WRITE	Randômico READ	Randômico WRITE	Sequencial READ	Sequencial WRITE	Randômico READ	Randômico WRITE
512	62.865665	61.525007	41.662019	31.081730	29.891919	50.060051	23.704524	24.164844
1024	67.229053	61.194446	44.418904	25.735051	28.312991	34.312695	24.794557	14.080518
2048	74.520236	56.710700	49.448646	20.485362	25.021844	20.813375	21.734965	10.851950
4096	73.518164	47.347277	57.826320	16.867756	21.500065	18.155325	19.313690	8.899292
8192	88.315293	41.303010	68.421222	16.857806	17.686481	16.892372	16.790063	7.805621
16384	94.261697	34.980084	83.683296	19.238506	13.993448	11.724102	13.729020	7.115928
32768	92.830573	31.724098	87.151165	21.100504	10.241477	8.489102	10.513678	6.622255
65536	97.124376	30.982236	90.098462	22.662323	8.014581	7.493536	7.754400	5.554726
	MEDIANA							
	81.4177645	44.3251435	63.123771	20.792933	19.593273	17.5238485	18.0518765	8.3524565

Tabela 13 – Cenário 2 (Utilização da CPU) – Resultados

Com base na Tabela 13, o protocolo iSCSI (Software) obteve um perfil geral de Utilização da CPU de 81.4177645% em operações de READ com acesso sequencial, 44.3251435% em operações de WRITE com acesso sequencial, 63.123771% em operações de READ com acesso randômico e 20.792933% em operações de WRITE com acesso randômico.

O protocolo iSCSI (HBA) obteve um perfil geral de Utilização da CPU de 19.593273% em operações de READ com acesso sequencial, 17.5238485% em operações de WRITE com acesso sequencial, 18.0518765% em operações de READ com acesso randômico e 8.3524565% em operações de WRITE com acesso randômico.

Observando os dados obtidos nas Figuras 45 a 47 e na Tabela 13 pode-se concluir que o protocolo iSCSI (HBA) obteve menor Utilização da CPU do que a obtida pelo protocolo iSCSI (Software). Essa diferença de Utilização da CPU é de aproximadamente 75.93% em operações de READ com acesso sequencial, 60.47% em operações de WRITE com acesso sequencial, 71.40% em operações de READ com acesso randômico e 59.83% em operações de WRITE com acesso randômico do protocolo iSCSI (HBA) em relação ao protocolo iSCSI (Software).

A utilização do protocolo iSCSI (HBA) demonstrou resultados gerais expressivos quanto a redução de processamento na CPU do Servidor de Aplicação em relação a utilização do protocolo iSCSI (Software).

Como mencionado anteriormente, o protocolo iSCSI (HBA) utiliza um adaptador de rede Gigabit Ethernet composto de um processador especial, com a principal finalidade de executar o processamento de todos os pacotes TCP/IP e quadros Ethernet, executando um ***offload*** de processamento para a CPU do Servidor de Aplicação. Entretanto, a utilização do protocolo iSCSI (Software) torna necessário a execução de todo o processamento dos pacotes TCP/IP pela CPU do Servidor de Aplicação, gerando, assim, um acréscimo significativo da Utilização da CPU no Servidor de Aplicação.

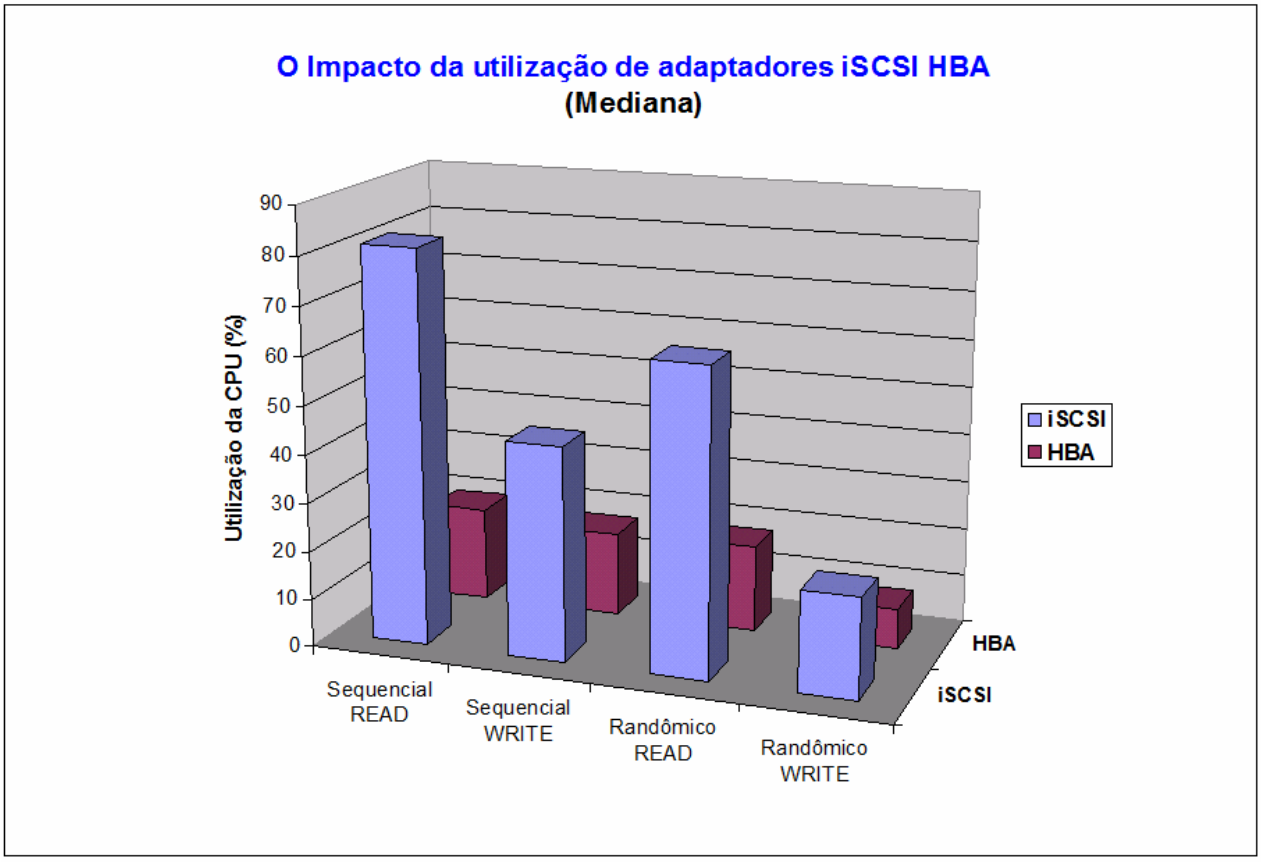


Figura 47 – Cenário 2 (Métrica Utilização da CPU) – Resultados Mediana

5.2.3 Comparação dos protocolos iSCSI (HBA) e Fibre Channel sujeitos à carga heterogênea

O terceiro experimento avalia as métricas Vazão e Utilização da CPU quando se utiliza os protocolos iSCSI (HBA) e Fibre Channel sujeitos a requisições heterogêneas com variações de 67% e 33% de operações de READ e WRITE com acesso randômico às informações localizadas no Dispositivo de Armazenamento.

5.2.3.1 Métrica Vazão – Acesso Randômico

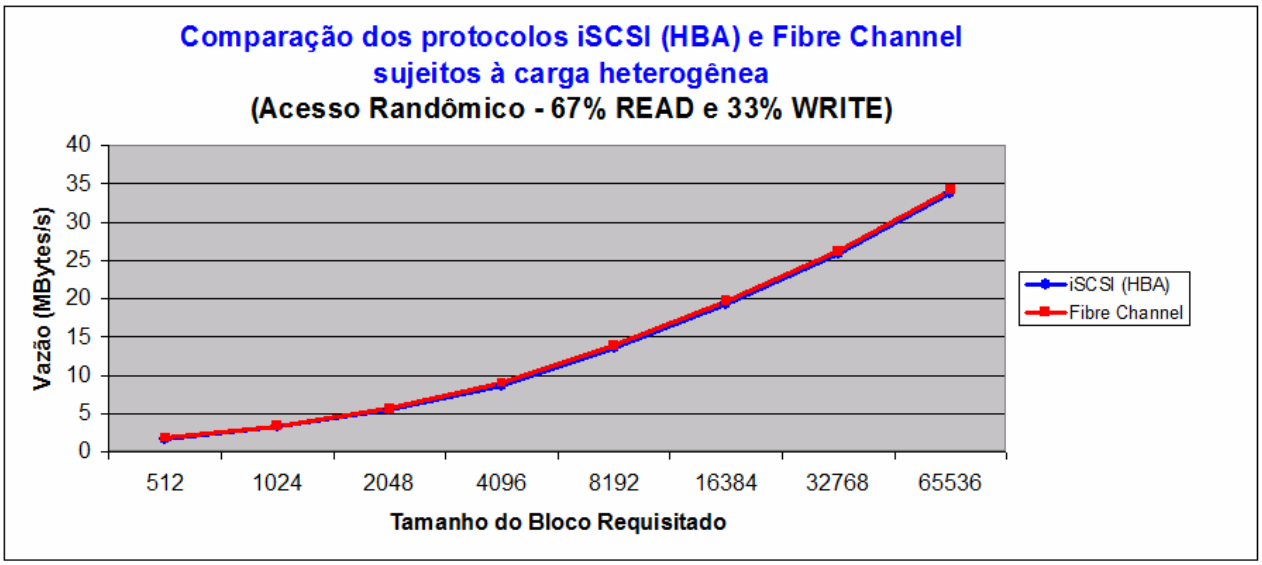


Figura 48 – Cenário 3 – Métrica Vazão – 67% READ e 33% WRITE

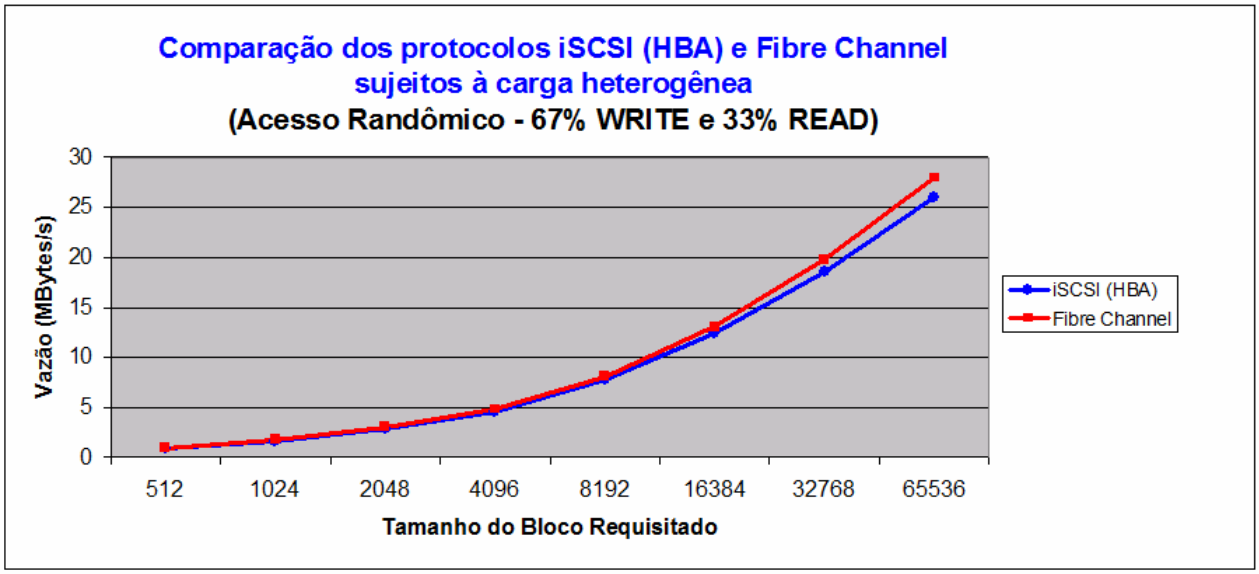


Figura 49 – Cenário 3 – Métrica Vazão – 67% WRITE e 33% READ

Como pode ser visto nas Figuras 48 e 49, os resultados obtidos nos experimentos do Cenário 3 (Métrica Vazão) sujeitos a requisições heterogêneas com variações de 67% e 33% de operações de READ e WRITE de acesso randômico dos protocolos iSCSI (HBA) e Fibre Channel (FCP).

A Tabela 14 ilustra os resultados dos experimentos do Cenário 3 (Métrica Vazão), onde os resultados obtidos estão classificados em seções (experimentos) baseados no tipo de requisição de E/S com perfil de acesso correspondente. Cada coluna possui oito (8) valores em função do tamanho de bloco transferido pela Aplicação. O resultado final de cada experimento corresponde ao resultado da operação mediana [33] dos valores obtidos de cada seção correspondente.

Cenário 3 (Resultados)				
Comparação dos protocolos iSCSI (HBA) e Fibre Channel sujeitos à carga heterogênea				
Métrica Vazão (MBytes/s)				
Tamanho Bloco Requisitado	Protocolo iSCSI (HBA)		Protocolo Fibre Channel	
	67% READ 33% WRITE	67% WRITE 33% READ	67% READ 33% WRITE	67% WRITE 33% READ
512	1.699026	0.84595	1.832145	0.931568
1024	3.256789	1.660799	3.378368	1.805601
2048	5.432719	2.863869	5.670795	3.063407
4096	8.611275	4.562858	9.016252	4.841952
8192	13.604280	7.805702	13.932805	8.127644
16384	19.293149	12.450435	19.732115	13.045668
32768	25.855244	18.589694	26.219523	19.823068
65536	33.774950	26.009260	34.221501	27.978388
	MEDIANA			
	11.1077775	6.18428	11.4745285	6.484798

Tabela 14 – Cenário 3 (Vazão) – Resultados

Com base na Tabela 14, o protocolo iSCSI (HBA) obteve um perfil geral de desempenho de 11.1077775 MBytes/s em operações com 67% READ e 33% WRITE em acesso randômico e 6.18428 MBytes/s em operações com 67% WRITE e 33% READ em acesso randômico.

O protocolo Fibre Channel (FCP) obteve um perfil geral de desempenho de 11.4745285 MBytes/s em operações com 67% READ e 33% WRITE em acesso randômico e 6.484798 MBytes/s em operações com 67% WRITE e 33% READ em acesso randômico.

Observando os dados obtidos nas Figuras 48 a 50 e na Tabela 14, pode-se concluir que a vazão obtida pelo protocolo iSCSI (HBA) é semelhante a vazão obtida pelo protocolo Fibre Channel. A diferença de vazão é de aproximadamente: 0.37 MBytes/s em operações com 67% READ e 33% WRITE de acesso randômico e 0.30 MBytes/s em operações com 67% WRITE e 33% READ de acesso randômico em relação ao protocolo Fibre Channel (FCP) na métrica Vazão.

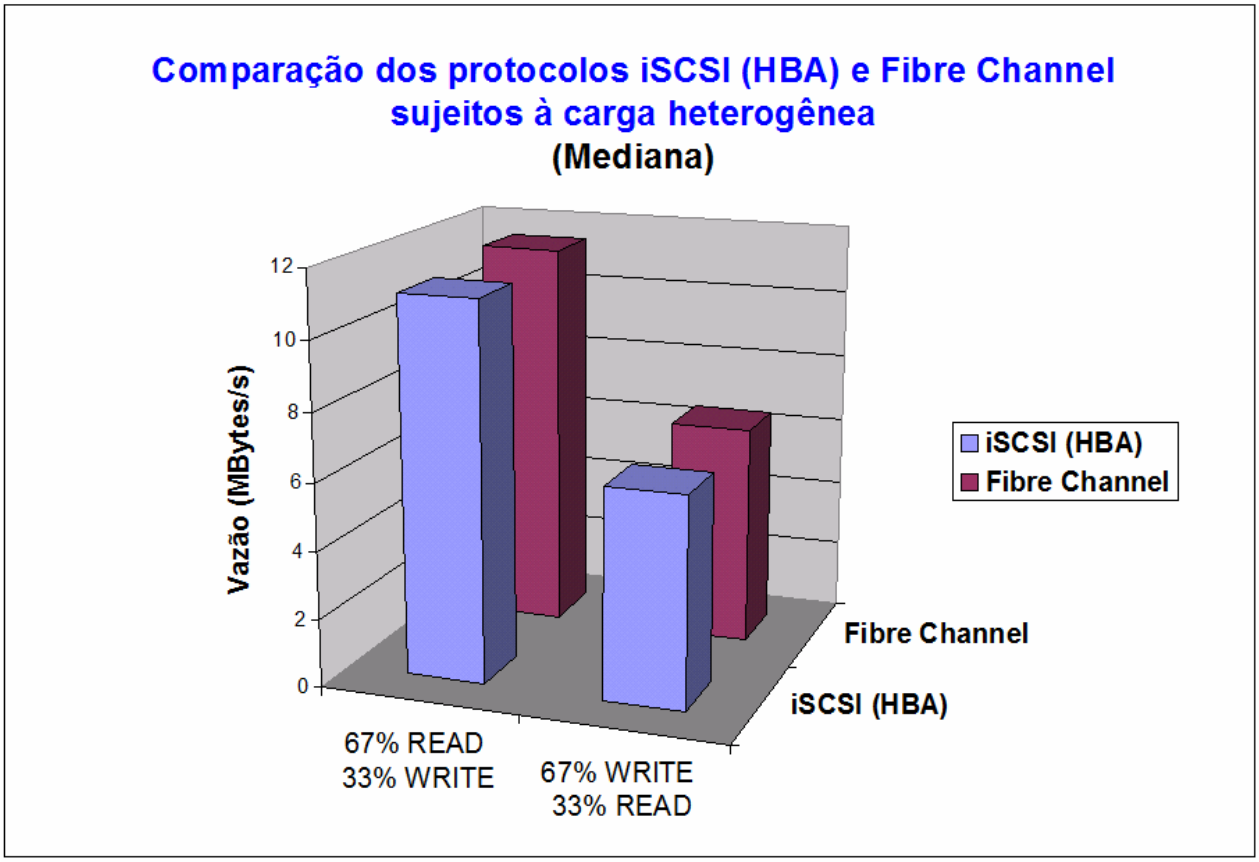


Figura 50 – Cenário 3 (Métrica Vazão) – Resultados Mediana

5.2.3.2 Métrica Utilização da CPU – Acesso Randômico

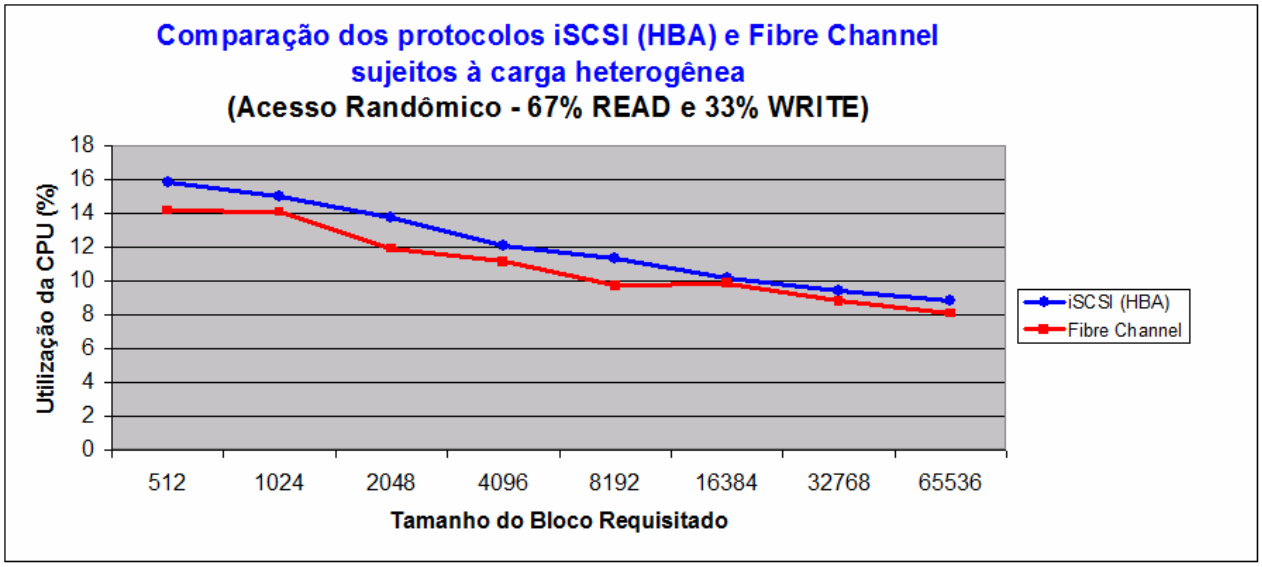


Figura 51 – Cenário 3 – Métrica Utilização da CPU – 67% READ e 33% WRITE

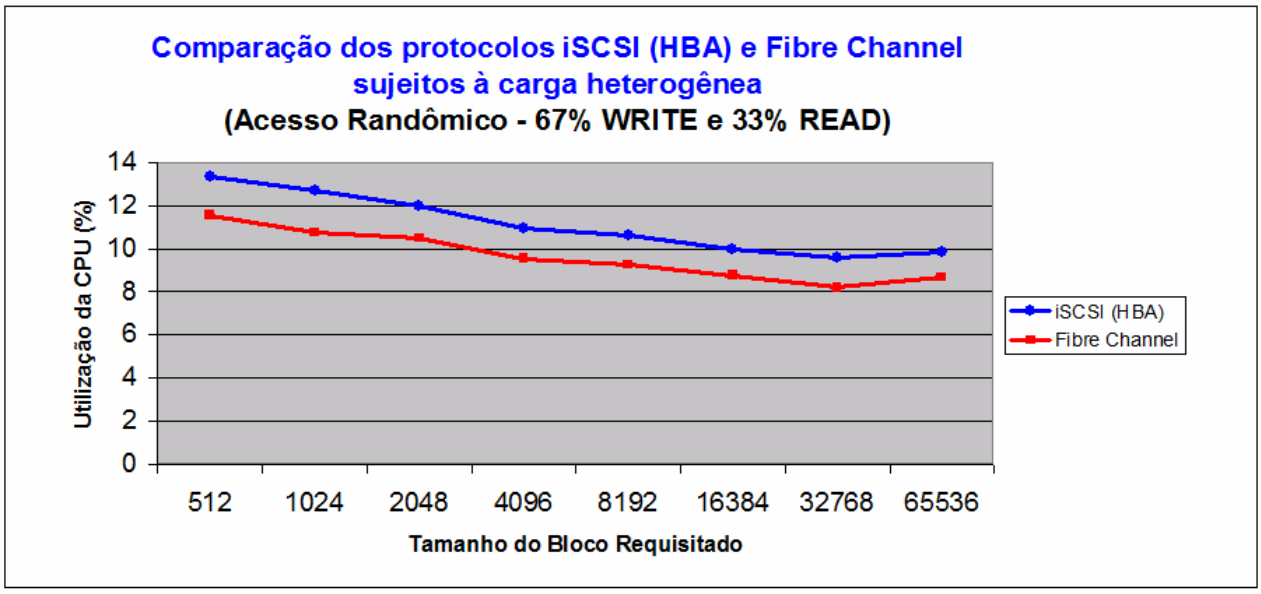


Figura 52 – Cenário 3 – Métrica Utilização da CPU – 67% WRITE e 33% READ

Nas Figuras 51 e 52, podem ser vistos os resultados obtidos nos experimentos do Cenário 3 (Métrica Utilização da CPU) para acesso sequencial e randômico dos protocolos iSCSI (HBA) e Fibre Channel (FCP).

A Tabela 15 ilustra os resultados dos experimentos do Cenário 3 (Métrica Utilização da CPU), onde os resultados obtidos estão classificados em seções (experimentos) baseados no tipo de requisição de E/S com perfil de acesso correspondente. Cada coluna possui oito (8) valores em função do tamanho de bloco transferido pela Aplicação. O resultado final de cada experimento corresponde ao resultado da operação mediana [33] dos valores obtidos de cada seção correspondente.

Cenário 3 (Resultados)				
Comparação dos protocolos iSCSI (HBA) e Fibre Channel sujeitos à carga heterogênea				
Métrica Utilização da CPU (% CPU)				
Tamanho Bloco Requisitado	Protocolo iSCSI (HBA)		Protocolo Fibre Channel	
	67% READ 33% WRITE	67% WRITE 33% READ	67% READ 33% WRITE	67% WRITE 33% READ
512	15.820796	13.319963	14.16905	11.557071
1024	14.988509	12.694022	14.089476	10.764757
2048	13.763727	11.969861	11.915494	10.491937
4096	12.084738	10.980276	11.160526	9.539962
8192	11.303762	10.637475	9.729904	9.259749
16384	10.159659	9.983515	9.846243	8.765091
32768	9.442568	9.609025	8.818279	8.206855
65536	8.845796	9.866918	8.089897	8.675831
	MEDIANA			
	11.69425	10.8088755	10.5033845	9.3998555

Tabela 15 – Cenário 3 (Utilização da CPU) – Resultados

Com base na Tabela 15, o protocolo iSCSI (HBA) obteve um perfil geral de Utilização da CPU de 11.69425% em operações com 67% READ e 33% WRITE em acesso randômico e 10.8088755% em operações com 67% WRITE e 33% READ em acesso randômico.

O protocolo Fibre Channel (FCP) obteve um perfil geral de Utilização da CPU de 10.5033845% em operações com 67% READ e 33% WRITE em acesso randômico e 9.3998555% em operações com 67% WRITE e 33% READ em acesso randômico.

Observando os dados obtidos nas Figuras 51 a 53 e na Tabela 15, pode-se concluir que a utilização da CPU gerada pelo protocolo iSCSI (HBA) difere pouco da utilização detida pelo protocolo Fibre Channel (FCP). A diferença de Utilização da CPU é de aproximadamente, 10.18% em operações com 67% READ e 33% WRITE em acesso randômico e 13.04% em operações com 67% WRITE e 33% READ em acesso randômico do protocolo iSCSI (HBA) em relação ao protocolo Fibre Channel (FCP).

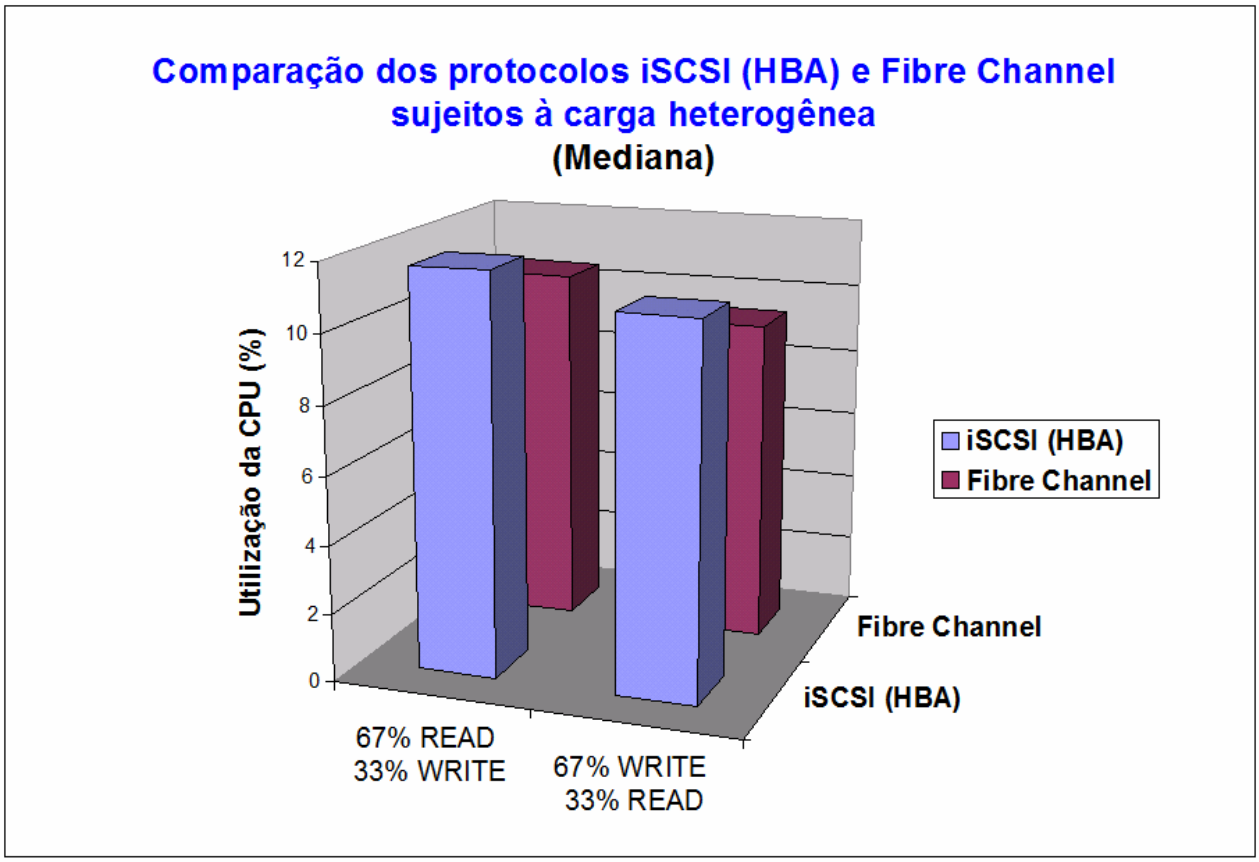


Figura 53 – Cenário 3 (Métrica Utilização da CPU) – Resultados Mediana

Capítulo 6

6.1 Conclusões

O protocolo iSCSI baseia-se em uma infraestrutura de rede IP, com isso tornando-o dependente de protocolos das camadas superiores para prover uma camada confiável para a entrega das mensagens SCSI ao Dispositivo de Armazenamento. O protocolo iSCSI está sujeito ao processamento e aos *overheads* introduzidos pela pilha TCP/IP e enquadramento Ethernet.

Dado a penetração deste protocolo nos dias de hoje, pode-se concluir que as corporações ainda estão inseguras quanto a adoção e utilização dessa arquitetura iSCSI (SAN IP) em relação a uma arquitetura mais consolidada e amplamente utilizada como a arquitetura Fibre Channel (SAN FCP) em redes de armazenamento de dados.

A grande maioria das corporações possuem um fator de escolha preconcebido favorável à adoção e utilização da arquitetura Fibre Channel, devido a diversos fatores relacionados como quadro sem nenhum *overhead* adicional ao seu enquadramento padrão, possibilidade de utilização de sequências de quadros, possibilidade de implementação de classes de serviços e possibilidade de implementação de enlaces a 2Gbits/s de desempenho.

A decisão pelas corporações quanto a escolha de uma arquitetura SAN IP ou uma arquitetura SAN FCP deve estar relacionado diretamente com a necessidade das Aplicações de Sistema, bem como o desempenho desses protocolos. No entanto, pelo resultados obtidos, nota-se que há poucas diferenças de desempenho entre os protocolos iSCSI e Fibre Channel (FCP).

Por exemplo, antes da padronização do protocolo iSCSI, se uma corporação necessitasse implementar uma rede de armazenamento de dados e atender as necessidades de uma Aplicação de Sistema, como por exemplo:

- ✓ **Aplicação de Sistema** – Correio Eletrônico
- ✓ **Tamanho de Bloco** – 4096 bytes
- ✓ **100% Requisições READ –** Vazão aproximadamente de 25MBytes/s
- ✓ **100% Requisições WRITE –** Vazão aproximadamente de 6MBytes/s
- ✓ **Acesso –** Randômico

A única alternativa existente para as corporações era basear em uma infraestrutura e protocolo Fibre Channel, implementando assim, uma arquitetura SAN FCP.

Entretanto, com a abrangência das redes TCP/IP, enlaces Gigabit Ethernet, e padronização do protocolo iSCSI as corporações atualmente possuem uma sólida alternativa à adoção da arquitetura SAN FCP. Essa alternativa engloba uma implementação de uma rede de armazenamento de dados utilizando uma infraestrutura IP (Gigabit Ehernet) e no protocolo iSCSI implementando, assim, uma arquitetura SAN IP. Além disso, o desempenho obtido do protocolo iSCSI, em todos os experimentos, demonstrou uma equivalência de comportamento em relação ao protocolo Fibre Channel.

Conforme pode ser ilustrado na Tabela 16, para a implementação de uma Aplicação de Sistema (Correio Eletrônico), o protocolo iSCSI (Software) atingiu taxas de 31.479853 MBytes/s em operações de READ e 7.872376 MBytes/s em operações de WRITE; o protocolo Fibre Channel atingiu taxas de 38.97383 MBytes/s em operações de READ e 7.695974 MBytes/s em operações de WRITE e o protocolo iSCSI HBA atingiu taxas de 27.38291 MBytes/s em operações de READ e 6.53953 MBytes/s em operações de WRITE. Portanto a Vazão necessária para a implementação e funcionamento da Aplicação de Sistema (Correio Eletrônico) que é de 25 MBytes/s (READ) e 6 MBytes/s (WRITE) pode ser atendido por todos os protocolos. (Tabela 16)

Aplicação de Sistema – Correio Eletrônico					
Tamanho de Bloco – 4096 bytes					
100% Requisições READ – Vazão aproximadamente de 25MBytes/s					
100% Requisições WRITE – Vazão aproximadamente de 6MBytes/s					
Acesso – Randômico					
Métrica Vazão					
iSCSI (Software)		**Fibre Channel**		**iSCSI HBA**	
READ	**WRITE**	**READ**	**WRITE**	**READ**	**WRITE**
31.479853	7.872376	38.97383	7.695974	27.38291	6.53953

Tabela 16 – Aplicação de Sistema (Correio Eletrônico) - Vazão

Entretanto, um fator importante a ser analisado na adoção de uma infraestrutura SAN IP é a Utilização da CPU. Conforme ilustra a Tabela 17, a Utilização da CPU para essa determinada Aplicação de Sistema (Correio Eletrônico), entre as arquiteturas iSCSI HBA e Fibre Channel possuem uma equivalência de valores de aproximadamente 25% de utilização da CPU em operações de READ e de aproximadamente 9% de Utilização da CPU em operações de WRITE.

Porém, com a adoção e utilização da arquitetura iSCSI (Software) a Utilização de CPU foi de aproximadamente 58% para operações READ e de aproximadamente 16% de utilização da CPU para operações WRITE.

Aplicação de Sistema – Correio Eletrônico					
Tamanho de Bloco – 4096 bytes					
100% Requisições READ – Vazão aproximadamente de 25MBytes/s					
100% Requisições WRITE – Vazão aproximadamente de 6MBytes/s					
Acesso – Randômico					
Métrica Utilização da CPU					
iSCSI (Software)		**Fibre Channel**		**iSCSI HBA**	
READ	**WRITE**	**READ**	**WRITE**	**READ**	**WRITE**
57.82632	16.86775	26.63936	8.512193	19.31369	8.89929

Tabela 17 – Aplicação de Sistema (Correio Eletrônico) – Utilização da CPU

A escolha de uma arquitetura iSCSI (Software) ou iSCSI (Hardware) está diretamente relacionada com a utilização de CPU do Servidor de Aplicação. Se a utilização da CPU no Servidor de Aplicação for viável em termos de manutenção da Qualidade de Serviço (QoS), a adoção da arquitetura iSCSI (Software) é mais recomendada. Se o aumento da utilização da CPU não for viável para a manutenção da Qualidade de Serviço, a recomendação é a utilização de adaptadores iSCSI HBA ou iSCSI TOE para a execução de *offload* de processamento.

Portanto com base nos experimentos realizados e na análise de comportamento de desempenho do protocolo iSCSI em relação ao protocolo Fibre Channel, esse "fator de escolha" adotado pelas corporações deve ser reavaliado, pois em todos os cenários o protocolo iSCSI demonstrou um fator de desempenho insignificativamente diferente ao desempenho do protocolo Fibre Channel.

O protocolo iSCSI apresenta também inúmeras vantagens em relação a custos de aquisição e operação, utilização de uma infraestrutura TCP/IP e a possibilidade de interoperabilidade com dispositivos conectados a Internet, podendo em um futuro próximo se tornar um padrão "de facto" para as redes de armazenamento de dados.

6.2 Trabalhos Futuros

Sugere-se como futuros desenvolvimentos e pesquisas sobre esse Estudo de Desempenho dos Protocolos iSCSI e Fibre Channel (FCP): Uma Análise de Aplicações Comerciais com base nos Experimentos realizados, em uma análise do comportamento geral do desempenho dos protocolos iSCSI e Fibre Channel com base em *benchmarks* padrões à nível da aplicação de sistema, como por exemplo:

- ✓ TPC-C : simulação ambientes OLTP (*Online Transaction Processing*);
- ✓ TPC-H e TPC-R : simulação ambientes DSS (*Decision Support System)*;
- ✓ TPC-W : simulação ambientes WEB (aplicações de e-commerce);

Sugere-se também uma Análise de Comportamento Geral do Desempenho dos Protocolos iSCSI e Fibre Channel com base em *benchmarks* específicos para cada aplicação de sistema (fabricante), como por exemplo:

- ✓ Microsoft Corporation: utilização de ferramentas específicas para simulação de ambientes de Banco de Dados (Microsoft SQL Server) e Correio Eletrônico (Exchange);
- ✓ Oracle Corporation: utilização de ferramentas específicas para simulação de ambientes de Banco de Dados (Oracle Database);
- ✓ IBM: utilização de ferramentas específicas para simulação de ambientes de Banco de Dados (IBM DB2);

A realização de experimentos como estes permitirá uma análise de desempenho específica da utilização do protocolo iSCSI (Software) ou iSCSI (Hardware) para aplicações de sistema.

Lista de Acrônimos

CDB	Command Descriptor Block
DAS	Direct Attached Storage
FC	Fibre Channel
FCP	Fibre Channel Protocol
HBA	Host Bus Adapter
iSCSI	internet Small Computer Systems Interface
JBOD	Just a Bunch of Disks
LU	Logical Unit
LUN	Logical Unit Number
RAID	Redundant Arrays of Independent Disks
SAM	SCSI Architecture Model
SAN FCP	Storage Area Network using Fibre Channel Protocol
SAN IP	Storage Area Network using IP Protocol
SAN	Storage Area Network
SCSI	Small Computer Systems Interface
TOE	TCP Offload Engine

Referências Bibliográficas

[1] "How Much Information? 2003"
http://www.sims.berkeley.edu/research/projects/how-much-info/

[2] "Storage - New Game New Rules"
A Paper by Fred Moore, Horison Information Strategies
StorageTek

[3] D. A. Patterson and G. Gibson and R. H. Katz, "A Case for Redundant Arrays of Inexpensive Disks (RAID)", Proc ACM SIGMOD Conf, Chicago, IL, Jun 1988.

[4] D. A. Patterson and P. Chen and G. Gibson and R. H. Katz, "Introduction to Redundant Arrays of Inexpensive Disks (RAID)", IEEE COMPCON 89, San Francisco, Feb-Mar 1989.

[5] IDC Market
http://www.netapp.com/news/press/2003/news_rel_20030606.html

[6] Technical Comittee T10
http://www.t10.org

[7] SCSI-1 Model – Documento X3.131-1986
http://www.t10.org/x3t9_2.htm

[8] SCSI-2 Model – Documento X3.131-1994
http://www.t10.org/drafts.htm

[9] SCSI-3 Model – Documento X3.270-1996
http://www.t10.org/scsi-3.htm

[10] Preston, W. Curtis, “Using SANs and NAS”, O’Reilly & Associates, 2002

[11] Cisco Systems; “iSCSI Protocol Concepts and Implementation”; White Paper; http://www.cisco.com

[12] Gupta, Meeta, “Storage Area Network Fundamentals”, Cisco Press, 2002

[13] Farley, Marc, “Building Storage Networks”, Osborne/McGraw Hill, 2002

[14] Clark, Tom, “IP SANs – A Guide to iSCSI, iFCP and FCIP Protocols for Storage Area Networks”, Addison-Wesley, 2001

[15] Clark, Tom, “Designing Storage Area Networks – A Practical Reference for Implementing Fibre Channel and IP SANs”, Second-Edition, Addison-Wesley, 2003

[16] J. Satran, K. Meth, C. Sapuntzakis, M. Chadalapaka, E. Zeidner – iSCSI (Internet Small Computer Systems Interface) – RFC 3720

[17] Orenstein, Gary, “IP Storage Area Network – Straight to the Core”, Addison-Wesley, 2003

[18] Toigo, Jon William, “The Holy Grail of Network Storage Management”, Prentice Hall PTR, 2003

[19] Tecnologia Gigabit Ethernet – IEEE 802.3z

[20] Kurose, James F. and Ross, Keith W., “Computer Networking – A Top-Down Approach Featuring the Internet”, Third Edition, Pearson/Addison-Wesley, 2005

[21] Forozan, Behrouz A.; “Data Comunications and Networking”, Second Edition Update, McGraw-Hill Higher Education, 2001

[22] Satran, Julian; Meth, Kalman Z.; “Design iSCSI Protocol”, IBM Haifa Research Laboratory, Proceedings of the 20 th IEEE/11 th NASA Goddard Conference on Mass Storage Systems and Technologies (MSS’03)

[23] Hufferd, John L.; “iSCSI – The Universal Storage Connection”, Addison-Wesley, 2003

[24] Aiken, Stephen; Grunwald, Dirk; Pleskun, Andrew R.; Willeke, Jesse; “A Performance Analysis of the iSCSI Protocol”; Colorado Center for Information Storage, University of Colorado; Proceedings of the 20 th IEEE/11 th NASA Goddard Conference on Mass Storage Systems and Technologies (MSS’03)

[25] IBM RedBooks; “Using iSCSI – Solutions’ Planning and Implementation”; Hernandez, Rowell; Carmichael, Keith; Malen, Diana; Moore, Barry; Lane, George; Earhart, Joe; IBM Corporation, 2002

[26] Spectra Logic; “IP SAN or Fibre Channel SAN?”, A Benchmark Test Comparing Backups of Microsoft Exchange® Server; White Paper; 2003
http://www.spectralogic.com

[27] Microsoft Corporation; “Microsoft Support for iSCSI”; White Paper; 2003
http://www.microsoft.com

[28] Network Appliance, Inc.
http://www.netapp.com

[29] Iometer User’s Guide; SourceForge Project
http://sf.net/projects/iometer

[30] Alacritech and Nishan Systems; “Achieving Wire-Speed iSCSI Performance”; Technical Note; 2001
http://www.nishan.com
http://www.alacritech.com

[31] Network Computing; “Don’t Sink Your IPSAN”; White Paper; 2003
http://www.nwc.com

[32] Cisco Systems; “SN 5420 Storage Router Performance Benchmark”; White paper; 2002
http://www.cisco.com

[33] Simitci, Huseyin; “Storage Network Performance Analysis”; Wiley Publishing; 2003

[34] Alacritech Inc; “Alacritech iSCSI SAN Performance White Paper”; White Paper
http://www.alacritech.com/html/iscsi_hba_SAN_paper.html

[35] Schmisseur, Mark; “RAID Benchmarking Workload Analysis”; InfoStor; 2000