Instituto de Física da Universidade de São Paulo

DEPARTAMENTO DE FÍSICA EXPERIMENTAL

Relatório Anual de Atividades
Programa PROCONTES
Processo Nº 03.1.27082.1.0

Rogério Luiz Iope

Nº USP 1712076

Dezembro 2006

Índice

1. Introdução

A Física de Altas Energias tem como principal objetivo desvendar a estrutura íntima da matéria, buscando determinar quais são os seus constituintes mais fundamentais e como eles interagem entre si. As grandes descobertas nessa área, feitas durante os últimos 100 anos, tornaram-se possíveis graças à construção de aceleradores de partículas cada vez maiores e mais sofisticados. Os aceleradores de partículas são hoje os maiores e mais complexos instrumentos de investigação científica já construídos. Nos aceleradores modernos, partículas estáveis, eletricamente carregadas, são injetadas no interior de um gigantesco duto circular, cujo interior é mantido em alto vácuo, e campos elétricos e magnéticos são usados para acelerá-las e mantê-las em órbitas circulares. Nos experimentos de maior energia, feixes de partículas e anti-partículas são injetados nesses dutos, acelerados em direções opostas e mantidos separados entre si, até o momento em que adquirem a energia cinética de interesse para o experimento. Nesse momento, os feixes que viajam em direções opostas são forçados a se alinhar, e colidem frontalmente nas regiões onde se situam os grandes detectores, nos quais os resultados das colisões são registrados pelos diversos sub-detectores, capazes de identificar as características das inúmeras sub-partículas produzidas durante as colisões. Este modo de operação dos modernos aceleradores recebe comumente a denominação “anel de colisão”.

Anéis de colisão são os principais instrumentos científicos para a investigação das interações fundamentais da Natureza e da estrutura íntima da matéria. O Tevatron do Fermilab e o Large Hadron Collider (LHC) do CERN são os principais exemplos. O anel de colisão do Tevatron possui 6,3 Km de circunferência, sendo o maior acelerador de partículas em operação atualmente. O Tevatron tem dois grandes detectores, CDF e DZero, que são capazes de observar os mais variados eventos ligados à Física de Altas Energias. O anel de colisão do LHC, que tem uma circunferência de 27 Km, está em fase final de construção, e entrará em operação em 2007. O LHC tem quatro grandes detectores: ATLAS, CMS, ALICE e LHCb. Ao entrar em operação, os resultados dos experimentos gerados no LHC desempenharão um papel fundamental para as investigações científicas na área de Física de Altas Energias durante os próximos 20 anos. Estes experimentos deverão produzir uma quantidade de dados sem precedentes, devendo atingir 10^9 GB durante a próxima década. Estes dados terão que ser armazenados, processados e analisados por milhares de pesquisadores ao redor do mundo. Para alcançar este objetivo de forma eficiente, as colaborações internacionais estão desenvolvendo sofisticadas arquiteturas de processamento distribuído, conhecidas como Grids computacionais.

Grids computacionais são ambientes de computação distribuída virtualizados. Tais ambientes buscam permitir a seleção dinâmica (i.e., em tempo de execução), o compartilhamento e a agregação de recursos computacionais autônomos e geograficamente distribuídos, pertencentes a domínios administrativos distintos, baseando-se na disponibilidade, capacidade computacional e de armazenamento, desempenho e custo desses recursos. A idéia de Grids computacionais se assemelha à rede de interconexão entre usinas geradoras e consumidores para a distribuição de energia elétrica, uma complexa malha de cabos elétricos de alta tensão, transformadores e usinas de geração de energia, capaz de levar energia elétrica sob demanda para os consumidores, desde pequenas residências até grandes indústrias siderúrgicas. No caso dos Grids computacionais, a visão é criar organizações virtuais dinâmicas, através do compartilhamento seguro e coordenado de recursos entre indivíduos ou instituições. A computação em Grid adiciona à computação distribuída tradicional não somente o acesso a recursos heterogêneos em localizações dispersas, mas também de organizações distintas, gerenciados por domínios administrativos distintos e independentes, de forma a agregar poder

computacional e de armazenamento, facilitar o trabalho colaborativo e o acesso rápido às informações a todos os usuários conectados a esse sistema.

Aplicações típicas em Grid em geral envolvem a manipulação de grandes volumes de dados ou a necessidade de grande poder computacional e geralmente necessitam de um compartilhamento de recursos seguro entre organizações distintas. Isto não é facilmente conseguido através das infra-estruturas atuais da Internet e da World Wide Web. O conceito chave é a habilidade do 'middleware' de negociar arranjos de compartilhamento de recursos entre um conjunto de entidades participantes, e então usar esse 'pool' de recursos resultantes para algum propósito comum. O compartilhamento de recursos não diz respeito apenas à troca de dados ou arquivos, mas sim o acesso direto aos computadores, software, armazenamento, como é necessário nas estratégias de resolução colaborativa de problemas complexos emergentes na ciência, engenharia e indústria. Esse compartilhamento deve ser, necessariamente, muito bem controlado, com provedores e consumidores de recursos definindo claramente o que exatamente é compartilhado, quem tem permissão de acesso, e as condições nas quais tal compartilhamento ocorre. Um conjunto de indivíduos ou instituições definido por tais regras de compartilhamento é o que se costuma chamar de "Organização Virtual". Enquanto a Internet é uma rede de comunicação de dados, um Grid é uma rede de computação ou processamento de dados: essa tecnologia fornece as ferramentas e protocolos para o compartilhamento de uma grande variedade de recursos de tecnologia de informação.

A computação em Grid teve início com a aplicação simultânea de recursos computacionais de computadores interligados em rede para a solução de um mesmo problema científico. De fato, a comunidade científica, da qual se destacam os físicos de altas energias, os astrônomos e os biomédicos, está liderando o desenvolvimento na área de Grids computacionais, devido à complexidade dos problemas que esses grupos têm que resolver. Os instrumentos científicos estão ficando cada vez maiores, os custos de construção altíssimos, e por isso estão sendo construídos de forma colaborativa por centenas – às vezes milhares – de pesquisadores espalhados ao redor do planeta. O acelerador de partículas LHC do CERN e os imensos telescópios do Projeto GEMINI são exemplos dessa tendência à construção de estruturas gigantescas e complexas por uma comunidade grande de pesquisadores que precisam colaborar entre si, trocar dados e informações, e ao mesmo tempo fazer uso compartilhado de um mesmo instrumento, construído em conjunto. Tais instrumentos, por sua vez, são capazes de capturar informações e gerar imensas massas de dados, que precisam ser analisados por sistemas computacionais de enorme capacidade de processamento, o que só se consegue se a tarefa de processamento for realizada de forma distribuída e colaborativa.

Como membros das colaborações DZero do Fermilab e CMS do CERN, os pesquisadores do São Paulo Regional Analysis Center – SPRACE, uniram-se aos esforços internacionais na implementação de centros regionais de análise distribuídos para o processamento dos dados gerados por estes experimentos. O SPRACE está provendo, desde 2004, os mais variados serviços de processamento de dados científicos, tais como distribuição do código computacional do DZero, simulações de Monte Carlo desse experimento, administração de submissão e execução de processos, acesso ao banco de dados do DZero e reprocessamento e análise dos dados obtidos. Além disso, desde agosto de 2005 este centro regional de processamento uniu-se ao Open Science Grid (OSG) americano, e está totalmente engajado nas iniciativas relacionadas ao Worldwide LHC Computing Grid (WLCG), como site ativo da colaboração CMS do CERN, que deverá estar produzindo dados a partir de 2007.

2. Descrição do SPRACE

2.1. Espaço físico e recursos computacionais

O SPRACE está instalado nas dependências do Centro de Ensino e Pesquisa Aplicada (CEPA) do Departamento de Física Experimental, no Instituto de Física da USP. O coordenador do CEPA, Prof. Dr. Gil da Costa Marques, disponibilizou uma área de aproximadamente 25 m^2, na qual foram feitas as reformas necessárias para a instalação do cluster e da infra-estrutura de rede do SPRACE. O SPRACE foi o pioneiro no processamento de dados em Física de Altas Energias em forma de Grid no Hemisfério Sul. O início das operações ocorreu em um tempo surpreendentemente curto: apenas 4 meses e meio após a concessão do auxílio pela FAPESP o grupo já estava produzindo os primeiros eventos de Monte Carlo para a Colaboração DZero.

O esquema geral de estruturação do espaço físico ocupado pelo centro de processamento do SPRACE é apresentado na Figura 1 a seguir.

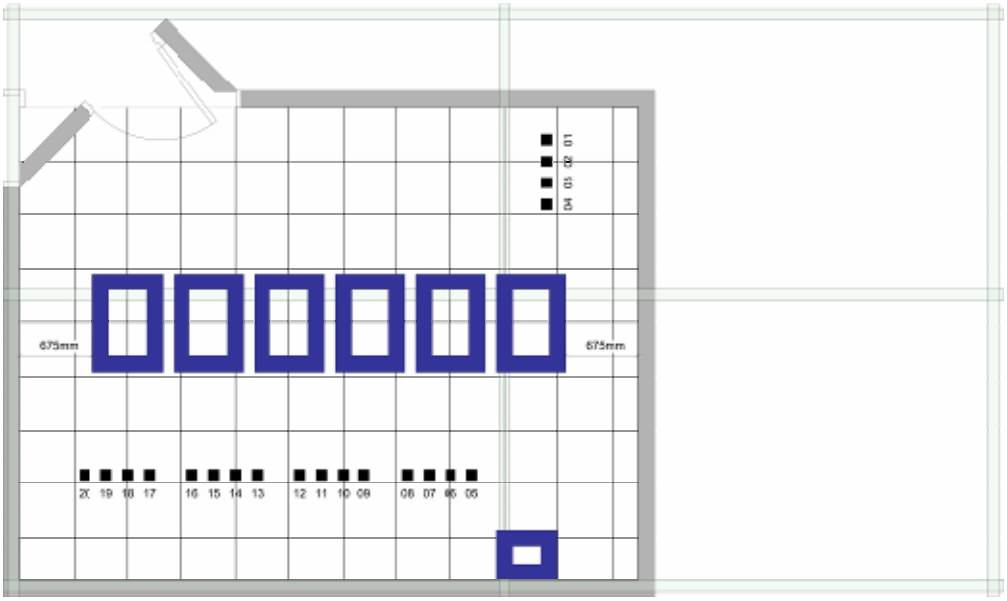

Figura 1: *Espaço físico de aproximadamente 25 m^2 ocupado pelo centro de processamento do SPRACE, ressaltando as posições ocupadas pelos 6 racks que alojam os servidores de processamento e 1 rack menor para os equipamentos de rede. Os pequenos quadros numerados indicam os pontos de tomada elétrica, instaladas na parte de baixo do piso elevado.*

Todo o dimensionamento das instalações físicas, planejado em 2004, foi feito levando-se em conta a implantação do cluster completo com 80 nós de processamento duais. Na verdade, o cluster, agora em sua versão final, conta com 86 nós de processamento e 4 servidores principais. Os servidores principais realizam as funções de 'front-end', 'gatekeeper', controle de armazenamento centralizado, e

controle de 'pools' de armazenamento distribuídos, e os nós de processamento, também servidores duais, são os que efetivamente executam o processamento de jobs, atuando como "number crunchers". Dentre os nós de processamento, os 32 mais recentemente adquiridos incorporam a tecnologia de núcleo duplo ('dual-core'). Logo, a somatória chega a 244 núcleos de processamento, cada um dos quais com 1 GB de memória disponível.

2.2. Infra-estrutura de rede de dados

(a) Link internacional

O CHEPREO (Center for High-Energy Physics Research and Education Outreach) [1], uma colaboração entre Caltech, Florida International University, University of Florida, Florida State University, UERJ, UNESP, USP e FAPESP, é um centro inter-regional criado para viabilizar programas de incentivo à pesquisa e educação, através do aprimoramento da "cyber-infrastructure", ou infra-estrutura de rede. No final de 2004, uma parceria entre o CHEPREO, a FIU e a FAPESP, com apoio financeiro da NSF, permitiu a ampliação da conexão de rede que integra a rede acadêmica do Estado de São Paulo (rede ANSP) à rede Abilene (Internet2) nos EUA. Esta conexão, administrada pela AMPATH, foi ampliada de 45 Mbps para 622 Mbps, através da troca de diversos equipamentos de rede e da renegociação da concessão do link. Posteriormente, este link subiu para 1,2 Gbps, devido ao projeto WHREN-LILA (detalhado adiante). Assim, este link consiste atualmente de dois circuitos STM-4 entre São Paulo e Miami, e faz uso de um cabo óptico submarino provido pela empresa LANautilus [2], do grupo Telecom Itália. Essa ampliação aumentou em quase 28 vezes a velocidade de comunicação das redes acadêmicas do Estado de São Paulo com os Estados Unidos.

Embora esta rede esteja atualmente operando a 1,2 Gbps, ela será em breve atualizada para 2,5 Gbps.. Durante o Bandwidth Challenge do evento SuperComputing 2004, que ocorreu em Pittsburg, esse link foi testado quase ao limite pela primeira vez, como parte da demonstração chamada "High Speed Terabyte Transfers for Physics", coordenada pelo Caltech. Essa demonstração tornou-se possível graças ao trabalho realizado pelo NARA (Núcleo de Apoio à Rede Acadêmica, que administra a rede ANSP) e pelo SPRACE, e ocorreu devido a uma singular união de esforços de várias entidades: o consórcio Internet2 e a empresa Qwest forneceram a conectividade entre Miami e Pittsburg, as empresas Terremark do Brasil e Impsat forneceram espaço físico e suporte técnico em seus data-centers; a Eletropaulo Telecom e a Iqara cederam fibras ópticas em São Paulo; a LANautilus, do grupo Telecom Itália, cedeu capacidade no cabo submarino entre São Paulo e Miami; a FPL (Florida Power & Light Company) FiberNet cedeu fibras ópticas em Miami; Cisco e Foundry supriram os equipamentos de rede necessários para o funcionamento da conexão. O link foi fechado desde o NAP do Brasil em Barueri até o "floor" do evento, e o tráfego atingiu 2 + 1 Gbps (tráfego de chegada + tráfego de saída) e foi sustentado durante 1 hora. Este é o recorde atual de transmissão de dados entre os hemisférios.

O link São Paulo-Miami é bastante estratégico; os principais objetivos de utilização definidos para ele são:

- Estabelecer um "peering" com a rede Abilene (Internet2) e outras redes de pesquisa e educação dos Estados Unidos através da AMPATH, possibilitando conectividade óptica de alta velocidade a todos os grupos de pesquisa do Estado de São Paulo com centros de pesquisa americanos;
- Fornecer conectividade ao backbone WHREN-LILA, um projeto proposto pela FIU e CENIC, e financiado pela NSF, que consiste de um anel óptico abrangendo todo o hemisfério oeste,

interligando as redes acadêmicas da América Latina com importantes pontos de troca de tráfego nos EUA;

- Fornecer conectividade ao backbone “Atlantic-Wave” (A-Wave), similar à iniciativa “Pacific Wave” na costa oeste americana, que prevê a interligação de vários pontos de troca de tráfego internacional por toda a costa leste americana, estendendo-se à Europa pela rede GEANT, e que permitirá futuro provisionamento de “lambdas” (tecnologia WDM).

A rede Abilene [3] foi o primeiro testbed óptico de extensão nacional construído nos Estados Unidos, sobre o qual foi implementada a rede Internet2. É uma corporação sem fins lucrativos, e iniciou-se através de uma sociedade entre as empresas Qwest, Cisco Systems, Nortel Networks e a Indiana University. Atualmente a rede Abilene (Internet2) é formada por um amplo consórcio de universidades americanas, denominado UCAID [4] (University Corporation for Advanced Internet Development). A AMPATH foi a responsável por prover acesso à rede Abilene aos países latino-americanos, através da disponibilização de canais de 45 Mbps para a ANSP em São Paulo, RNP no Rio de Janeiro, RETINA em Buenos Aires, REUNA em Santiago, REACCIUN em Caracas e SENACYT no Panamá. A rede ANSP recebeu o canal de 45 Mbps em março de 2002.

O padrão internacional atual de comunicação óptica está em 10 Gbps. A considerar que até 2004 os pesquisadores no Estado de São Paulo contavam com um canal de 45 Mbps com a rede Abilene americana (menos de 0,5% do padrão internacional), a evolução que está ocorrendo é bastante significativa, pois o mesmo canal já está pronto para operar a 2,5 Gbps (25% do padrão internacional).

É evidente que na América Latina ainda não existe a mesma integração na área de redes ópticas avançadas para pesquisa e educação, como já existe nos EUA, no Canadá, na Europa, e nos países da Ásia Oriental e Austrália, onde já existem há mais tempo fortes enlaces intra-regionais e organizações regionais correspondentes, como a rede Abilene nos EUA, a CA*Net4 [5] no Canadá, a TERENA - Trans-European Research and Education Network [6] na Europa, e a APAN – Ásia-Pacific Advanced Network [7] na região da Ásia e Austrália.

Esta situação começou a mudar em 2005, em virtude de uma iniciativa da NSF, lançada em abril de 2004. Naquele ano, a NSF abriu um programa de financiamento de novas conexões entre as redes avançadas dos EUA e redes semelhantes em outras partes do mundo, denominado IRNC - International Research Network Connections. Este programa dará apoio a grandes projetos de conexão durante 5 anos, com investimentos anuais da ordem de US$ 5 milhões, e envolvem diretamente a Europa, Japão, Austrália, Rússia e América Latina. No caso da América Latina, a proposta de projeto submetida pela FIU e CENIC, em conjunto com a rede paulista ANSP e com as redes de pesquisa nacionais latino-americanas RNP, REUNA, RETINA e CUDI, foi uma das contempladas pelo programa, que foi aprovado em dezembro de 2004.

A proposta submetida pela FIU e CENIC criou uma rede óptica abrangente para todo o hemisfério oeste. O projeto, denominado WHREN-LILA [8] (da sigla em inglês Western Hemisphere Research and Education Network – Links Interconnecting Latin America), foi iniciado oficialmente em maio de 2005, e resultou de uma parceria entre a NSF (Award # 0441095), FAPESP (Projeto N^{o} 04/14414-2), Corporation for Education Initiatives in California (CENIC) e Florida International University (FIU), esta última a responsável pelo projeto AMPATH [9]. Esta rede já está em fase de implantação, e está conectando pontos de troca de tráfego internacional já bem estabelecidos da América do Norte (Miami, FL, Seattle, WA, Los Angeles, CA, Chicago, IL e New York, NY) com pontos de troca de tráfego emergentes na América Latina (São Paulo no Brasil, Santiago no Chile e Tijuana no México). O resultado desse trabalho será uma rede para produção e pesquisa de alta velocidade e alta disponibilidade para as Américas. A proposta WHREN-LILA aumentou de imediato o link São Paulo-Miami para 1,2 Gbps, e adicionou três novos links: um link de 155 Mbps entre São Paulo e Santiago, um

link de 310 Mbps entre Santiago e Tijuana, e um link em fibra apagada entre Tijuana e San Diego, na Califórnia. Durante o período de 5 anos de duração do financiamento da NSF, cada um desses links evoluirá para 2,5 Gbps, de acordo com o projeto proposto.

O diagrama a seguir mostra maiores detalhes da conexão atual entre Miami e São Paulo dentro do contexto do projeto WHREN-LILA. O par de roteadores Cisco ONS 15454 são os elementos fundamentais dessa conexão. O ONS 15454 é um equipamento sofisticado e muito versátil, com suporte a serviços multi-camada e multi-protocolo, além de permitir um aumento incremental da largura de banda do link, através da agregação de múltiplos circuitos STM-4, ou da utilização de circuitos STM-16 (2,5 Gbps). Pode-se ver no diagrama que a configuração adotada fornece conectividade em gigabit-Ethernet exclusiva para a rede ANSP, e conectividade em gigabit-Ethernet compartilhada entre as redes CLARA e RNP.

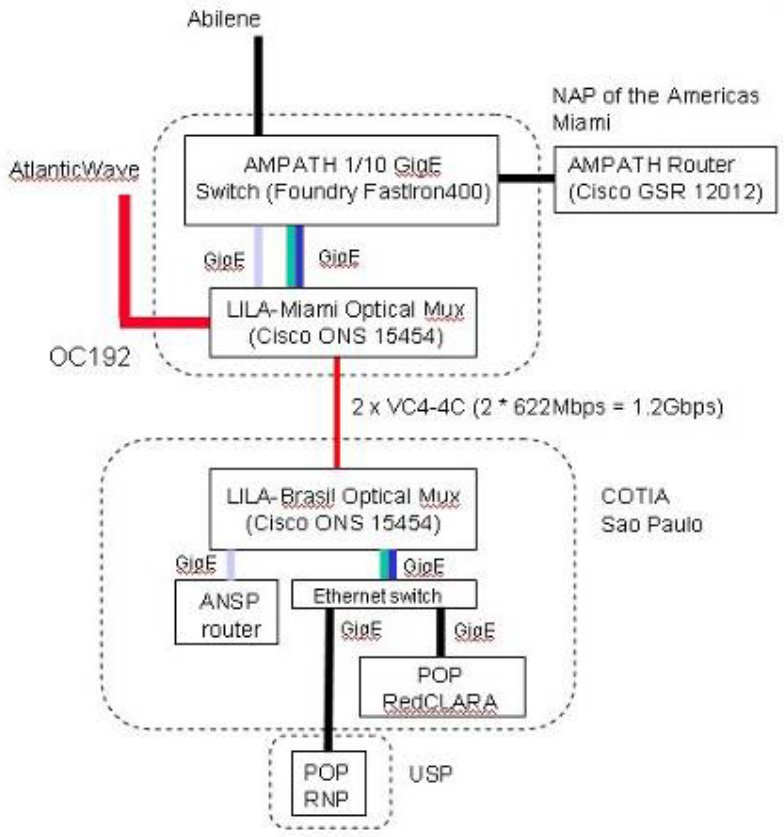

Figura 2: *Conexão São Paulo – Miami definida pelo projeto WHREN-LILA*

Outra iniciativa de relevância é o backbone “Atlantic-Wave” (A-Wave), um “peering” internacional que está interligando Canadá, Estados Unidos, América do Sul e Europa, com o objetivo de fornecer serviços multi-camada e multi-protocolo entre as redes participantes. Dentre as funcionalidades previstas, estão:

- Serviços de “peering” de camada 3 sobre Ethernet com suporte a Jumbo-Frame
- Serviços de comutação de lambdas e provisionamento de “lightpaths” para o GLIF – Global Lambda International Facility.

O GLIF [10] é um laboratório virtual em escala global, estabelecido em 2003, para o desenvolvimento de “middleware” e aplicações para uma área de pesquisa emergente conhecida como lambdaGrids, em que aplicações para Grids computacionais baseiam-se no controle dos comprimentos de onda dos sinais ópticos (“lambdas”) em redes ópticas WDM dinamicamente reconfiguráveis. A comunidade GLIF compartilha a visão de se construir um novo paradigma de redes, que usa a tecnologia de geração e controle de “lightpaths” para dar suporte ao transporte de dados para aplicações em e-Science que demandam grande volume de troca de dados.

Nas Américas, os pontos de distribuição do “Atlantic-Wave” estarão em New York, NY, Washington, DC, Atlanta, GE, Miami, FL, e São Paulo, SP. Em New York e em São Paulo, o anel se fechará com a Europa, através da rede européia GEANT. Novamente, o link São Paulo-Miami desempenha um papel fundamental nessa rede.

Com base no panorama de crescimento das redes ópticas apresentado, pode-se facilmente concluir que São Paulo está emergindo como um grande ponto de troca de tráfego de rede de nível internacional, agregando o backbone da rede CLARA, conectando-se à rede GEANT européia, e participando como ramo importante dos projetos WHREN-LILA e Atlantic-Wave. Neste contexto, a situação do SPRACE (assim como a de qualquer centro de pesquisa ou laboratório do Estado de São Paulo integrado à rede ANSP), é bastante privilegiada, por estar ligado a um ponto de troca de tráfego ao qual estão se conectando redes de alta velocidade trans-nacionais (com os EUA, Chile, México e Europa), financiadas principalmente por projetos acadêmicos. Essa ampliação da conectividade já está mostrando resultados muito positivos, possibilitando ao Brasil participar, pela primeira vez, de eventos internacionais de demonstração de tráfego de rede compatível com padrões internacionais, absolutamente necessários para eventos dessa natureza. O SPRACE tem aproveitado essas oportunidades para testar e aprimorar as configurações de seus equipamentos, e estreitar os laços de cooperação mútua com a comunidade internacional.

Em virtude das transformações advindas da ampliação do link entre São Paulo e Miami, o laboratório SPRACE viu-se obrigado a atualizar sua infra-estrutura de rede. Graças a uma importante doação do Caltech, que ocorreu em maio de 2005, o laboratório conta hoje com um switch/router Cisco Catalyst 3750, com software EMI (“Enhanced Multilayer Image”), que dá suporte a roteamento IP dinâmico completo. A doação também incluiu um par de conversores ópticos (SFPs 1000BASE-LX e 1000BASE-ZX), que permitiram conectar o switch diretamente aos roteadores da rede ANSP, situados a uma distância de quase 70 Km. O switch Catalyst 3750 é o “default gateway” de toda a rede. Um bloco de endereços (200.136.80.0/24) foi disponibilizado pela FAPESP para uso do laboratório, e estabelecida uma rede ponto-a-ponto entre o laboratório e o roteador de entrada da rede ANSP em Barueri. Com esta alteração, o switch Catalyst tornou-se o equipamento responsável pelo roteamento de tráfego entre as seguintes redes:

- Rede 200.136.80.0/24 , à qual estão ligados os servidores do laboratório
- Rede 143.108.254.240/30 , link ponto-a-ponto ligado diretamente à rede ANSP (e, conseqüentemente, ao link internacional)
- Rede 10.24.46.0/24 , ligada aos grupos de pesquisa em física de altas energias no Rio de Janeiro (rede GIGA)

O switch Catalyst é também o responsável pela troca de tráfego da rede interna ao cluster do SPRACE, interligando os switches instalados em cada rack, servidores de ‘front-end’ e ‘gatekeeper’ e servidor de armazenamento. O diagrama a seguir ilustra a configuração atual da rede do laboratório (evidenciando o switch Catalyst como o equipamento central da rede):

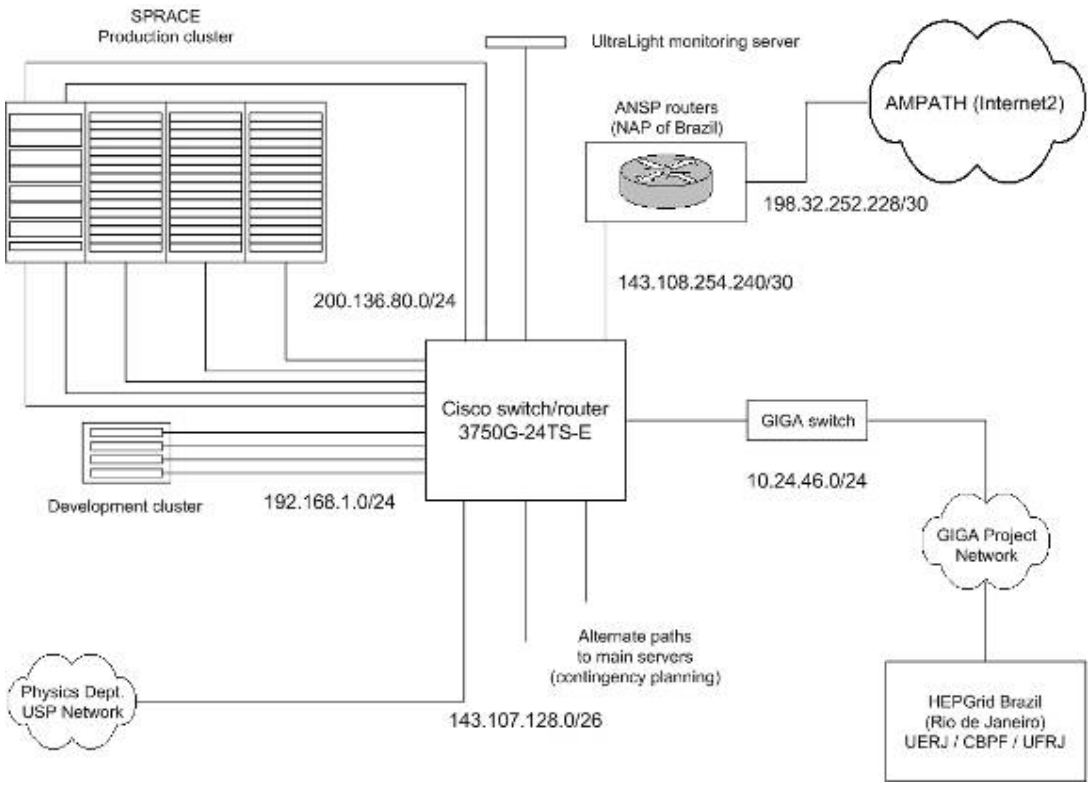

Figura 3: *Configuração de rede do laboratório SPRACE*

O switch Catalyst também fornece informações de tráfego para o sistema de monitoração do projeto UltraLight, via SNMP. Esse sistema de monitoração é baseado no software MonALISA, desenvolvido pelo Caltech [11].

O diagrama a seguir ilustra a conexão entre o cluster no laboratório SPRACE e os roteadores da rede ANSP, e detalha a conexão com a UERJ pela rede GIGA, através do switch Catalyst.

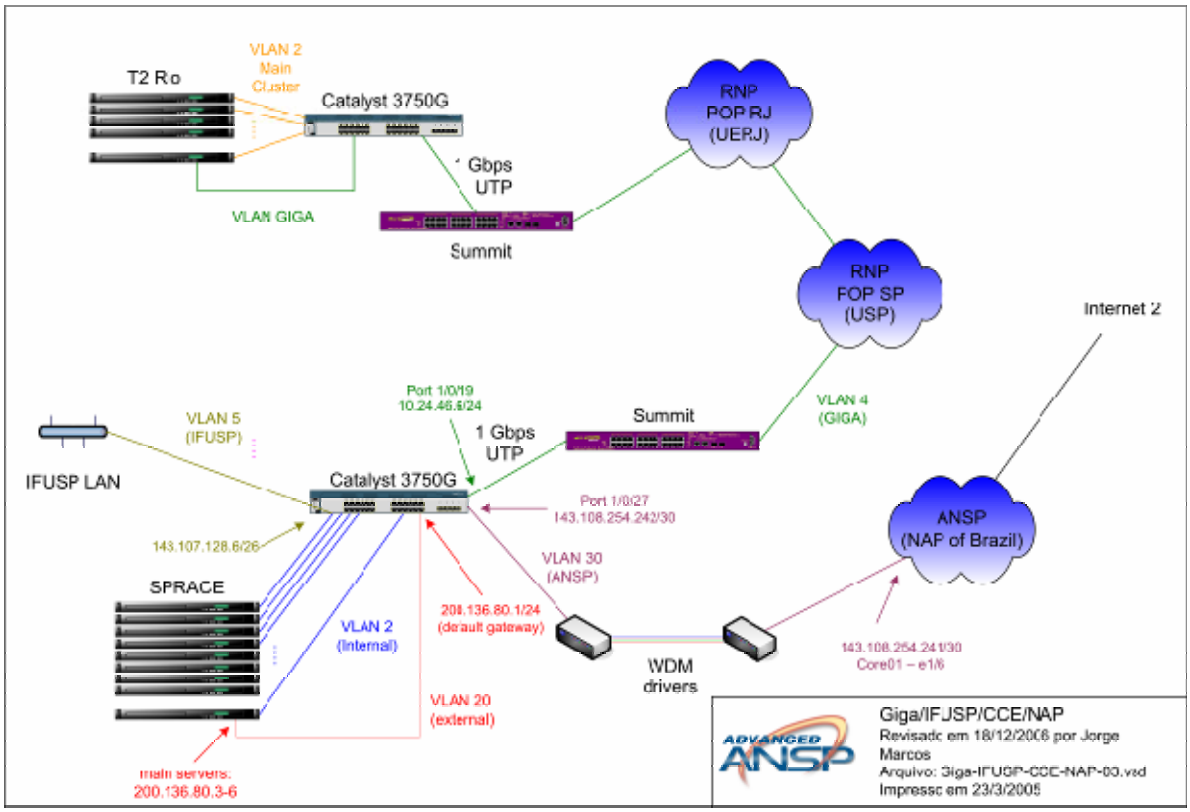

Figura 4: *Rede do laboratório, ligada à rede ANSP e à rede GIGA*

(b) Link interestadual: Projeto GIGA

O projeto GIGA, uma iniciativa da RNP e do CPqD, é uma rede experimental de alta velocidade. Consiste na implementação e uso de uma rede óptica voltada para o desenvolvimento de tecnologias de rede óptica, aplicações e serviços de telecomunicação associados à tecnologia IP sobre WDM (Wavelength Division Multiplexing), baseada em comutação de “lambdas”, com suporte a aplicações avançadas. Esta tecnologia associa sinais ópticos a diferentes freqüências de luz (comprimentos de onda ou “lambdas”), o que permite separar, dentro de um mesmo meio físico (a fibra óptica), canais diversos para tráfego de dados.

A rede experimental do Projeto GIGA foi implementada em maio de 2004, e tem atualmente 735 Km de extensão e capacidade de 1 Gbps, com previsão de se chegar a 10 Gbps. A rede interconecta 17 universidades e centros de pesquisa do eixo Rio-São Paulo, abrangendo os municípios de Campinas, São Paulo, São José dos Campos, Cachoeira Paulista, Rio de Janeiro, Niterói e Petrópolis. Há um projeto em andamento para estender a rede até o Nordeste. O uso da rede GIGA é exclusivo a subprojetos de pesquisa e desenvolvimento selecionados, com foco em uma das quatro áreas temáticas (definidas no lançamento do projeto): redes ópticas; serviços experimentais de telecomunicações; protocolos e serviços de rede; e serviços e aplicações científicas.

A infra-estrutura da rede GIGA é formada por equipamentos de núcleo baseados em roteadores BlackDiamond 10K e 6808, equipamentos de distribuição baseados em roteadores BlackDiamond 6808, da empresa Extreme Networks. A interligação entre regiões metropolitanas usa tecnologia DWDM (mais

cara, porém mais eficiente), enquanto a conectividade dentro das regiões metropolitanas é feita através de tecnologia CWDM. Os equipamentos de acesso, instalados nos laboratórios, são baseados em switches Summit 200-24, também da Extreme Networks. Por questões de segurança, existem dois centros de controle da rede (NOCs – Network Operations Centers): um na RNP no Rio de Janeiro e outro no CPqD em Campinas.

No Estado de São Paulo, o Projeto KyaTera da FAPESP apresenta características semelhantes ao projeto GIGA. Trata-se de uma plataforma óptica de alta velocidade, com equipamentos de última geração, destinada exclusivamente para ensino e pesquisa científica e tecnológica, sendo a mais avançada do hemisfério sul. As fibras do Projeto KyaTera são também instaladas diretamente nos diversos laboratórios de pesquisa participantes. Como ambos projetos surgiram na mesma época, as fibras dos cabos ópticos lançados em diversos campi são compartilhados. O laboratório SPRACE submeteu recentemente um subprojeto para participação no KyaTera, o qual foi aprovado.

Iniciativas semelhantes estão surgindo em outros países da América Latina. No Chile, a rede nacional de pesquisa REUNA está implementando um testbed óptico de 250 Km entre as cidades de Valparaíso e Santiago, conectando 3 Universidades, além da própria REUNA, também usando a tecnologia IP sobre WDM. Este projeto, que se iniciou em 2003, é suportado pela agência chilena de fomento à pesquisa FONDEF, e inclui o desenvolvimento de tecnologias nos níveis óptico, rede IP e aplicações [12].

Um dos subprojetos aprovados para uso da rede GIGA é o projeto de N^{o} 2446, “HepGrid Brazil”, do qual participam o DFNAE/UERJ, o LAFEX/CBPF e o LPC/IF/UFRJ no Rio de Janeiro, o SPRACE/UNESP e o IF/USP em São Paulo. O objetivo do subprojeto apresentado é justamente utilizar a infra-estrutura de rede de alta velocidade disponibilizada pelo GIGA para interligar esses laboratórios de pesquisa, viabilizando assim o compartilhamento de recursos computacionais necessários à construção de um Grid regional, para a execução conjunta de aplicações científicas em física experimental de altas energias.

Através da configuração adequada dos equipamentos de rede que compõem a rede GIGA, os engenheiros de rede estabeleceram uma rede local virtual ou VLAN (virtual local area network) interligando os laboratórios das instituições acima citadas. Os equipamentos de rede foram configurados de modo a rotear o tráfego através desta VLAN, provendo a interligação necessária entre as diversas redes de cada uma das instituições. O diagrama a seguir apresenta a configuração definida e disponibilizada pelos engenheiros da rede GIGA.

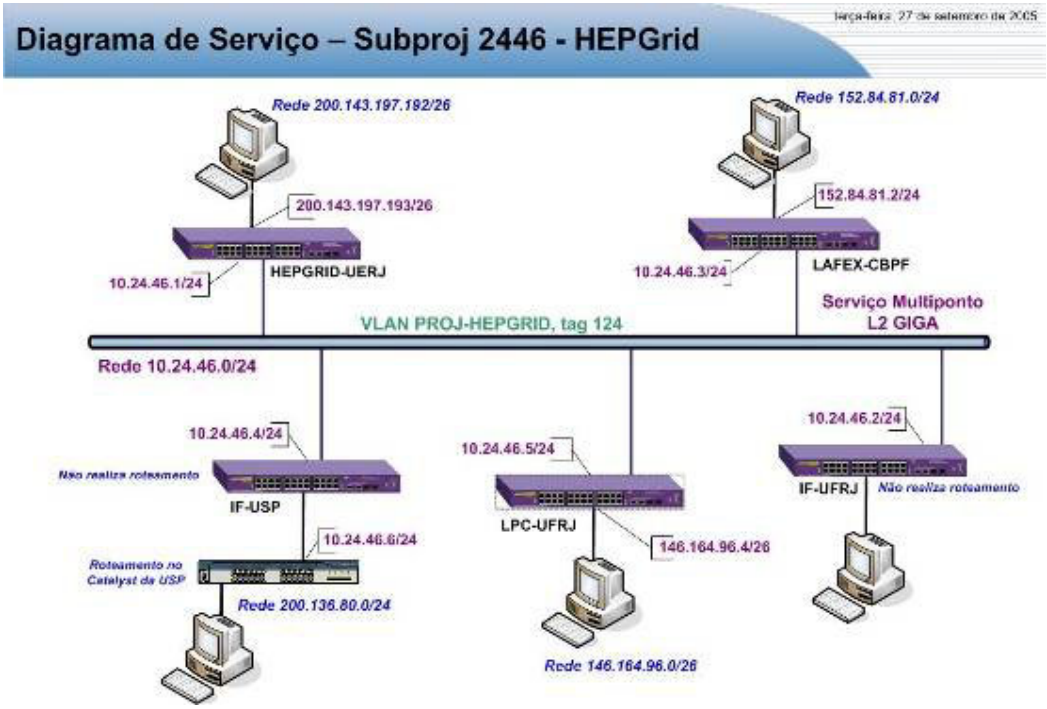

Figura 5: *Rede HEPGrid do Subprojeto 2446 do GIGA, que interliga CBPF, UERJ, UFRJ e SPRACE*

No SPRACE em particular, definimos uma VLAN específica para o projeto GIGA no switch Cisco Catalyst 3750 (doado pelo Caltech), e a mesma foi conectada à VLAN da rede GIGA, como pode ser verificado na figura anterior. Essa configuração foi feita para o switch Catalyst atuar como roteador, permitindo um controle local do roteamento, independente do projeto GIGA. Tal procedimento foi adotado por apresentar as seguintes vantagens:

- dá maior flexibilidade ao laboratório, tornando-o independente (não há necessidade de solicitar possíveis alterações de configuração aos engenheiros do projeto GIGA);
- permite a monitoração local do tráfego pela rede GIGA, usando as facilidades de monitoramento presentes no switch Catalyst, e de forma independente do projeto GIGA;
- mantém inalteradas as configurações de rede dos servidores do cluster (quaisquer alterações de configuração serão sempre efetuadas no ativo de rede, preservando os servidores).

Como se nota na Figura 5, no caso específico de São Paulo, o switch de acesso Summit 200-24, embora desnecessário devido à presença do Catalyst, foi mantido para garantir a independência de configuração (os engenheiros do GIGA configuram o Summit; os pesquisadores e analistas do SPRACE e engenheiros da ANSP configuram o Catalyst).

A configuração final da rede apresentada no diagrama anterior foi completada em setembro de 2005. A rede assim disponibilizada pelo projeto GIGA permitiu a formação de um Grid regional entre os principais clusters do projeto HEPGrid Brazil (em São Paulo e no Rio de Janeiro).

A participação do SPRACE nos Projetos GIGA e KyaTera foi fundamental para prover o laboratório de uma infra-estrutura de rede bastante avançada, capaz de atender às necessidades de conectividade óptica dos próximos anos. Graças a esses projetos, foi lançado um cabo óptico monomodo, do fabricante Lucent, de excelente qualidade (atenuação menor do que 0.2 dB / Km), com

24 fibras (12 pares), cobrindo a distância de 1400 metros entre o Centro de Computação Eletrônica (CCE) e o Instituto de Física, dentro do campus da USP. Este cabo foi doado pelo Projeto KyaTera, e o custo de lançamento do mesmo (de R$ 1653,59) foi custeado com a reserva técnica do temático do SPRACE. O trabalho foi contratado e administrado pelo CCE, e executado pela mesma empresa que lançou os cabos ópticos dos projetos GIGA e KyaTera por todo o campus da USP.

As fibras desse cabo óptico foram divididas em 3 grupos, e destinadas para os projetos GIGA e KyaTera, sendo algumas reservadas para aumentar a conectividade óptica entre o Instituto de Física e o CCE. A divisão ficou da seguinte forma: 2 pares para o projeto GIGA, 6 pares para o projeto KyaTera, e 4 pares para o Instituto de Física. Assim, como resultado secundário desse trabalho, o Instituto de Física quintuplicou sua capacidade de comunicação via rede óptica com o Centro de Computação Eletrônica.

Um dos pares do projeto GIGA foi usado para fechar a conexão entre o switch Summit 200-24 do SPRACE e o roteador BlackDiamond 6808 localizado no CCE. O segundo par é o responsável por fechar a conexão entre o switch Catalyst 3750 e o CCE. Um “patch-cord” instalado no “Wiring-Closet” do CCE provê a conexão entre o bastidor do projeto GIGA e o bastidor da empresa Iqara Telecom, que provê o serviço de conexão em fibras ópticas do CCE até o NAP do Brasil, em Barueri (onde estão instalados os equipamentos de rede que formam o “backbone” da rede ANSP).

Como o cabo óptico Lucent é um cabo “outdoor”, não podendo avançar para dentro do laboratório, foi também instalado um outro cabo óptico, do tipo “indoor”, com 16 fibras (8 pares), da Metrocable, doado também pelo projeto KyaTera, entre o Centro de Computação do Instituto de Física da USP (CCIFUSP) e o laboratório SPRACE, cobrindo uma distância de 50 metros. O mini-rack da GKC instalado no laboratório, que abriga os ativos de rede, também foi doado pelo projeto KyaTera. O projeto GIGA, por sua vez, disponibilizou dois painéis de distribuição óptica para 24 fibras. Um deles foi instalado no CCIFUSP e outro no mini-rack do laboratório. As emendas ópticas (fusões) e conectores ópticos foram também disponibilizados pelo projeto GIGA. No distribuidor do CCIFUSP, foram efetuadas fusões em 4 pares das fibras que chegam do CCE através do cabo óptico “outdoor”. Tais fibras foram disponibilizadas para uso do Instituto de Física. As fibras restantes do cabo óptico “outdoor” (8 pares) foram fundidas diretamente às fibras do cabo “indoor”, e levadas diretamente ao painel de distribuição óptica do laboratório, que foi instalado dentro do mini-rack.

2.3. Infra-estrutura de rede elétrica

O quadro de distribuição de energia elétrica do SPRACE é alimentado por 3 fases de 120V, cada uma protegida por um fusível de 160A, e distribui 2 fases + terra para um total de 20 tomadas, instaladas abaixo do piso elevado da sala de servidores. Dessas 20 tomadas, 13 delas alimentam os 'no-breaks': 12 'no-breaks' na sala de servidores e um 'no-break' externo (que alimenta o servidor sprace e uma estação de trabalho). Os disjuntores do quadro de distribuição são de 20 A x 380V, e protegem cada uma das 20 tomadas. A numeração dos disjuntores está marcada na fiação interna ao quadro, começando em 01 no canto superior esquerdo, e terminando em 20 no canto inferior direito, sendo que os ímpares estão à esquerda e os pares à direita, olhando o quadro de frente, conforme mostra o esquema apresentado na Figura 6 a seguir. Este levantamento foi efetuado em janeiro de 2006. Através dele, e verificando a distribuição das fases no quadro de força, conforme indicado na Figura 6, foi possível fazer uma melhor distribuição da carga total em cada uma das 3 fases, bastando para isso cuidar de ligar 4 ‘no-breaks’ em cada par de fases, que corresponde à configuração atual, em que as 3 fases do projeto já estão totalmente implantadas.

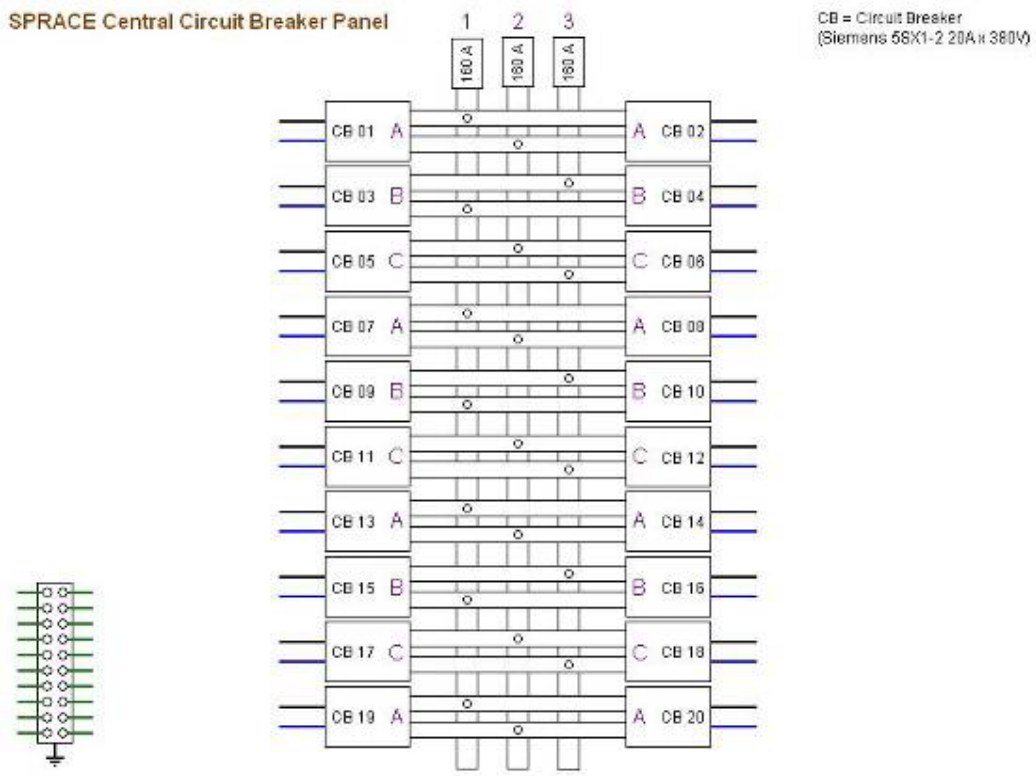

Figura 6: *Representação do quadro de força que alimenta os sistemas computacionais do SPRACE*

Os 'no-breaks' da APC têm uma série de itens configuráveis. Uma das configurações possíveis é chamada "Capacidade mínima antes de voltar a ligar", que faz com que o 'no-break' carregue suas baterias até uma percentagem especificada antes de ser religado (default é 0 - religa imediatamente, e pode ser configurado para 15, 50 ou 90% de carga nas baterias). A capacidade de corrente nominal de cada 'no-break' em regime é de 16A; logo os disjuntores estão com a especificação correta (20A). Os 'no-breaks' foram configurados de modo que eles façam uma carga de 15% nas baterias antes de religar os servidores, uma tentativa de evitar o surto de corrente que desarma os disjuntores quando a energia é restabelecida após uma queda de energia prolongada. O que provoca o desarme é o surto de corrente quando volta a energia após uma queda prolongada, pois cada 'no-break' religa 8 servidores ao mesmo tempo, quando as baterias ainda estão descarregadas.Infelizmente, porém, essa configuração não resolveu o problema de desarme dos disjuntores, um problema que ainda permanece sem solução. O desarme dos disjuntores tem impedido a possibilidade de se religar os sistemas remotamente após uma queda de energia (num final de semana, por exemplo).

Existe a possibilidade de se controlar remotamente os servidores (ligar/desligar, dar reset, monitorar as temperaturas internas e até velocidades das ventoinhas), de forma independente do sistema operacional. A Supermicro tem um módulo que se instala dentro do servidor, que permite realizar essas tarefas remotamente, através da rede, e de forma independente do sistema operacional, de acordo c/ o padrão IPMI da Intel (http://www.intel.com/design/servers/ipmi/).

Todos os servidores do cluster do SPRACE usam ‘mainboards’ da Supermicro, então seria possível instalar esses módulos em todos eles. Os servidores da fase III foram previstos com estes módulos já instalados.

3. Atividades realizadas no período

Exercendo a função de Especialista de Laboratório do SPRACE, as atividades profissionais que executo estão relacionadas ao suporte aos sistemas computacionais e à infra-estrutura de rede do centro de processamento de dados o SPRACE, e a de auxílio às atividades de pesquisa do grupo. As principais atividades que executei durante o período a que se refere este relatório são detalhadas a seguir.

3.1. Ampliação do SPRACE

(a) Implantação da Fase III

No decorrer de 2006 foi realizada a implantação da terceira fase do centro de processamento de dados do SPRACE, conforme proposto no projeto temático submetido à FAPESP. A descrição das características dos equipamentos adquiridos é dada a seguir.

32 Nós de Processamento:

- Plataforma LX 211 XEON EM64T
- Processador 2 x Xeon Dual Core 2,0 GHz FSB 1333 MHz
- Memória: 4 GB FBDIMM DDR2 (4 pentes de 1GB)
- Disco: 1 x SATA 160 GB
- Rede: 2 interfaces Intel 10/100/1000 (1Gbps)
- CDROM 24X / Floppy
- Gabinete tipo rack com 1U de altura

2 Racks:

- Largura padrão (19”) com 23 U's de altura
- Réguas de tomadas na parte traseira
- Ventilação forçada na parte superior

Para a aquisição dos equipamentos da fase III, foram realizadas tomadas de preço com 4 fornecedores nacionais, que foram convidados a apresentar proposta de venda, conforme especificações técnicas detalhadas, enviadas através de carta-convite. Os resultados das tomadas de preço são apresentados na tabela 1 a seguir. A empresa Itautec foi a que apresentou proposta de menor preço e melhor técnica.

Itautec				
Item	**Descrição**	**Qtd.**	**Valor**	**Total**
1	Nó de Processamento	32	11.990,00	383.680,00
2	Racks	2	6.430,00	12.860,00
	Total			**R$ 396.540,00**

IBM				
Item	**Descrição**	**Qtd.**	**Valor**	**Total**
1	Nó de Processamento	32	21.014,67	672.469,69
2	Racks	2	4.343,6	8.687,20
	Total			**R$ 681.156,89**

HP				
Item	**Descrição**	**Qtd.**	**Valor**	**Total**
1	Nó de Processamento	32	12.855,00	411.360,00
2	Racks	2	9.320,00	18.640,00
	Total			**R$ 430.000,00**

Dell				
Item	**Descrição**	**Qtd.**	**Valor**	**Total**
1	Nó de Processamento	32	19.482,42	623.437,44
2	Racks	2	16.506,60	33.013,20
	Total			**R$ 656.450,64**

Tabela 1: *Preço obtido junto aos fornecedores para compra dos equipamentos da terceira fase do projeto*

Graças a uma negociação cuidadosa, a proposta da Itautec incluiu também uma extensão da garantia "on site" de três anos para todos os servidores Itautec do SPRACE, incluindo aqueles adquiridos em fevereiro de 2004 (1 servidor Infoserver 1251 + 16 nós de processamento Infoserver 1252) e em junho de 2005 (32 nós de processamento LX210).

Após a aquisição e instalação dos servidores da fase III, o SPRACE ficou com a seguinte configuração:

- Capacidade de processamento: 244 processadores Xeon (‘cores’), perfazendo um total de 720 GHz de poder de processamento. Todos os processadores possuem 1 GB de memória RAM dedicada.
- Capacidade de Armazenamento: Dos 86 nós de processamento, 54 possuem discos de 36 GB SCSI, e 32 deles discos de 160 GB SATA. Cada um dos 2 servidores de armazenamento possui dois discos de 36 GB e o servidor de gerenciamento possui 4 discos de 72 GB. Dois RAID’s possuem 2.016 GB cada e outros 2 RAID’s tem capacidade de 4.200 GB cada, perfazendo um total de 14.8 TB de capacidade de armazenamento.

O trabalho realizado para a implantação da terceira fase do SPRACE foi cercado de cuidados. Foram tomadas decisões no sentido de se conseguir adquirir a tecnologia mais recente, dentro do orçamento disponível, o qual foi elaborado em 2005. Devido à variedade e complexidade das novas tecnologias disponíveis, com destaque para o surgimento no mercado dos primeiros processadores de núcleo duplo (‘dual-core’), percebemos que seria essencial fazer uma escolha criteriosa dos novos servidores, de forma a maximizar desempenho e flexibilidade e ao mesmo tempo minimizar os custos de aquisição e de operação, incluindo consumo de energia e necessidade de refrigeração. A escolha da tecnologia foi uma decisão estratégica: custo versus desempenho e eficiência no consumo de energia e dissipação de calor. Tal escolha levou em conta a necessidade de otimização do ‘data-center’ do SPRACE, aumentando a densidade de poder computacional por volume ocupado pelos servidores e, ao mesmo tempo, evitando o consumo excessivo de energia elétrica e novos investimentos na capacidade de refrigeração. De forma a aproveitar ao máximo o orçamento disponível para a fase III, decidimos dividir a aquisição em 3 partes: orçamentos de servidores e racks foram enviados para fabricante de computadores, de switches para fabricantes de equipamentos de rede, e de nobreaks para revendas especializadas nesse tipo de produto. Esta estratégia permitiu negociar o melhor preço diretamente com as empresas especializadas em cada tipo de equipamento, evitando o repasse de lucro entre fornecedores. A estratégia foi bem sucedida, pois foi possível adquirir equipamentos topo de linha dentro do orçamento disponível. Tal ocorreu por exemplo com os switches de rede. Através de negociação com apoio da 3Com, foi possível obter um preço diferenciado junto ao distribuidor CNT Brasil, que em geral não efetua venda a consumidor final. Porém, a pedido da 3Com, e com a intermediação da revenda Ziva Tecnologia e Soluções Ltda., foi possível fechar a aquisição de um par de switches 3Com modelo 3824 (3C17400), um excelente equipamento, a um preço cerca de 40% inferior ao praticado no varejo.

O levantamento técnico para a aquisição dos servidores iniciou em abril, através de contatos com representantes da Intel Brasil. Em maio, tomamos a decisão de escolher os novos processadores Intel série 5100, de codinome ‘Woodcrest’, para a implementação da Fase III. Resultados preliminares publicados em abril indicavam que essa nova plataforma apresentava inovações tecnológicas que resultavam numa maior eficiência no consumo de potência e num maior desempenho do poder de processamento, quando comparada aos mais rápidos processadores de núcleo único (‘single-core’) da geração anterior, e a um custo muito próximo destes.

O processador Woodcrest consiste de dois núcleos de processamento, encapsulados em um chip único, compartilhando um mesmo cache nível 2 de 4 MB. O espaço de memória no cache compartilhado é alocado dinamicamente de acordo com as necessidades de cada processador, e também serve como meio de troca de dados entre eles. A nova plataforma que dá suporte aos processadres Woodcrest (denominada plataforma Bensley) fornece uma infra-estrutura muito melhorada para servidores ‘low’ e ‘mid-range’, a faixa de servidores que atende aos requisitos de processamento do SPRACE, com uma largura de banda muito maior para cada processador e para a memória, em relação a plataformas anteriores da própria Intel. Esta plataforma também adotou a nova tecnologia de memórias do tipo ‘Fully Buffered DIMMs’, que consistem de um arranjo de memórias DDR2 padrão unidas por um chip que funciona como um buffer, o qual se comunica com o controlador de memória através de um link

serial. Os novos servidores passam a dispor de 4 canais FB-DIMM que permitem acessos concorrentes aos módulos de memória, permitindo uma largura de banda de acesso à memória superior a 20 GB/s.

Comparando-se servidores com 4 núcleos de processamento ('dual-processor / dual-core') com servidores com 2 núcleos (dual-processor / single-core'), poder-se-ia esperar um desempenho 2 vezes superior. No entanto, resultados de testes publicados em abril deste ano [13] já mostravam que o desempenho de servidores com 2 processadores Woodcrest chega a ser 3 vezes superior para operações aritméticas e lógicas com números inteiros (SPECint2000 Standard Benchmark), e quase 4 vezes superior para operações em ponto flutuante (SPECfp2000 Standard Benchmark), em relação a servidores com 2 processadores da geração anterior, devido principalmente à melhoria na tecnologia de cada núcleo de processamento, ao aumento na velocidade do 'front-side-bus' e a melhorias no chipset.

Em maio deste ano apostamos numa arquitetura que ainda não havia sido lançada (o processador Woodcrest foi lançado oficialmente em 26 de junho). Os resultados dos diversos 'benchmarks' publicados recentemente comprovam que a escolha foi acertada. Conforme fomos informados pela Intel Brasil, o SPRACE foi o primeiro projeto envolvendo os novos processadores dual-core série 5100 da América Latina, e esteve entre os primeiros do mundo.

(b) Ampliação da capacidade de memória dos servidores da fase I

Devido à grande demanda por memória dos aplicativos do CMS, vimo-nos obrigados a aumentar a memória dos nós de trabalho adquiridos na fase I, passando de 512MB por processador, para 1 GB por processador, configuração na qual já haviam sido adquiridos os nós de processamento da fase II. O nós de trabalhos adquiridos na fase III possuem também 1 GB de memória RAM por núcleo de processamento.

Da mesma forma, a enorme quantidade de transmissão de dados simultâneos que ocorrem durante a operação do processamento dos dados do CMS, nos obrigou a aumentar a memória do servidor de disco, passando também de 512MB para 1 GB por processador.

Após este 'upgrade', todos os componentes do cluster passaram a possuir 1GB de memória por núcleo de processamento, ajustando-se às especificações requeridas para o processamento dos dados do experimento CMS.

Foram adquiridos 24 módulos de memória DDR PC-2100 de 1 GB, do tipo 'ECC registered', marca Itaucom, modelo 01GE266R54. Foram feitas 3 tomadas de preço, junto a fornecedores renomados.

Através de negociação com apoio da Itautec, foi possível obter preço diferenciado junto ao distribuidor Alcatéia, que em geral não realiza venda a consumidor final. Por intermediação da Itautec, foi possível fechar a compra com este distribuidor a um preço cerca de 30% mais em conta do que o praticado no varejo.

(c) Aquisição de equipamento de videoconferência

No decorrer de 2006, percebeu-se que a instalação de um ambiente de comunicação colaborativa virtual através de recursos de videoconferência passou a ser absolutamente necessária,

para atender à demanda crescente por comunicação de melhor qualidade com grupos do exterior. A participação do grupo em seminários e workshops passou a ocorrer com maior freqüência, e com um número cada vez maior de participantes. Um equipamento de videoconferência de última geração foi então adquirido para uso do grupo, após consulta técnica realizada com representantes da Polycom, uma das empresas líderes no segmento de áudio e videoconferência. Foi escolhido o modelo VSX-7000s, para manter compatibilidade com os sistemas de videoconferência usados pelos diversos grupos da colaboração. Tal equipamento permitirá ao grupo estabelecer comunicação multiponto, com áudio e vídeo de alta qualidade, com grupos americanos através de conexão à MCU da ESNet, com sede na Califórnia, e com grupos europeus através de conexão à MCU do IN2P3, com sede em Lyon. O equipamento adquirido tem interface para conexão em redes IP e possibilidade de expansão futura para conexão com rede telefônica ISDN. É capaz de interoperar com outros sistemas de videoconferência baseados em hardware ou software (por exemplo, o VRVS), incluindo suporte a QoS e multicast, seguindo a especificação H.323, que estabelece padrões para a comunicação multimídia em redes que não dispõem de qualidade de serviço, como a Internet. Equipamentos que implementam a especificação H.323 suportam vídeo, áudio e troca de dados em tempo real.

3.2. Manutenção do SPRACE

As atividades do dia-a-dia consistem basicamente na manutenção da operação dos sistemas de forma ininterrupta, através da monitoração contínua dos servidores (hardware) e dos serviços (software). O processo de monitoração consiste nas atividades rotineiras executadas por um analista de sistemas da área de suporte, tais como verificar, continuamente, se todos os servidores encontram-se em operação, se os serviços principais que são executados nos mesmos estão operantes, se não há ocupação excessiva de espaço nos discos, se nenhum servidor está operando além da temperatura normal, além das atividades relativas à segurança e integridade dos sistemas, como o acompanhamento dos logs de sistema de cada servidor para verificar se há algum processo executando de forma irregular ou se houve alguma tentativa de acesso indevido aos sistemas. Em caso de falha em algum dos sistemas, constitui atividade correlata o diagnóstico preliminar, de forma a se buscar a origem da falha, e em caso de defeito de hardware o contato com os fornecedores para comunicar o problema e solicitar o reparo adequado. Constitui também atividade rotineira a conservação da limpeza e organização da sala dos servidores, bem como a organização dos documentos de manutenção dos fornecedores e os registros das atividades realizadas (logbook). Essas são atividades permanentes e rotineiras, e foram realizadas diariamente ao longo de todo o ano.

3.3. Participação na Colaboração CMS

O LHC deverá enfrentar um grande desafio para processar os dados produzidos em seus detectores. Deverá ocorrer cerca de um bilhão de interações próton-próton por segundo, gerando uma média de 30 a 40 novas partículas. Esta taxa de produção jamais foi atingida em outro detector e deverá exigir que cada experimento tenha que armazenar dados, depois do trigger de nível 3, à razão de 100 MB/seg. O processamento destes dados só será possível com a implementação da arquitetura de Grid. Neste contexto, todos os laboratórios que participarem de um experimento do LHC poderão contribuir para o processamento, colocando à disposição seus recursos computacionais. Passa, portanto, a ser imprescindível participar desta iniciativa para que se possa fazer pesquisa científica nesta área.

Como parte dos preparativos para participar do processamento para a Colaboração CMS, o SPRACE passou a fazer parte do *Open Science Grid* (OSG)[1]. O OSG é a iniciativa americana de processamento distribuído que apóia a computação científica através da colaboração entre pesquisadores, desenvolvedores de software e engenheiros de rede. O OSG vem operando uma rede internacional de recursos computacionais que permite acesso aos pesquisadores de diversas áreas a esses recursos compartilhados.

Como qualquer outra infra-estrutura de Grid, o OSG pode ser definido como um sistema que coordena recursos sem a utilização de um sistema central de gerenciamento, no qual o sistema combinado é significativamente maior que suas partes. Para tornar-se universal e possibilitar seu desenvolvimento também de forma distribuída e descentralizada, o OSG é baseado em código-fonte aberto e utiliza protocolos padronizados. O SPRACE participa do Open Science Grid contribuindo com os seguintes elementos:

(i) Elemento de Computação:

O Elemento de Computação do Open Science Grid no SPRACE está localizado no servidor spgrid.if.usp.br. No OSG este elemento de computação está registrado como sendo a localidade SPRACE do parque UNESP. Ele funciona como o 'gateway' dos nós de trabalho do cluster, e também como servidor de 'login' dos usuários locais. Ele serve ao cluster as seguintes funcionalidades:

Middleware do OSG: Está atualmente instalada no nosso elemento de computação a versão 0.4.1 do conjunto de serviços do OSG, o mais recente. Este conjunto de serviços contém os seguintes pacotes:

- Globus Toolkit – Middleware básico do sistema de manuseio de trabalhos do Grid
- MonALISA – Ferramenta de monitoramente global.
- GUMS (Grid User Membership Service) – Verifica a autenticidade dos requerentes dos serviços de grid consultando seus registros nas diversas Organizações Virtuais atendidas pelo SPRACE e os mapeia em contas locais, determinando suas autorização e política de uso dos recursos do SPRACE.
- GIP (Generic Information Provider) – Provê aos serviços de informação do OSG as informacoes relativas à existência e disponibilidade dos recursos de processamento do SPRACE.
- BDII (Berkeley Database Information Index) – Provê ao EGEE, Grid utilizado pelos membros do LHC Computing Grid localizados na Europa, as mesmas funcionalides do GIP para o OSG. Necessário para compatibilização dos recursos computacionais mundialmente distribuídos do experimento CMS.

Condor: Sistema de gerenciamento, ordenamento e distribuição de trabalho aos nós de processamento. Recebe os trabalhos do middleware do OSG e os executa nos recursos existentes no cluster.

Ganglia: Sistema de monitoramento do funcionamento do cluster. Provê informação em html do estado dos componentes do cluster, tais como carga de trabalho, ocupação de memória, tráfego de dados etc. Atualiza as informações a cada minuto e mantém um histórico de até um ano.

CMSSW: Conjunto de pacotes do experimento CMS, utilizado para a realização das simulações de Monte Carlo e processamento dos trabalhos de análise de dados do experimento.

NFS: Serviço de sistema de arquivos de rede que exporta o software do OSG, o Condor e o CMSSW aos nós de trabalho, bem como o 'home' dos usuários locais aos diversos elementos do cluster que o necessite.

---

1 http://www.opensciencegrid.org/

(ii) Nós de Trabalho

O SPRACE já está disponibilizando atualmente todos os nós de trabalho, adquiridos nas Fases I, II e III do projeto, ao elemento de computação do OSG. Os serviços vindos do Grid são executados em partições do disco local dos nós de trabalhado e apagados quando do término da execução do serviço. Os nós de trabalho estão localizados na rede local do cluster e se comunicam com a rede externa (link internacional WHREN-LILA) através do firewall localizado no gateway do cluster (o servidor spgrid).

(iii) Elemento de Armazenamento

O Elemento de Armazenamento do OSG no SPRACE está localizado na máquina spdc00.if.usp.br e registrada como sendo a localidade SPRACE:sem_v1 do parque UNESP. Nele foram instalados as seguintes funcionalidades e serviços.

pNFS: Sigla de *Perfectly Normal File System*, o pNFS é um sistema de arquivo em que recursos de armazenamento fisicamente distribuídos, como discos em máquinas distintas, aparecem como sendo parte de um único sistema de arquivos contínuo. Ele é constituído basicamente por um banco de dados *postgresql* com interface de acesso similar aos comandos *posix* usuais de manipulação de arquivos. Nele está também incorporado o servidor NFS de modo que as informações sobre seus arquivos possam ser exportadas aos demais elementos do cluster.

dCache: Sistema de catalogo de arquivos. Utiliza o pNFS para acesso aos recursos de armazenamento em disco, e pode ser utilizado também para armazenamento em fita.

SRM: O Storage Resource Manager faz parte do middleware do OSG para prover acesso uniforme aos recursos computacionais do Grid. Sua função básica é traduzir o *Logical File Name* com o qual os arquivos universalmente distinguidos no Grid em *Physical File Name*, que é como os arquivos são identificados e localizados pelos sistemas locais. É também o serviço do Grid encarregado de prestar informações e realizar operações com os arquivos dos diversos Elementos de Armazenamento do Grid. Faz a interface entre o Grid e o dCache local ao cluster.

PhEDex: Este é o sistema utilizado pela colaboração CMS para a movimentação de dados entre os diversos elementos de processamento e armazenamento utilizados pelo experimento. Este serviço é encarregado de manter o catálogo com a localização das réplicas dos dados globalmente distribuídos e realizar a sua movimentação para o processamento dos dados que os requisitar. Utiliza o SRM para realizar suas operações.

Squid: É o sistema do CMS de acesso e distribuição das informações do banco de dados da calibração do detector e do sistema de aquisição de dados, fundamental para a realização da análise dos dados adquiridos.

(iv) Nós de Armazenamento do dCache

Assim como o Elemento de Computação possui os nós de trabalho que lhe disponibilizam poder de processamento, da mesma forma o Elemento de Armazenamento possui seus nós de armazenamento que lhe disponibiliza capacidade de armazenamento. No Elemento de Armazenamento do SPRACE cada nó de armazenamento executa seus próprios agentes de transferência de dados e possui conectividade com a rede mundial de modo a aumentar sua capacidade integrada de transferência de dados. Nele estão instalados os ‘pools’ do dCache para o compartilhamento do armazenamento, e os clientes do SRM para a execução da transferência dos dados para aquele determinado ‘pool’. Os nós de armazenamento do SPRACE são:

Servidor de Disco: Este é o servidor de armazenamento massivo do SPRACE ao qual estão conectados os módulos RAID, com uma capacidade total de 12 TB. Além de fornecer 4 *pools* ao dCache de 1,5 TB cada, este servidor exporta via NFS aos demais elementos do cluster a área onde estão localizados os aplicativos específicos de cada Organização Virtual.

Nós de Processamento: Devido ao custo decrescente e aumento da capacidade dos discos locais dos servidores de processamento, estes podem ter ao mesmo tempo a função de processar dados e disponibilizar seus discos para se incorporarem ao sistema de arquivos distribuídos do cluster. Em cada um dos 32 nós de processamento adquiridos na fase III deverá ser instalado um disco SATA de 500GB perfazendo um total de 16TB de espaço distribuído entre os nós. Com a sua incorporação ao Elemento de Armazenamento do SPRACE nos moldes descritos, o SPRACE contará com cerca de 28 TB de disco e será capaz de realizar 33 sessões de transmissão de dados em paralelo, com capacidade de até 1Gbps cada.

Em 2006, o SPRACE tornou-se um Tier2 plenamente funcional do experimento CMS. O T2_SPRACE vem se juntar aos Tier2 de Caltech, MIT, Winsconsin, Florida, Purdue e Nebraska, formando o sistema de processamento das Américas do CMS, sob a centralização do Tier1 do Fermilab. O SPRACE foi o primeiro Tier2 do hemisfério Sul a atuar plenamente sob a arquitetura de Grid do Worldwide LHC Computing Grid.

3.4. Participação em Treinamentos e Workshops

(a) I Brazilian LHC Workshop

O “I Brazilian LHC Computing Workshop” [14] foi organizado pelo próprio grupo de pesquisa, sob a coordenação do Prof. Sérgio Novaes, e ocorreu no Instituto de Física da USP no dia 8 de maio de 2006. A finalidade deste workshop foi discutir as necessidades computacionais de cada um dos experimentos do CERN e estabelecer a demanda de recursos computacionais dos grupos brasileiros que fazem parte destes experimentos para estarem preparados para o ínicio das operações do LHC em 2007. Um objetivo secundário foi também traçar um plano estratégico para garantir que os diferentes middlewares e softwares possam rodar em harmonia, compartilhando os recursos disponíveis. A principal proposta debatida foi a possibilidade de integração dos recursos nacionais dos pesquisadores envolvidos nos experimentos do CERN em uma única Tier distribuída do WLCG.

O workshop contou com 20 participantes do Brasil e do exterior e teve a seguinte programação:

09:45 – 10:00	Welcome	
10:00 – 10:30	Computação no Alice e Grid	Alexandre Suaide
10:30 – 11:00	Atlas Computing	Fernando Marroquim
11:00 – 11:30	Coffee Break	
11:30 – 12:00	Grid Computing in CMS	Andre Sznajder
12:00 – 12:30	LHCb Computing Model	Miriam Gandelman
12:30 – 14:00	Lunch	
14:00 – 14:30	The EELA Project	Diego Carvalho
14:30 – 15:00	OSG-EGEE Interoperability for WLCG Jobs	Ruth Pordes
15:00 – 15:30	Coffee Break	
15:30 – 16:00	Atlas Distributed SW Tier 2 Center	Horst Severini
16:00 – 18:00	Discussion: Perspectives & Outlook	Sérgio Novaes (moderator)
20:00	Dinner	

(b) 3rd EELA Tutorial

De 26 a 30 de junho deste ano, participei do "3rd EELA Tutorial for Users and System Administrators" [15], evento relacionado ao Projeto EELA [16] que ocorreu nas dependências do Instituto de Física da Universidade Federal do Rio de Janeiro. Neste evento, tivemos a oportunidade de conhecer mais detalhadamente toda a infra-estrutura de software necessária para um site operar como um nó do middleware de Grid mais difundido na Europa, e base do middleware atualmente usado pelo LHC Computing Grid, o EGEE [17]. As sessões teóricas foram todas sucedidas por atividades práticas em laboratório, durante as quais foi possível instalar todos os pacotes necessários para a operação do EGEE. O projeto EELA, coordenado pelo CIEMAT na Espanha e pela UFRJ no Brasil, é um consórcio formado por diversas instituições de pesquisa européias e latino-americanas, cujo objetivo principal é estabelecer uma infra-estrutura de Grid computacional piloto na América Latina, interoperável com o projeto EGEE já existente e consolidado na Comunidade Européia. O middleware EGEE é extensivamente usado, de modo que um objetivo secundário deste projeto é promover a difusão dessa infra-estrutura para além da Europa. A Figura 7 a seguir detalha os principais participantes do EELA no estágio atual do projeto.

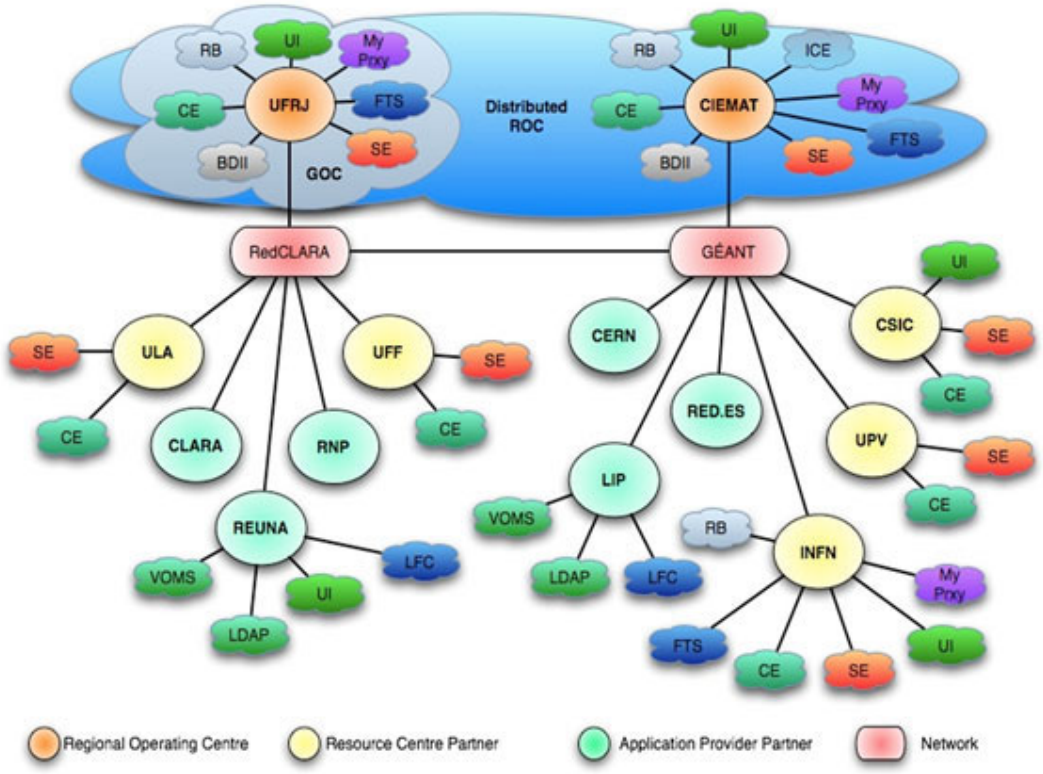

Figura 7: *Infra-estrutura atual do projeto EELA, ressaltando as instituições européias (CIEMAT, CSIC e UPV na Espanha, INFN na Itália, CERN na fronteira Franco-Suíça, e LIP em Portugal, além das redes de educação e pesquisa RED.ES e GÉANT2) e latino-americanas (UFRJ e UFF no Brasil, ULA na Venezuela, UNAM no México, UTFSM e UDEC no Chile, CUBAENERGIA em Cuba, UNLP na Argentina, além das redes nacionais de educação e pesquisa REUNA, RNP, e do consórcio redCLARA).*

(c) CERN School of Computing 2006

De 21 de agosto a 01 de setembro, participei do CERN School of Computing [18], realizado em Helsinki na Finlândia. Trata-se de um treinamento intensivo nas áreas de Tecnologia de Computação em Grid, Tecnologias de Software e Física Computacional aplicada à Aquisição e Análise de dados experimentais. A seguir são detalhadas cada uma dessas atividades, que foram desenvolvidas ao longo das duas semanas de treinamento, com 8 horas de atividades diárias.

(i) Tecnologias de Computação em Grid

Foram discutidos vários aspectos da computação em Grid, além de experiência prática no uso e configuração de modernas ferramentas usadas nesta área. Uma boa parte do treinamento foi dedicada ao 'middleware' de Grid em uso pelos projetos LCG (LHC Computing Grid) e EGEE (Enabling Grids for e-Science). Esses projetos têm como objetivo a construção de uma infra-estrutura computacional distribuída de larga escala para apoio à pesquisa científica, que atualmente já engloba mais de 100 centros de pesquisa espalhados pela Europa, Ásia e Américas. Essa estrutura está sendo usada não apenas pelos experimentos do LHC, mas também por diversas outras comunidades científicas como biomedicina, astrofísica, química e ciências atmosféricas. Os tutoriais que compõem o treinamento

apresentam a arquitetura de software na qual se baseiam as infra-estruturas do LCG e do EGEE. Foram também discutidos outros 'midllewares' de alto nível, bem como softwares de aplicação que se apoiam nesta infra-estrutura. Os tutoriais foram complementados por dois seminários, que versaram sobre aspectos operacionais das infra-estruturas de Grid e respectivas técnicas de otimização. Novas tecnologias de Grid baseadas em extensões dos conceitos de Web Services (chamados Grid Services) foram também discutidas. O treinamento teve um enfoque eminentemente prático, através de vários exercícios centrados no uso de aplicações computacionais intensivas.

(ii) Tecnologias de Software

Foram apresentadas modernas técnicas de projeto de software, bem como ferramentas e tecnologias para a compreensão e aperfeiçoamento de softwares já desenvolvidos. Foi dada ênfase em projetos de software de larga escala, comuns em Física de Altas Energias. A primeira série de seminários cobriu um conjunto de ferramentas e técnicas que foram exemplificadas durante as aulas de exercícios. Os seminários incluiram tópicos como engenharia de software, projeto, metodologia e testes. Uma outra série de seminários foi centrada na tecnologia de Web Services; nos quais foram discutidas técnicas básicas que dão suporte a serviços de mais alto nível, como aqueles oferecidos pelas tecnologias de Grid. O treinamento foi complementado por dois tópicos essenciais: métodos e técnicas de segurança de computadores, e qualidade de serviços na Internet e melhoria de desempenho em redes de computadores.

(iii) Física Computacional - Análise e Aquisição de dados experimentais

Nesta parte do treinamento foram apresentados alguns conceitos fundamentais de Física Computacional, com ênfase nos apectos relacionados à simulação e visualização, incluindo simulações dos ajustes necessários para otimizar detetores, testes e aperfeiçoamentos dos softwares de reconstrução de dados, além de uma compreensão mais detalhada dos dados obtidos.

A primeira série de seminários apresentou os componentes de software e hardware necessários para o processamento de dados experimentais, desde a fonte dos dados - o detector - até a análise física. Foi dada ênfase nos conceitos, mas alguns detalhes de implementação foram discutidos. O conceito chave discutido foi a questão da redução de dados, tanto em termos de taxa de aquisição quanto em termos de densidade de informação. Foram apresentados vários algoritmos usados para redução de dados, tanto 'online' quanto 'offline'. Uma segunda série de seminários apresentou uma introdução ao domínio da simulação de experimentos em física de altas energias, com exemplos práticos obtidos do experimento CMS do LHC. As apresentações mostraram como executar a simulação de eventos através da ferramenta GEANT4, que é um 'toolkit' que simula a passagem de partículas através da matéria. Ele inclui um conjunto completo de funcionalidades, incluindo o 'tracking', a geometria, os modelos físicos,e os 'hits'. Os processos físicos oferecidos por esse 'toolkit' abrangem interações eletromagnéticas, hadrônicas e ópticas, um grande conjunto de partículas de longa vida, materiais e elementos, sobre uma grande faixa de energias, desde 250 eV até da ordem de TeV. Esse 'toolkit' foi projetado e construído para expor os modelos físicos utilizados, para manusear geometrias complexas, e para permitir uma fácil adaptação para uso ótimo em diferentes conjuntos de aplicações. Requisitos básicos de simulações foram explicados, bem como o setup experimental em termos da geometria, materiais e campos eletromagnéticos externos, princípios de processos físicos, seleção e configuração de processos físicos, conceitos de eventos, e extração de informações de colisões. Foram mostrados na prática como estes requisitos são obtidos usando a ferramenta GEANT4. Uma terceira série de seminários teve como ponto central as técnicas e sistemas de aquisição de dados online, com foco nos 4 experimentos do LHC, incluindo princípios de fluxo de dados, requisitos qualitativos e quantitativos e arquitetura dos sistemas de aquisição de dados.

(d) EELA Grid School

De 04 a 15 de dezembro, participei do evento 1st EELA Grid School [19], um programa de treinamento intensivo na área de Computação em Grid, ainda inédito no Brasil nos moldes planejados pelo comitê organizador, formado por renomados pesquisadores do CIEMAT (Espanha), INFN (Itália) e UFRJ (Brasil). Para nos tornar elegíveis a participar deste treinamento, formamos uma equipe composta de especialistas da USP, UNESP e UNICAMP, cujo objetivo foi adaptar uma aplicação científica para uso em Grids computacionais. A aplicação em questão é um simulador de múltiplas falhas em redes puramente ópticas, escrito em Java e desenvolvido pelo Dr. Gustavo S. Pavani da UNICAMP. O evento contou com a participação.de 8 equipes formadas por alunos vindos de diversos países da Europa e América Latina, que portaram aplicações para execução no 'middleware' do EEELA/EGEE. Nosso time, do qual participaram dois analistas do Centro de Computação Eletrônica da USP, além de mim e do Dr. Gustavo Pavani da Unicamp, conseguiu adaptar inteiramente a aplicação submetida previamente para análise, para execução no middleware do EELA/EGEE.

Além de guardar estreita vinculação com as atividades profissionais que atualmente exerço como funcionário do Instituto de Física, os conhecimentos que adquiri neste treinamento serão de grande valia para enriquecer as discussões em andamento no Grupo Assessor de Grid e Computação de Alto Desempenho, do qual sou participante ativo. Este grupo foi criado pela Coordenadoria de Tecnologia de Informação para discutir a implantação de uma infra-estrutura de Grid Computacional na Universidade de São Paulo.

3.5. Seminários e palestras técnicas apresentados

"Algumas Iniciativas em 'e-Science': Grids e Redes Ópticas Avançadas", I Seminário sobre Inovações Tecnológicas com a aplicação de Cluster e Grid no âmbito do Governo Federal. Brasília, em 16 de dezembro de 2005.

"Scheduling Lightpaths: The Network as a Managed Resource in Grid Systems", BELIEF (Bringing Europe's eLectronic Infrastructures to Expanding Frontiers) videoconference Workshop on Grid Research. São Paulo, em 5 de abril de 2006.

Palestra técnica "KyaTera-WebLabs", proferida a um grupo de pesquisadores e analistas do Instituto de Física, em 09 de maio de 2006.

"Brazilian HEP Grid Initiatives in São Paulo: The São Paulo Regional Analysis Center", 2nd EELA Workshop. Rio de Janeiro, em 24 de junho de 2006.

Palestras técnicas "Projeto GridUNESP" e "USPGrid: Planejamento dd infra-estrutura física", proferidas em reuniões do Grupo Assessor em Grid e HPC da Coordenadoria de Tecnologia de Informação (CTI)

da USP, como exemplo de projeto externo já em andamento para fomentar a discussão referente à implantação de uma infra-estrutura de Grid computacional na Universidade de São Paulo, em 18 de julho e 08 de agosto de 2006, respectivamente.

3.6. Demonstração SC06

A SuperComputing é uma conferência internacional de computação de alto desempenho, redes de computadores, armazenamento e análise de dados, evento que ocorre anualmente. Este ano a cidade de Tampa na Flórida sediou a conferência, de 11 a 17 de novembro [20]. O "show floor" do evento, no Tampa Convention Center, foi conectado às principais redes de pesquisa americanas, em particular a Abilene, ESNet, NLR PacketNet, NLR FrameNet, HOPI e AMPATH.

O 'Bandwidth Challenge' tem sido uma atividade presente no SuperComputing dos últimos 6 anos. Este ano, o 'Bandwidth Challenge' estabeleceu regras que definiram uma abordagem nova em relação às redes de dados de alta velocidade. O objetivo para este ano foi demonstrar a viabilidade de se utilizar links de rede fim a fim, desde o 'show floor' do evento até cada uma das instituições participantes do 'challenge', usando as redes de dados de produção das próprias instituições, ao invés de redes temporárias, especialmente construídas para o evento. O objetivo definido pelo programa deste ano estabeleceu que os participantes deveriam não apenas demonstrar, mas também publicar todas as configurações, análises de erros, ajustes e políticas, visando não só a demonstração, mas também a geração de documentos que permitam que qualquer pesquisador pertencente a uma das instituições participantes alcance os mesmos resultados. Este é um objetivo diferente dos 'challenges' de anos anteriores.

Caltech, CERN, FNAL e SLAC, com o apoio de outras instituições, montaram uma sofisticada infra-estrutura de WAN e sobre ela operaram um conjunto de sistemas distribuídos através da tecnologia de Grid Services. A demonstração, denominada "High Speed Data Gathering, Distribution and Analysis for Physics Discoveries at the Large Hadron Collider", envolveu a movimentação e análise de conjuntos de dados da ordem de terabytes, comumente usados na pesquisa em física de altas energias. Foram demonstradas transferências otimizadas de dados sobre redes de 10 Gbps conectando servidores e sistemas de armazenamento, usando a mais recente tecnologia de processadores e sistemas de armazenamento. O desempenho da rede WAN montada foi monitorado usando o sistema MonALISA, um 'framework' de monitoramento distribuído baseado em agentes inteligentes, desenvolvido por pesquisadores do Caltech. Foi também usado um conjunto de ferramentas de software para análise apropriados para a execução em sistemas interligados em Grid, desenvolvido por pesquisadores do Caltech, University of Florida e University of Michigan. A demonstração também usou ferramentas de gerenciamento de dados do OSG (Open Science Grid) e do EGEE (Enabling Grids for E-sciencE): SRM, dCache, FTS, e PhEdEx. Um protótipo da versão mais recente do pNFS (parallel Network File System) foi também demonstrado, buscando evidenciar o excelente desempenho dessa nova versão de 'file system' distribuído na execução de tarefas de análise e movimentação de dados em conexões de 10 Gbps. O diagrama a seguir apresenta detalhes da infra-estrutura de WAN montada para esta demonstração. O SPRACE participou através da conexão via AMPATH, usando o link WHREN-LILA.

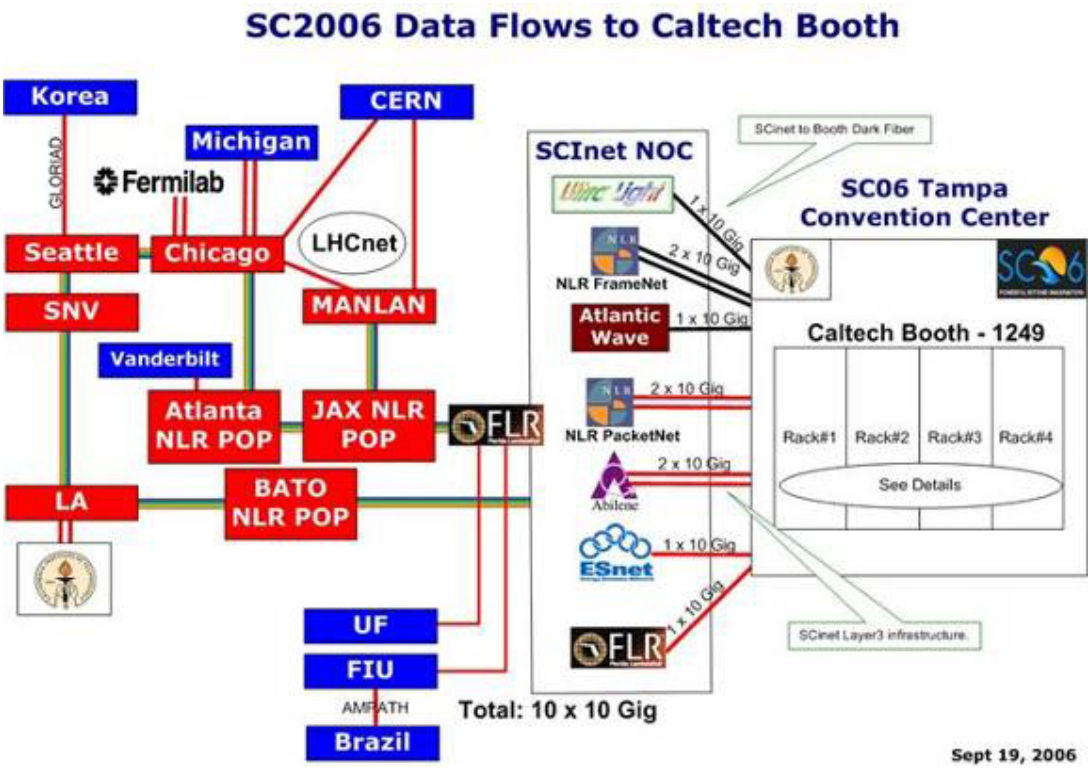

Figura 8: *Esquema da rede WAN montada na demonstração do Caltech, CERN, FNAL, SLAC: "High Speed Data Gathering, Distribution and Analysis for Physics Discoveries at the Large Hadron Collider".*

Os preparativos para a participação do SPRACE na demonstração liderada pelo Caltech incluíram o término da ligação do switch Cisco Catalyst 3750 - 'gateway' de todos os sistemas computacionais do laboratório - aos roteadores do NAP do Brasil em Barueri. Desde final de outubro deste ano, está em plena produção a conexão direta com o uso de um link exclusivo de 1 Gbps, através do uso de um par de equipamentos CWDM instalados pela ANSP na USP e em Barueri, no NAP do Brasil. Cada um desses equipamentos dispõe de um módulo MUX/DEMUX CWDM e dois módulos com 2 portas de 1 Gbps cada. Assim, o Centro de Computação Eletrônica da USP e o NAP do Brasil estão conectados através de 4 canais de 1 Gbps, logicamente independentes entre si, porém compartilhando um mesmo meio físico (um par de fibras fornecido pela CTBC-Telecom). Os canais de 1 Gbps serão disponibilizados para os seguintes usuários: USP, RNP, KyaTera e SPRACE.

Embora o link de 1 Gbps exclusivo seja permanente, para as demonstrações do SC06 tivemos a oportunidade de utilizar temporariamente um segundo link de 1 Gbps, de forma a se gerar uma conexão agregada de 1 + 1 Gbps entre o SPRACE e o NAP do Brasil. Porém, a saída do link internacional WHREN/LILA está localizado em Cotia, não em Barueri, e atualmente existem apenas dois links de fibra ligando Barueri e Cotia: um deles está disponível, enquanto pelo outro trafega parte do tráfego de Internet 'commodity' entre a rede ANSP e o exterior. Este último link, porém, não está operando a plena capacidade. Assim, em princípio seria possível utilizar aproximadamente 0.5 Gbps desse segundo link de fibra, de forma que o SPRACE esteve apto a gerar tráfego até o limite de 1,5 Gbps durante a SC06.

O diagrama a seguir mostra o caminho atualmente usado tanto pelo grupo de São Paulo como pelos grupos do Rio de Janeiro para se conectar ao Cisco ONS 15454 em Cotia, que é o equipamento de saída do link WHREN/LILA. O diagrama também mostra a conexão através do par de equipamentos

CWDM que disponibiliza um link exclusivo de 1 Gbps ao SPRACE. Esta nova conexão já está operacional, desde o final de outubro.

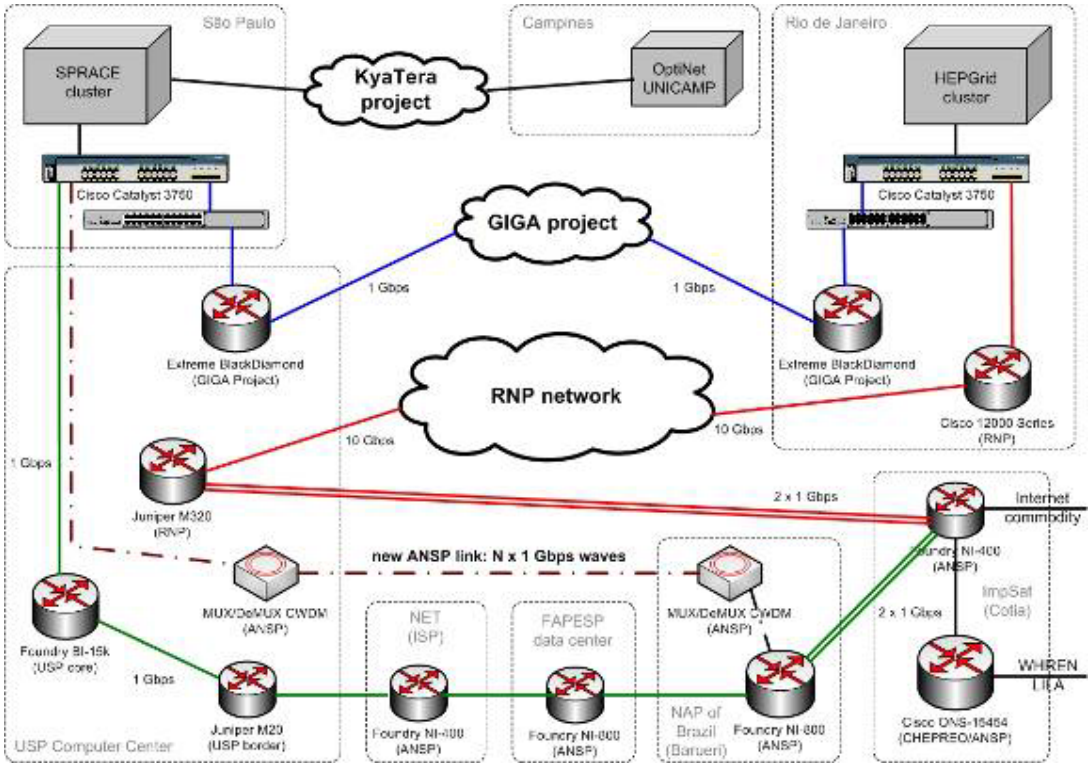

Figura 9: *Diagrama de interconexões de rede do SPRACE*

Na figura anterior, as linhas em azul indicam a conexão entre o SPRACE e os grupos de pesquisa no Rio de Janeiro (UERJ, CBPF e UFRJ, que integram o HEPGrid Brasil), através do projeto GIGA. As linhas em verde indicam a conexão temporária entre o SPRACE o NAP do Brasil em Barueri, sede da ANSP, através do backbone da USPnet, em operação até o final de outubro de 2006. As linhas em vermelho indicam o caminho que os grupos do Rio de Janeiro se conectam ao link internacional WHREN/LILA, através da RNP. As linhas tracejadas representam a conexão direta entre USP e NAP do Brasil em operação desde 31 de outubro, usando um par de switches CWDM da empresa MRV, modelo LD-800. A conexão entre o SPRACE e o laboratório OptiNet na Unicamp através da rede estável do projeto KyaTera está em fase de implantação.

O mapa a seguir mostra os principais links de rede óptica nas Américas: rede CLARA, CAnet4, WHREN/LILA (links leste e oeste), TransLight, Pacific Wave, e Atlantic Wave. A Atlantic-Wave, em fase de implantação, será um 'peering fabric' internacional com o objetivo de interconectar a costa leste dos Estados Unidos com o Canadá, Europa e América do Sul, através de pontos de troca de tráfego localizados em New York, Washington D.C., Atlanta, Miami e São Paulo. Percebe-se através deste mapa a importância do ponto de troca de tráfego em São Paulo, que troca dados com os Estados Unidos via LILA, com a América Latina via rede CLARA, e com a Europa via rede GEANT2. Tal mapa evidencia a posição estratégica do SPRACE, que dispõe de conexão através de diversos pares de fibras de alta qualidade com o coração do 'exchange point' de São Paulo, atualmente localizado no Centro de Computação Eletrônica da USP.

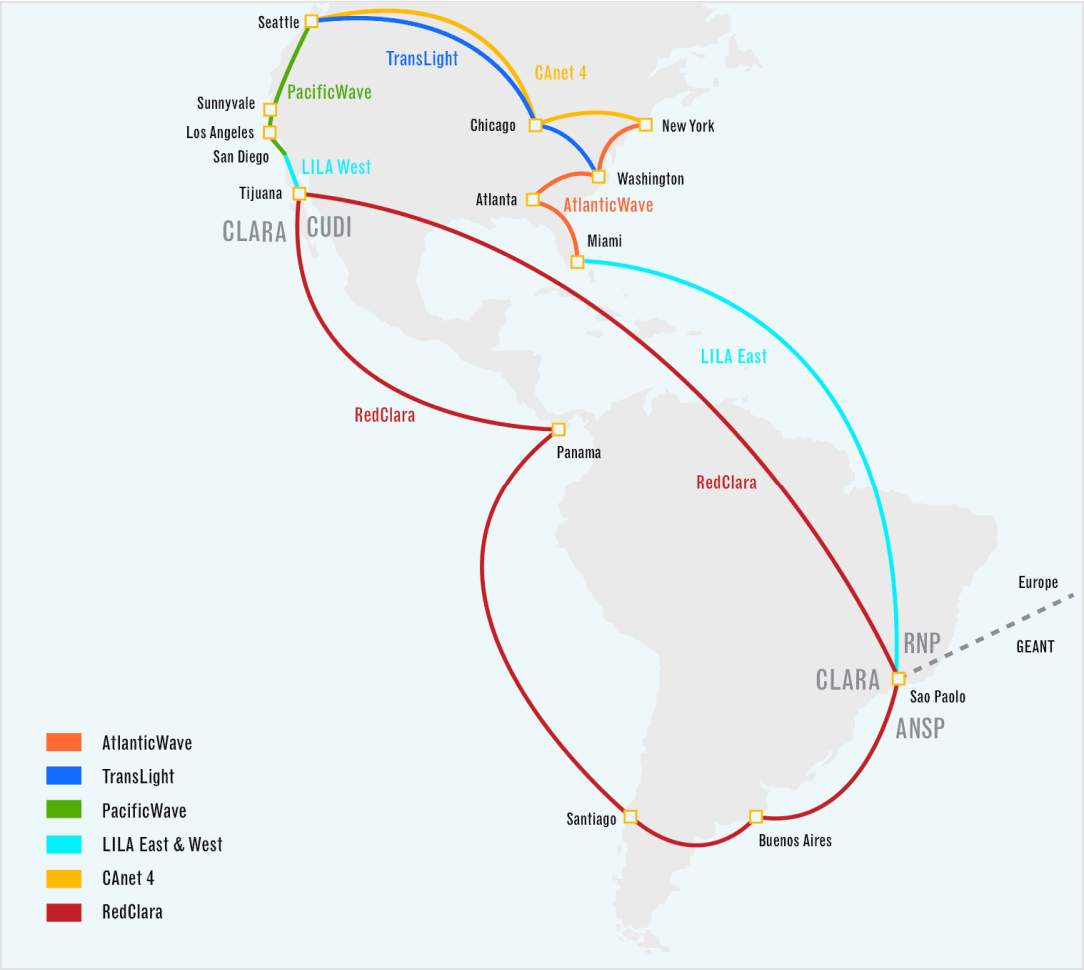

Figura 10: *Principais links de rede óptica nas Américas*

3.7. Outras atividades

- Participação nas reuniões semanais do Grupo Assessor em Grid e HPC da Coordenadoria de Tecnologia de Informação;
- Participação nas reuniões técnicas de discussão da implantação do projeto GridUNESP, como membro convidado;
- Participação em reuniões no Instituto de Física para discutir a inclusão do Instituto no projeto KyaTera através da submissão de projeto de automatização de experimentos (WebLabs);
- Instalação da ferramenta 'pinger' (ping end-to-end performance measurement) no servidor de monitoração, que permite que pesquisadores do SLAC possam coletar informações, monitorar e gerar estatísticas de conectividade da Internet na América Latina a partir da rede do SPRACE;
- Organização e acompanhamento da visita do Dr. Xun Su do Caltech ao SPRACE;
- Auxílio na escrita e preparação de diversos 'press-releases', a pedido de fornecedores;
- Auxílio na implementação e uso da ferramenta Wiki do grupo;
- Levantamento detalhado dos recursos do SPRACE para elaboração de inventário;
- Elaboração de diversos diagramas lógicos e esquemas, referentes à infra-estrutura computacional e de rede do laboratório, além de layouts de disposição dos racks na sala dos servidores, e layouts da estrutura da laje sobre a qual se apóiam os racks;
- Contatos com engenheiros da COESF para estudo de carga da laje do CEPA, para conferir se a estrutura física da laje teria capacidade para receber a fase III do projeto;
- Aquisição de 'headsets' para uso em videoconferências via VRVS e de mesa móvel para a sala de servidores;
- Aquisição de conversor de mídia 1000BaseLX <-> 1000Base-T da marca Planet, modelo GT702s, para permitir conectar servidores diretamente às fibras ópticas disponíveis no laboratório;
- Contatos técnicos permanentes com fornecedores e colegas de outras unidades da USP, e participação em feiras e eventos relacionados à tecnologia da informação, para atualização tecnológica.
- Submissão de artigo científico ao evento "Agent-based Grid Computing Workshop", com o título "Classes of Service for Grid Scheduling using Ant Colony Optimization", a ser realizado em maio de 2007. Artigo escrito em conjunto com o Dr. Gustavo Pavani da Unicamp (ainda em análise).

5. Planejamento para o próximo período

As atividades do dia-a-dia, ou seja, a manutenção da operação dos sistemas de forma ininterrupta, o processo de monitoração dos servidores e serviços, as manutenções eventuais e programadas, sendo atividades rotineiras e permanentes, deverão ocorrer no próximo período de forma bastante semelhante ao que foi realizado no decorrer de 2006.

Essas atividades diárias têm me permitido usar todos os conhecimentos práticos que adquiri ao longo de 4 anos exercendo as funções de analista de sistemas e de analista de redes junto ao Centro de Computação Eletrônica da USP. Além disso, minha formação como bacharel em Física habilita-me a compreender os conceitos fundamentais envolvidos nos experimentos e nas análises físicas. Os complexos sistemas eletrônicos e computacionais que fazem parte do experimento do CERN são por demais atraentes e representam um convite ao estudo, graças à experiência que adquiri no trato com sistemas eletrônicos como técnico em eletrônica, cargo que ocupei na Itautec durante 7 anos. As tecnologias de computação em Grid e as interconexões dos sistemas através de redes ópticas trans-nacionais encaixam-se com perfeição às minhas atividades acadêmicas atuais, como estudante de pós-graduação da Escola Politécnica.

Minha participação no projeto de implantação do GridUNESP deverá se tornar mais intensa no próximo período, bem como minha participação e contribuições nas discussões referentes à implantação de uma infra-estrutura de Grid computacional na Universidade de São Paulo.

Para o início de fevereiro, está previsto um Workshop interno do grupo, em que cada membro irá fazer apresentações dos trabalhos em que estão envolvidos. Minhas apresentações terão como objetivo fazer um resumo dos treinamentos que participei no ano de 2006.

Na função que exerço atualmente, consigo explorar plenamente todas as habilidades e experiências que adquiri ao longo de minha vida profissional, o que pode ser traduzido por uma sensação de realização e de grande apreço pelo trabalho que atualmente exerço, como auxiliar de pesquisa de um grupo em plena ascensão.

6. Referências

(a) Bibliografia consultada:

- Foster, I., “The Grid: A New Infrastructure for 21st Century Science”, *Physics Today*, 55(2):42-47, 2002.
- Foster, I., Kesselman, K.,“The Grid: Blueprint for a New Computing Infrastructure”, Morgan Kaufmann, 2004.
- Berman, F., Hey, A. J. G., Fox, G. C., “Grid Computing: Making the Global Infrastructure a Reality”, John Wiley & Sons, 2003.

(b) Citações:

[1] http://www.chepreo.org

[2] http://www.lanautilus.com

[3] http://abilene.internet2.edu

[4] http://www.ucaid.org

[5] http://www.canarie.ca/canet4/

[6] http://www.terena.nl

[7] http://www.apan.net

[8] http://whren.ampath.net

[9] http://www.ampath.fiu.edu

[10] http://www.glif.is/

[11] http://monalisa.caltech.edu/

[12] http://redesopticas.reuna.cl

[13] http://www.intel.com/design/intarch/dualcorexeon/benchmark_brief.pdf

[14] http://hep.ift.unesp.br/ILHCWorkshop/index.html

[15] http://indico.eu-eela.org/conferenceTimeTable.py?confId=37

[16] http://www.eu-eela.org/

[17] http://www.eu-egee.org/

[18] http://csc.web.cern.ch/CSC/

[19] http://www.eu-eela.org/egris1/

[20] http://sc06.supercomputing.org/

São Paulo, 18 de dezembro de 2006

_____

Rogério Luiz Iope

Funcional Nº 1712076

_____

Prof. Gil da Costa Marques
Coordenador
Centro de Ensino e Pesquisa Aplicada
Departamento de Física Experimental
Universidade de São Paulo