




Projeto Kyatera – Camada de Rede

Tereza Cristina M. B. Carvalho
CCE-USP, LARC-EPUSP
carvalho@larc.usp.br

Agenda

- 
- A large, semi-transparent orange circle with a white center is positioned on the left side of the slide.
- Parte I
 - Projeto Lógico das Redes Experimental e Estável.
 - Parte II
 - Redes Emuladas – Emulab
 - Parte III
 - PlanetLab
 - Considerações Finais

Projeto Kyatera

Rede GT2 e Rede GT3

Tereza Cristina M. B. Carvalho

Marcio Augusto de Lima e Silva

Fernando Frota Redigolo, Flávio Urschei

Marcos Tork Souza

<carvalho,msilva,fernando,furschei,mtork@larc.usp.br>

LARC/PCS/Escola Politécnica – USP

Jorge Futoshi Yamamoto, José Pedro

<yamamoto,jpolive@ansp.br>, NARA/ANSP

Pablo Siré - <pazsire@ifi.unicamp.br>

Optical Systems Management – Kyatera Optical Testbed



Agenda

- Disposição Geográfica
- Disposição Funcional
- Princípios Diretores GT2 (Rede Experimental)
- Princípios Diretores GT3 (Rede Estável)
- Premissas Iniciais
- Apresentação do Projeto



Disposição Geográfica



Disposição Funcional

Rede Experimental
Grupo de Trabalho 1

Fibra 0 - N

Rede Experimental
Grupo de Trabalho 2

Roteadores/
Comutadores/Redes
Emuladas Kyatera

Rede Estável
GT 2 (Camada 1)
Fibra N+1 – N+M

Rede Experimental
Grupo de Trabalho 3

Weblabs\E-learning

Rede Estável
GT 3 (Camada 2 - 7)
Roteadores/Comutadores
NARA/ANSP

Rede Estável
GT 3 (Camada 1)
Fibra N+M+1 – N+P

Princípios Diretores GT2 (Rede Experimental)

- A rede deve permitir que se disponham os equipamentos em **múltiplas topologias** (i.e. ponto a ponto, anel e malha).
- Os **equipamentos** necessários para implementação de circuitos de comunicação, virtuais ou físicos, devem ser **distribuídos** entre as localidades, possuindo todas as **funcionalidades** requeridas pelas **tecnologias habilitantes de pesquisa**.
- Deve haver possibilidade de **trocas de tráfego** entre localidades pelo roteamento e/ou comutação nas **camadas de rede, de enlace e física** da pilha de protocolos OSI (e.g. VPNs, VLANs e “lightpaths”).
- A rede deve ser arquitetural e topologicamente passível de acomodação de **equipamentos de rede, óticos ou eletrônicos** (e.g. Ethernet/IP) representativos do “**estado da arte**” em quaisquer configurações.
- O núcleo da rede deve ser provido de **equipamentos** que possibilitem o **aumento do número de circuitos de comunicação**, bem como o **aumento da capacidade** de cada um **destes circuitos**, da maneira menos custosa permitida pela tecnologia atual.
- Deve ser possível **acrescentar novas localidades** à rede da maneira menos custosa e disruptiva provida pela tecnologia atual.
- A rede física deve **servir de suporte**, e se **integrar plenamente**, aos equipamentos (“clusters” e geradores/analísadores de tráfego) que implementam as **redes emuladas**.



Princípios Diretores GT3 (Rede Estável)

- A rede deve **minimizar ou mesmo eliminar interrupções na sua operação**, obedecendo a rígidos controles para estabelecimento de janelas de **manutenção preventiva** e para estabelecimento de tempos de resposta mínimos para **manutenção emergencial**.
- Os **equipamentos** necessários para implementação de circuitos de comunicação, virtuais ou físicos, **de maior complexidade e com maior número de funcionalidades**, serão **concentrados apenas** nas localidades onde já há laboratórios e equipamentos da rede **GT2**.
- Nas **localidades de concentração de equipamentos**, deve haver possibilidade de troca de tráfego pelo **roteamento e/ou comutação nas camadas de rede e de enlace** do modelo OSI (e.g. VPNs e VLANs).
- Para garantir alta capilaridade a partir das **localidades de concentração**, todas as localidades do GT3 deverão dispor de **equipamentos para a troca de tráfego, ao menos, na camada de enlace** (i.e. VLANs).
- Deve haver **possibilidade de conexão**, sob demanda e automatizada, entre **as redes GT3 e GT2**.



Premissas Iniciais (I)

- Em virtude de restrições de toda ordem (orçamentária, política, operacional e tecnológica) **as redes GT2 e GT3** deverão **compartilhar os meios de transmissão** (i.e. os mesmos pares de fibras) e os **componentes de transmissão óptica**.
- É **facultada** aos projetistas a decisão de **compartilhamento** também **dos equipamentos** de rede (e.g. comutação de quadros e roteamento de pacotes).
- Em decorrência da premissa 1 e 2, as **redes GT2 e GT3** deverão ser implementadas como **subdivisões lógicas** (virtuais) de uma única **rede real** (i.e. física) **Kyatera**.
- Uma **entidade específica** será responsável pelo **gerenciamento** (e.g. monitoração e operação) **da rede**.

Premissas Iniciais (II)

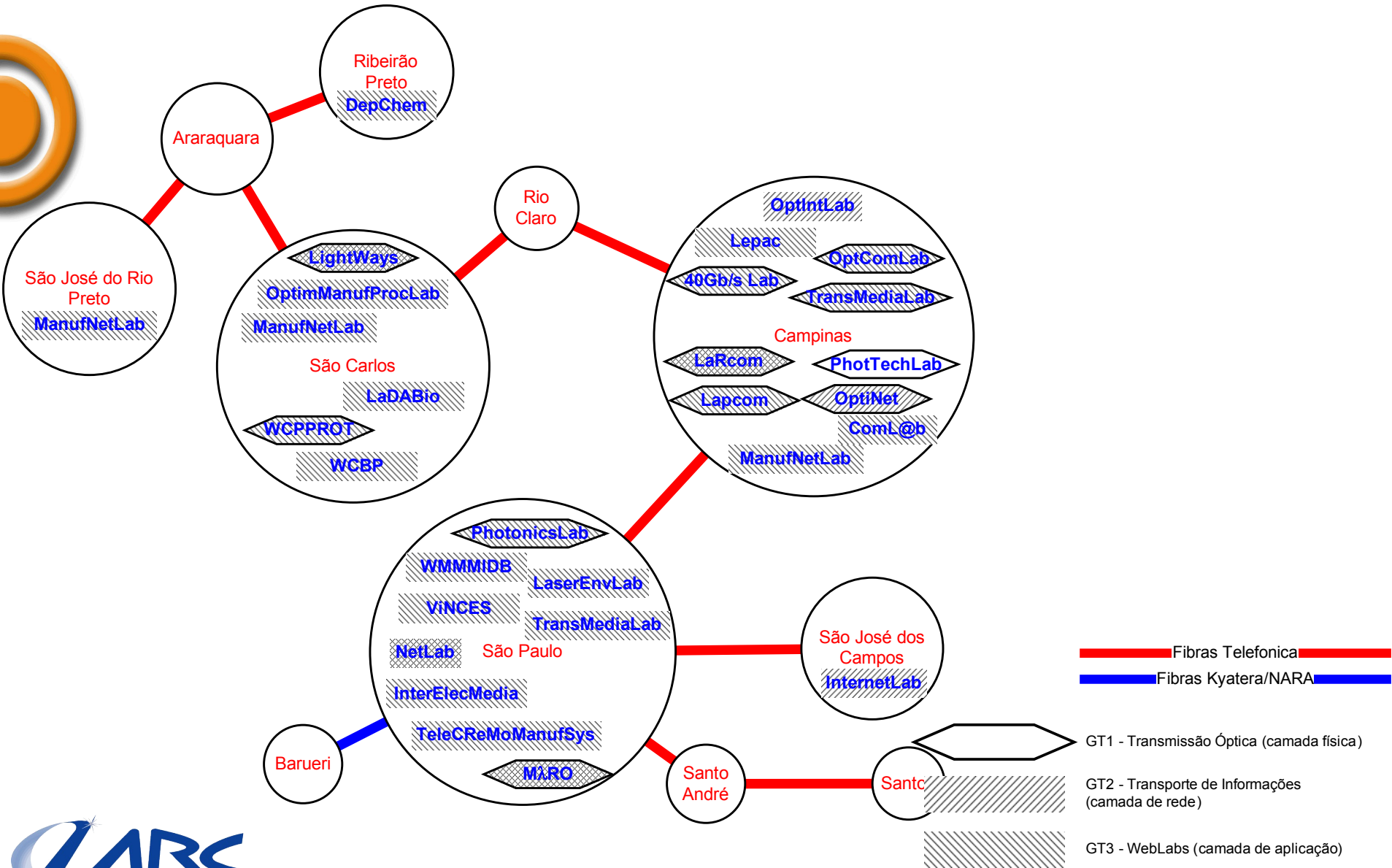
- Os laboratórios e instituições conectados a rede ora definida são aqueles participantes do **TIDIA – Kyatera (GT2/GT3)**, **TIDIA – Aprendizado Eletrônico** e **TIDIA – Incubadora Virtual**. Estes dois últimos serão considerados, para os fins deste projeto, como equivalentes àqueles pertencentes ao TIDIA – Kyatera – GT3.
- A **rede lógica GT3** (e.g. o “lambda” que a implementa em um equipamento DWDM) deverá ter um **núcleo com taxa de transmissão mínima de 1 Gbps**.
- A **rede lógica GT2** deverá ter um **núcleo com taxa de transmissão mínima de 10 Gbps**.
- Todo o **tráfego gerado** pela rede deverá ser **concentrado em três pontos específicos (SP, SC e Campinas)**. Esta premissa baseia-se nos princípios diretores das redes GT2 e GT3.
- Não existem restrições sobre as topologias das redes conectadas** a cada **ponto de concentração de tráfego**, admitindo-se a sua disposição em malha, anel ou árvore.



Premissas Iniciais (III)

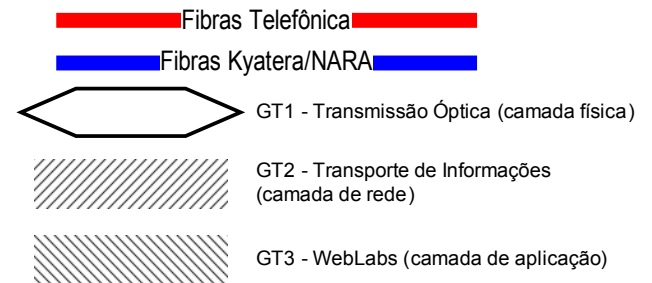
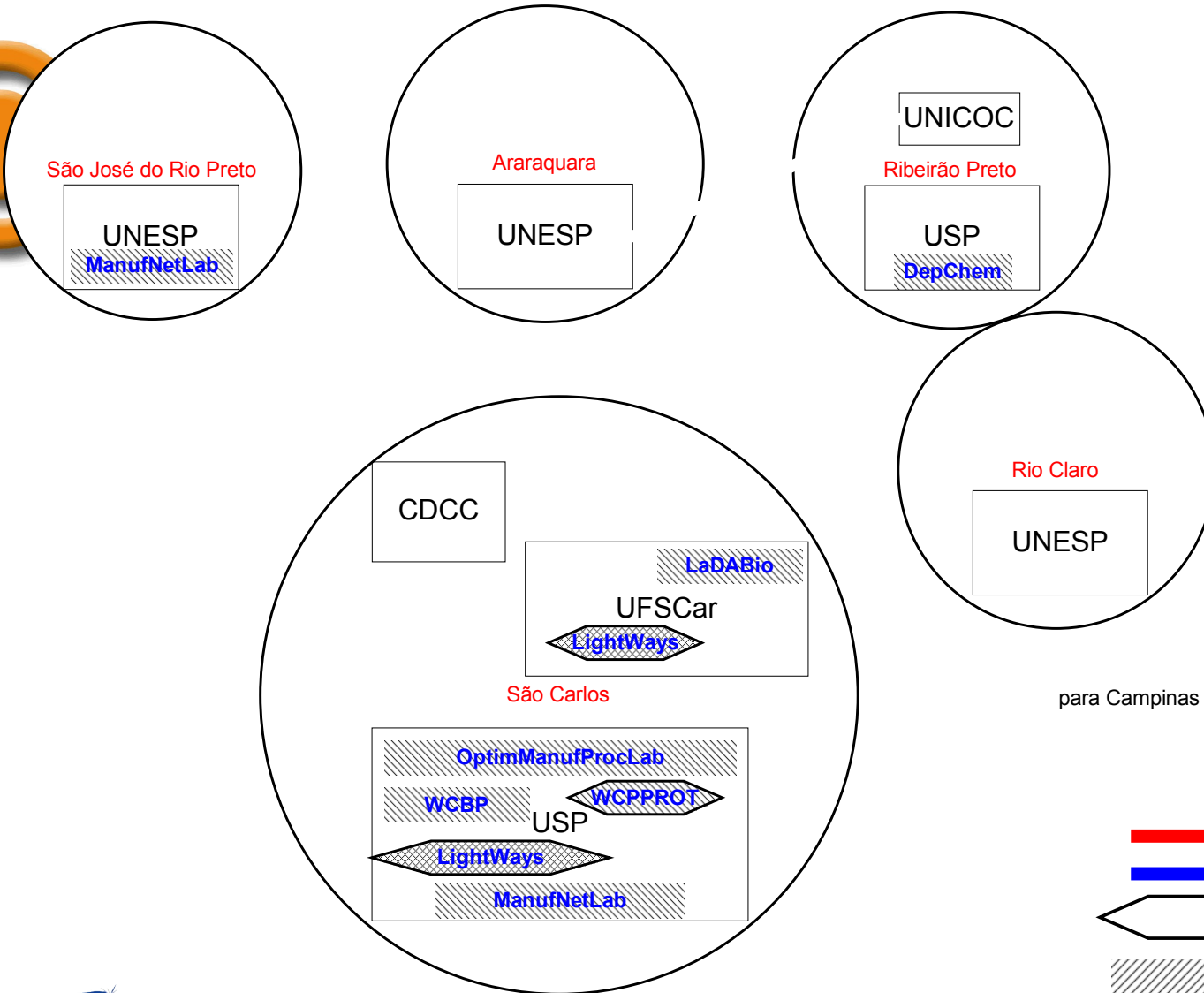
- Cada um dos laboratórios e instituições participantes do KyaTera se **conectará** à rede real através de **um ou mais pontos de concentração** citados, por meio de **equipamentos** que operam na **camada física** (e.g. conversores de mídia), na **camada de enlace** (e.g. comutadores) ou na **camada de rede** (e.g. roteadores) do modelo OSI.
- A **rede KyaTera** real, nos trechos implementados com o **uso de fibras cedidas pela Telefonica** (adequadamente identificadas no texto) **não poderá transportar qualquer tráfego destinado à Internet Pública, ou dela originado.**

Distribuição Geográfica dos Laboratórios – Nível 0



Distribuição Geográfica dos Laboratórios – Nivel 1 (I)

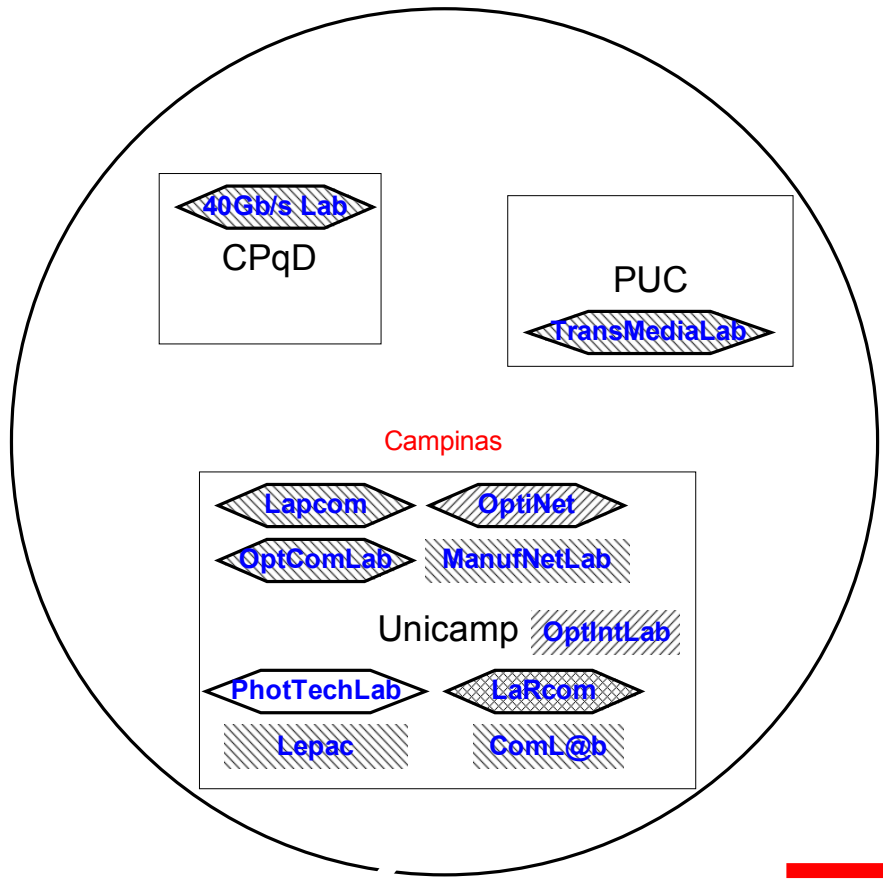
São Carlos



Distribuição Geográfica dos Laboratórios – Nível 1 (II)

Campinas

para Rio Claro



para São Paulo

Fibras Telefônica

Fibras Kyatera/NARA

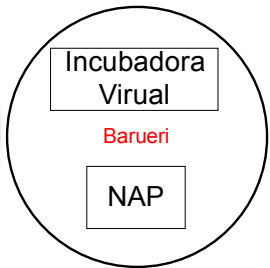
GT1 - Transmissão Óptica (camada física)

GT2 - Transporte de Informações (camada de rede)

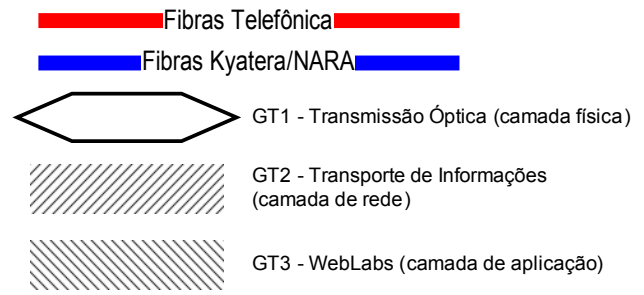
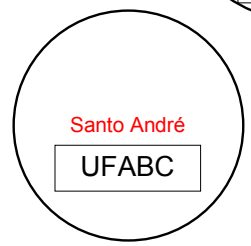
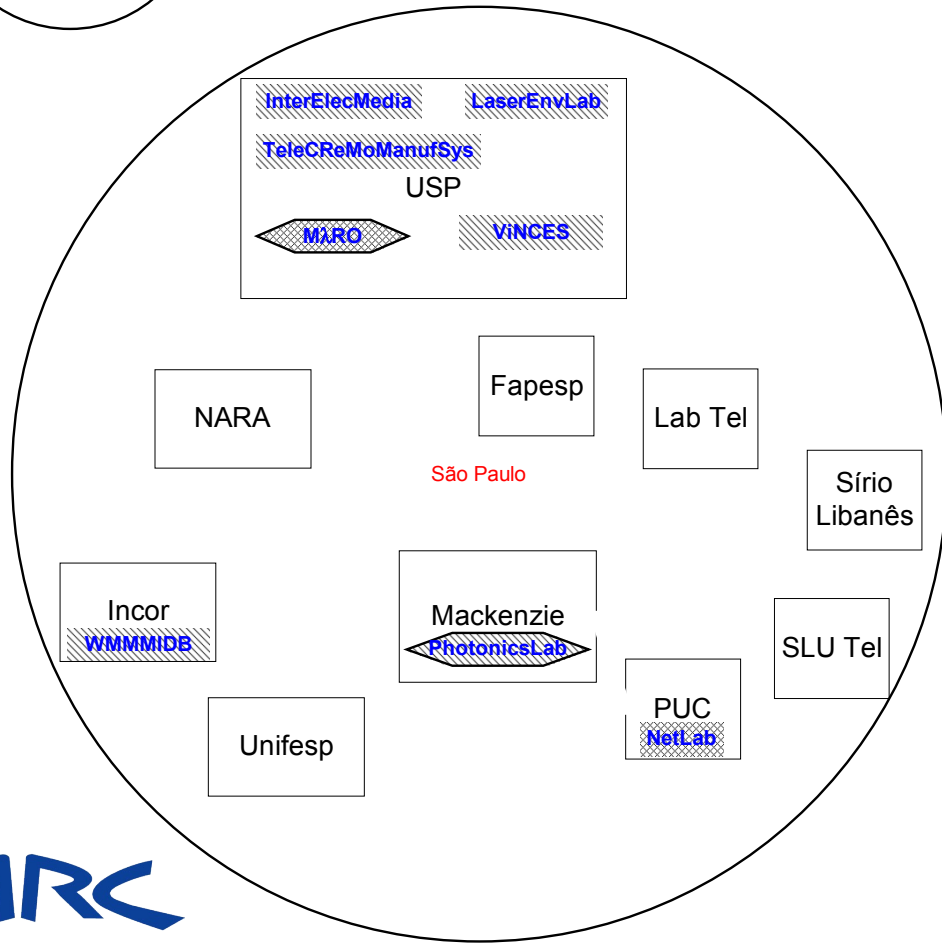
GT3 - WebLabs (camada de aplicação)

Distribuição Geográfica dos Laboratórios – Nível 1 (III)

São Paulo



para Campinas

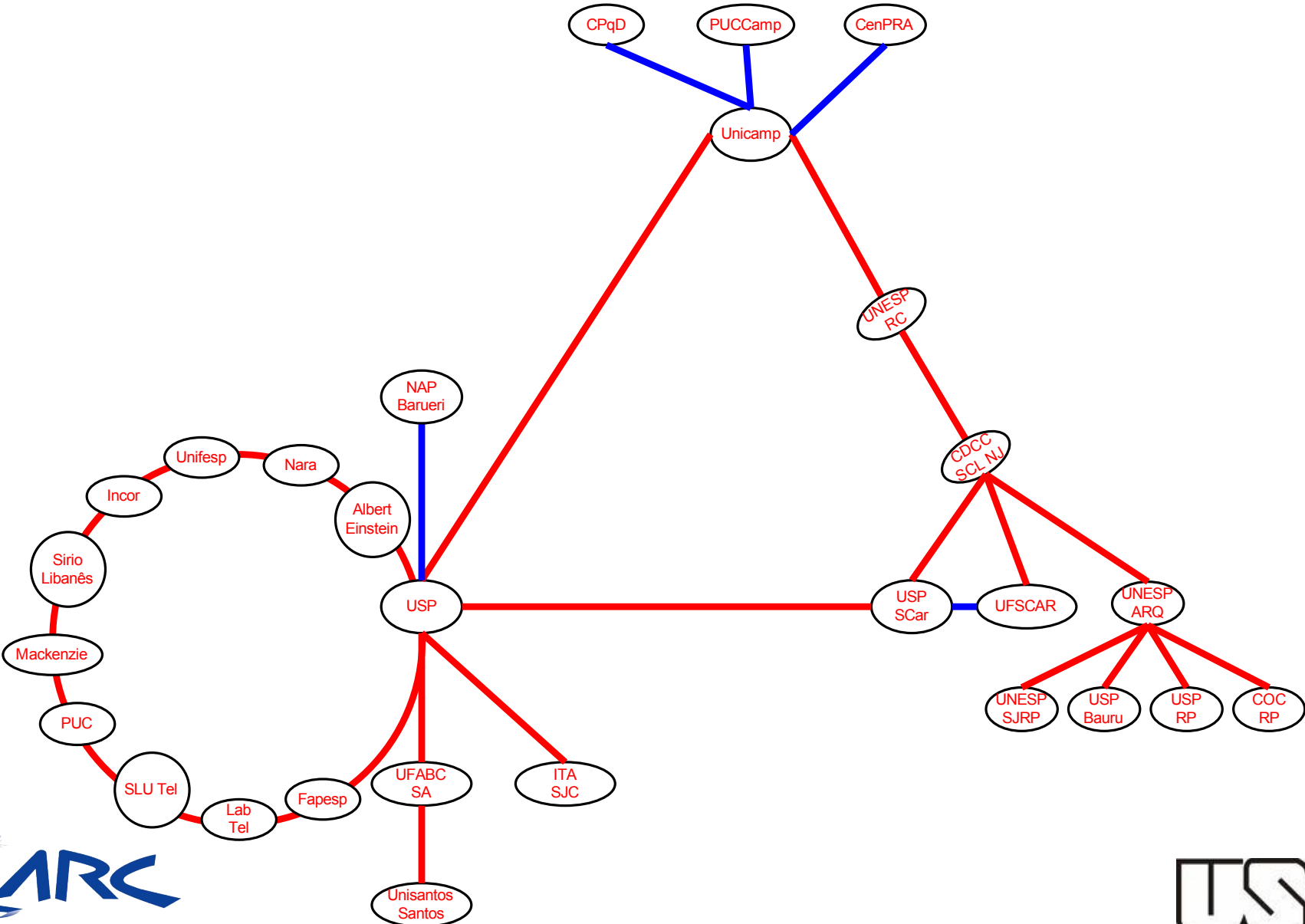


Explicação de Projeto







- Com base nos princípios diretores GT2/GT3, nas premissas iniciais e nas informações da rede física obtidas, as decisões de projeto serão explicitadas.
- As decisões de projeto são aquelas concernentes a vários aspectos da rede, cuja divisão arbitrária é ora declinada:
 - Decisão de Projeto A: Topologia
 - Decisão de Projeto B: Segmentação
 - Decisão de Projeto C: Roteamento
 - Opção de Acesso 1
 - Opção de Acesso 2
 - Opção de Acesso 3
 - Decisão de Projeto D: Equipamentos
- Em adição aos aspectos que são objetos das decisões de projeto, premissas adicionais a estas serão oportunamente explanadas.



Decisão de Projeto A - Topologia

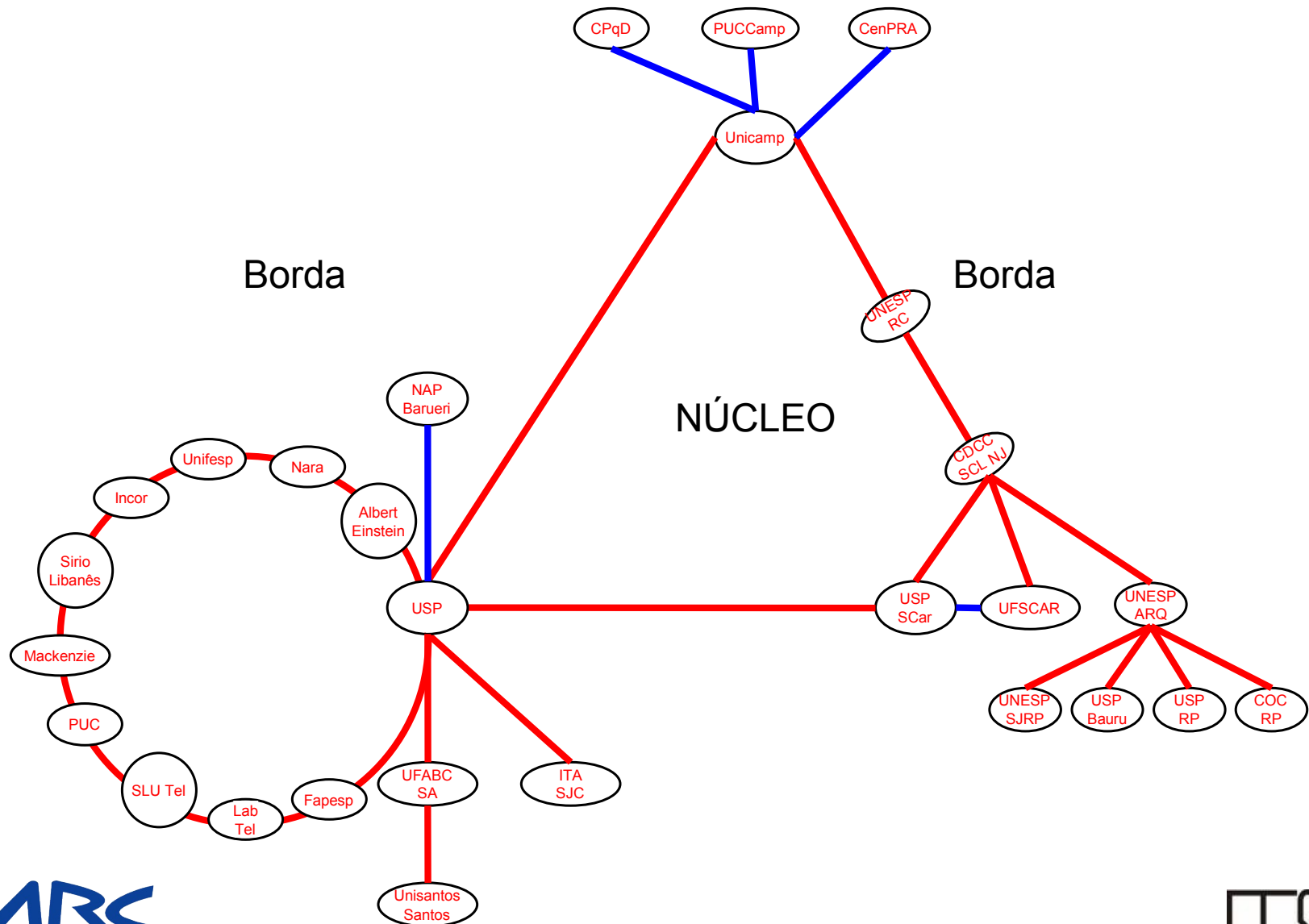


Decisão de Projeto A (Topologia) - Premissas

- 
 A **topologia** a ser implementada é a de um **anel**, formado a partir da interconexão das **três localidades de concentração de tráfego**.
- 
 O **anel** principal é aquele formado pelas localidades **São Paulo, Campinas e São Carlos**. Neste anel, **dois circuitos logicamente separados**, respectivamente capazes de transmitir **1 (GT3)** e **10 (GT2)** Gbps, serão estabelecidos.
- 
 Em adição ao anel principal, um **segundo anel**, compreendendo **laboratórios e instituições na localidade de concentração de tráfego São Paulo**, será implementado.
- 
 A **topologia em todas as outras regiões de borda** será em **estrela**.
- 
 A **rede (lógica) GT3** deve ser considerada uma via opaca de dados (i.e. uma rede para grupos que a tenham como ferramenta, e não como objeto de pesquisa). Como consequência direta deste fato, **não será facultada qualquer tipo de reconfiguração e/ou interrupção** da mesma.
- 
 A **rede (lógica) GT2**, na qualidade de embrião de uma futura rede física GT2, deve **permitir reconfigurações e/ou interrupções**, desde que as mesmas não afetem a disponibilidade da rede GT3.



Decisão de Projeto B - Segmentação



Decisão de Projeto B (Segmentação)

Premissas

- A **rede** será doravante dividida em **duas regiões**, a região de núcleo, que compreende o **anel principal** e a **região de borda**, a qual compreende todos os outros laboratórios e instituições.
- Cada ponto **componente do núcleo** deverá dispor de **Equipamentos de Núcleo (EN)**, que serão aqueles com a **maior capacidade de comunicação** e **maior rol de funcionalidades**.
- Em cada ponto **componente da borda** serão instalados os **Equipamentos de Borda (EB)**, cujo tipo pode variar entre:
 - **Conversor de mídia** (camada 1 OSI)
 - **Comutador** (camada 2 OSI)
 - **Roteador** (camada 3 OSI)

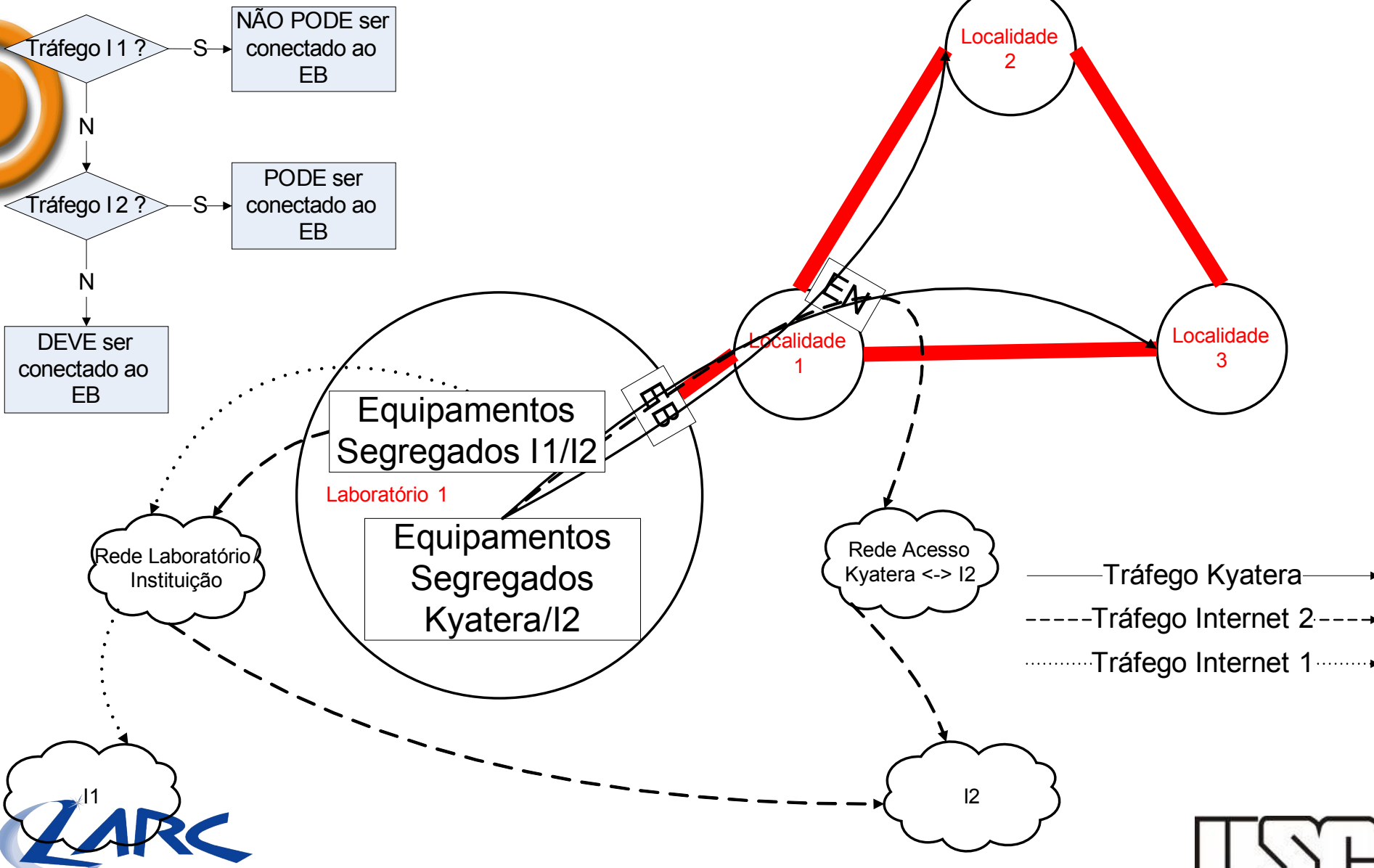


Decisão de Projeto C – Roteamento

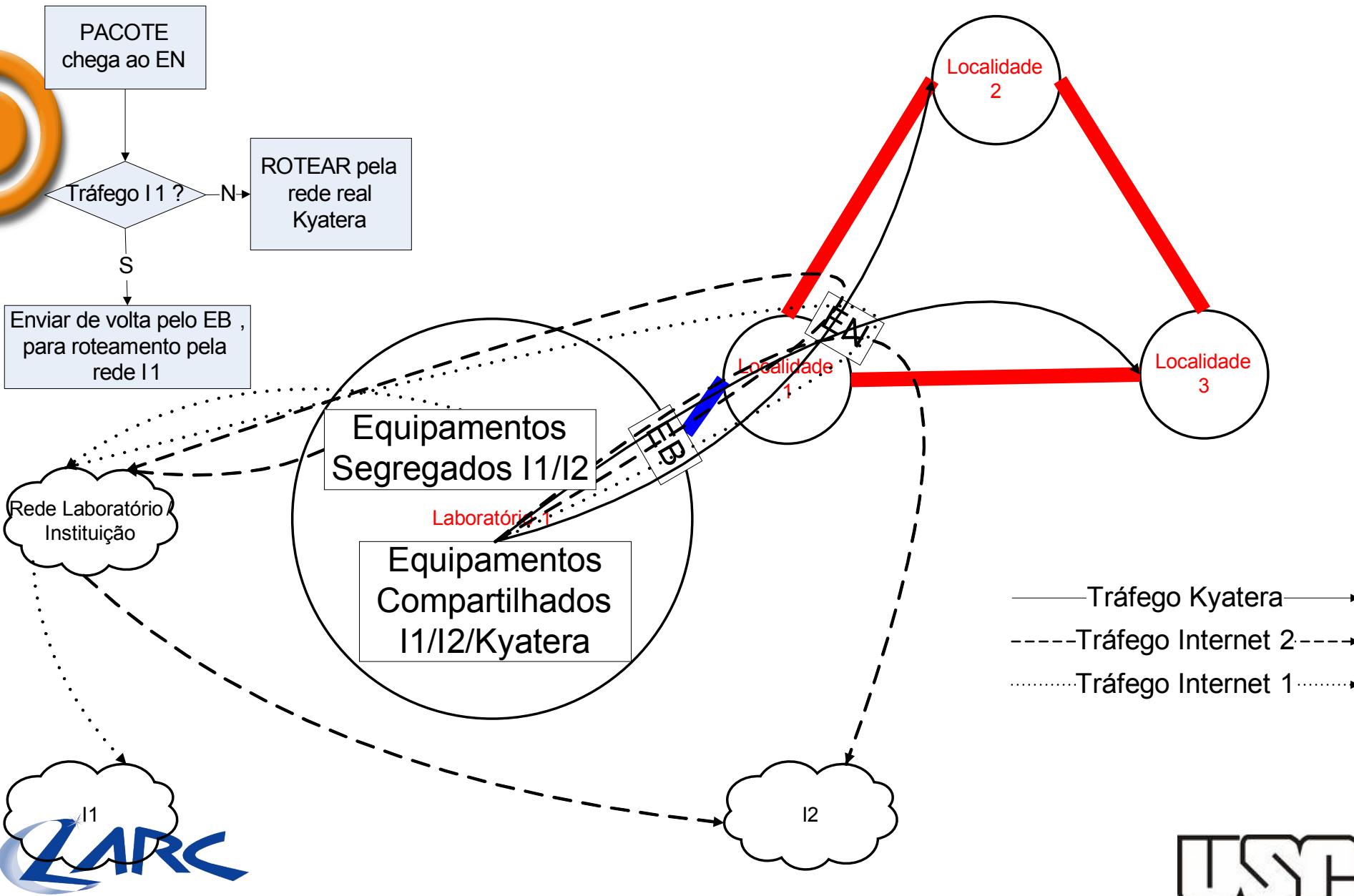
- Três possíveis **casos de acesso**:
 - Acesso à Internet Pública (i.e. I1)
 - Acesso à Internet 2 (i.e. I2)
 - Acesso à Rede Kyatera (i.e. K, representando GT2 e GT3)
- **Categorias de tráfegos** permitidos no **anel** Kyatera: K e I2 somente
- Três possíveis **classes de equipamentos** nas bordas:
 - Conversor de Mídia
 - Computador
 - Roteador



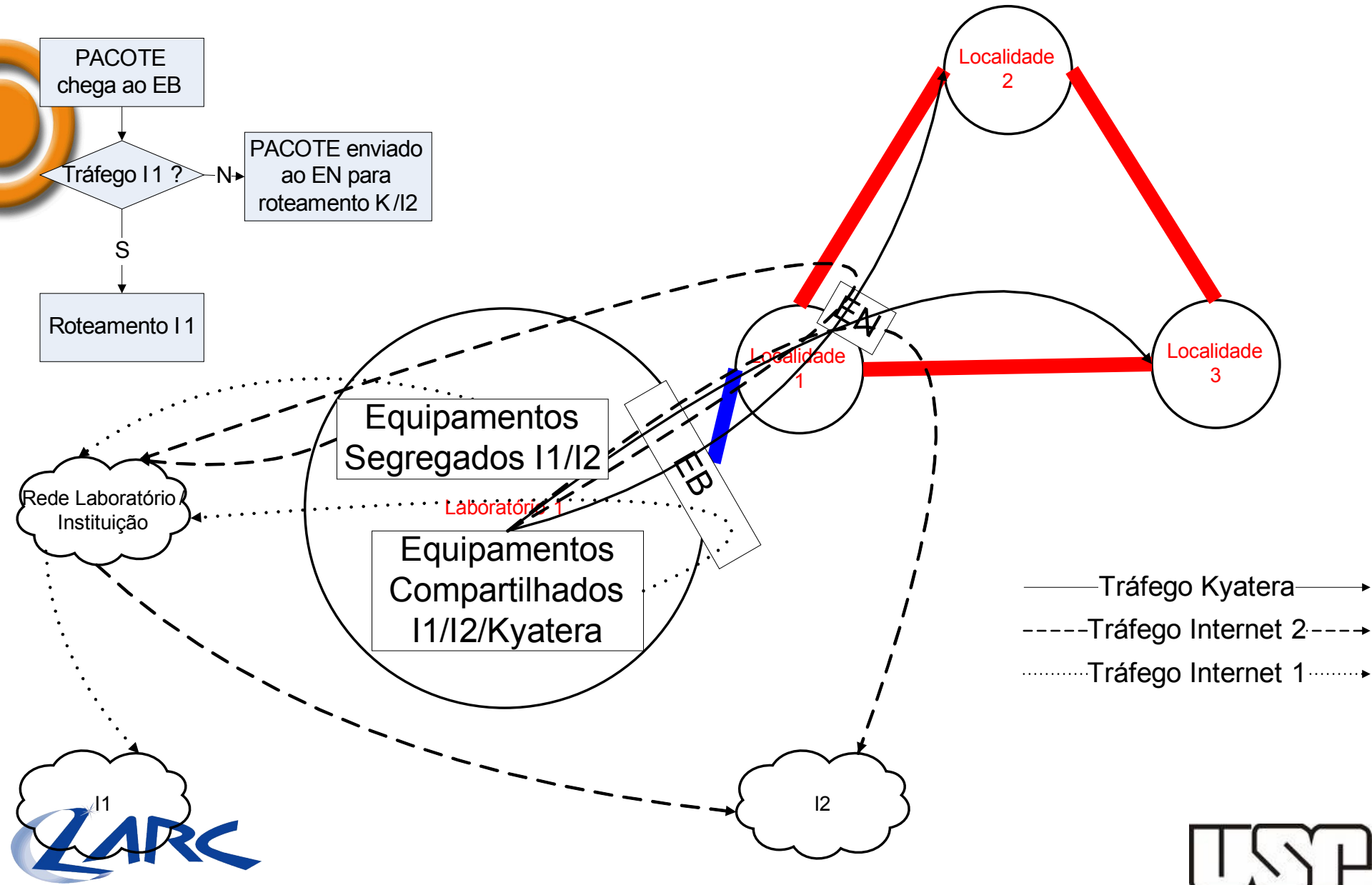
Decisão de Projeto C (Roteamento/Opção 1) – Croquis



Decisão de Projeto C (Roteamento/Opção 2) – Croquis



Decisão de Projeto C (Roteamento/Opção 3) – Croquis

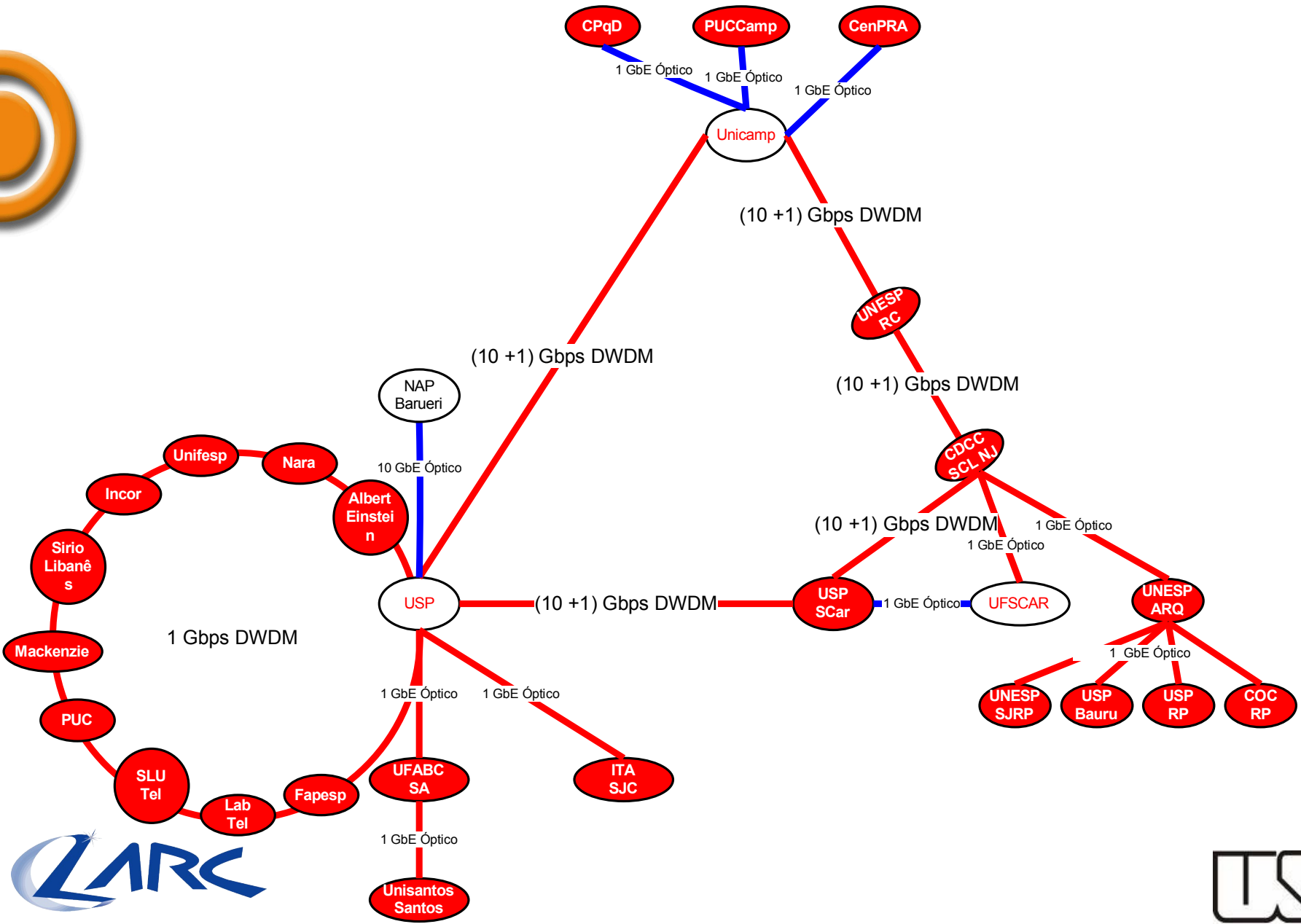



Decisão de Projeto D – Equipamentos

- Núcleo
 - a. Dois anéis lógicos DWDM.
 - Anel com taxa de transmissão de 10 Gbps, embrião da rede GT2.
 - Anel com taxa de transmissão de 1 Gbps, compreendendo a rede GT3.
 - Equipamentos de Núcleo para transmissão óptica: Multiplexadores/Demultiplexadores DWDM acoplados a “*transponders*” Gigabit Ethernet (1 GbE) e 10 Gigabit Ethernet (10 GbE), para conversão de sinal Óptico(850/1310nm)/Elétrico em comprimento de onda DWDM (grade ITU-T).
 - Equipamentos de Núcleo para conexão de rede: Roteador ou Comutador Multicamada com portas 1/10 GbE ópticas/elétricas para ligação ao anel, além de portas 1 GbE ópticas/elétricas para coleta de tráfego de borda.
- Borda:
 - a. Laboratórios e Instituições: Conversor de Mídia, Comutador ou Roteador/Comutador Multicamada, com porta 1 GbE óptica/elétrica.
 - b. Anel secundário em São Paulo: Em adição ao equipamento acima citado, equipamentos de transmissão óptica DWDM, a semelhança dos explicitados para o núcleo, com taxa limitada 1 Gbps.



Decisão de Projeto D (Equipamentos) - Croquis





Emulab: uma Plataforma Pública para Pesquisa em Sistemas Distribuídos e Redes

Tereza Cristina Melo de Brito Carvalho – carvalho@larc.usp.br

Fernando Frota Redígolo – fernando@larc.usp.br

Joelle Quaini Sousa – joelle@larc.usp.br

Márcio Augusto de Lima e Silva – msilva@larc.usp.br

Marcos Tork – mtork@larc.usp.br

O que é o Emulab ?



○ Simulação em redes

- Largamente utilizado
- Elementos abstratos representando o mundo real
- Ex.: ns-2, opnet

○ **Emulab**: bancada de testes para emulação de redes

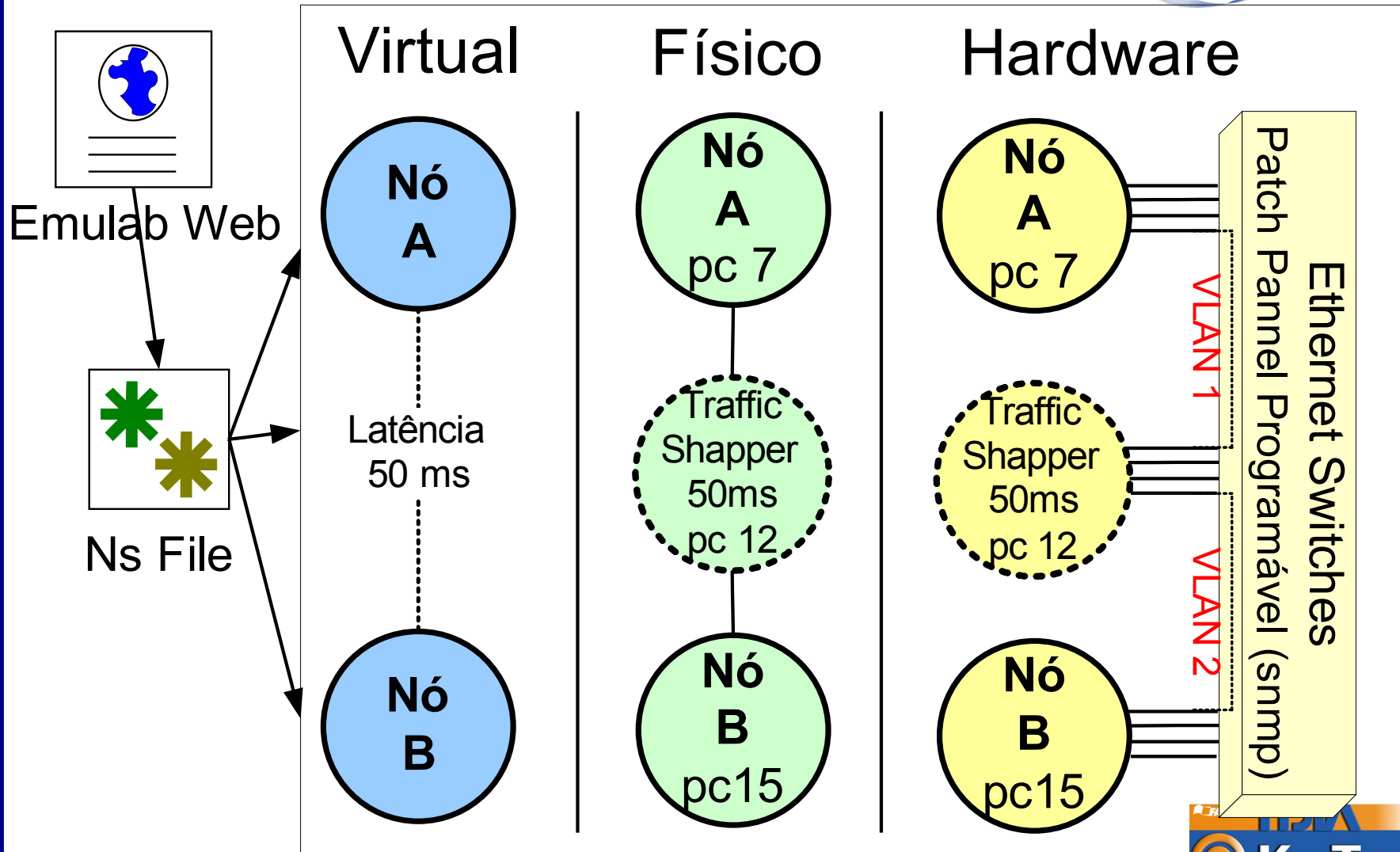
- **Solução híbrida, envolvendo elementos abstratos e reais**
- **Núcleo composto** por hardwares: **nós**, **switches** e **roteadores**;
- **Softwares**: de configuração, gerenciamento, armazenamento
- Usuários recebem nós dedicados com acesso irrestrito (root access)
- Acesso completamente remoto (ssh e serial console)
- **Segurança**
- **Criação automática da topologia**, nós e links
- **Experiências remotas em escala e condições reais**



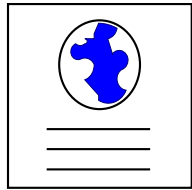
- Utilizado para ensino e pesquisa
- 14 em operação ao redor do mundo (+ 5 em construção)
 - EUA:
 - Berkeley, Georgia Tech, Cornell, Univ. Texas (Austin)...
 - Patrocínio: NSF
 - Outros: Canadá, Malásia, Taiwan, Coreia
 - América Latina: Brasil
- Maior: University of Utah
 - 350 Nós experimentais – Aberto para pesquisadores externos e ensino.
- Brasil: LARC – Laboratório de Arquitetura e Redes de Computadores – USP
 - 16 Nós experimentais
 - Disponível em www.emulab.larc.usp.br

Processo de Criação de Experimentos

Experimento

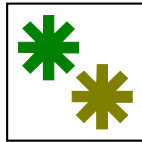


Processo de Criação de Experimentos



Web

Compilar
ns File



Mapeamento

Prover Acesso

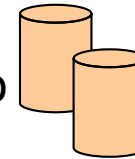


Alocação

Nomeação

Endereços
IP

Carregamento



IP

Configuração

DISCOS

VLAN

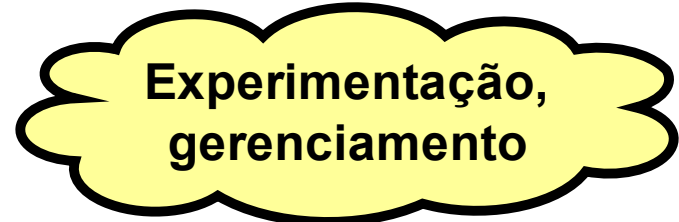
Configuração

Reboot

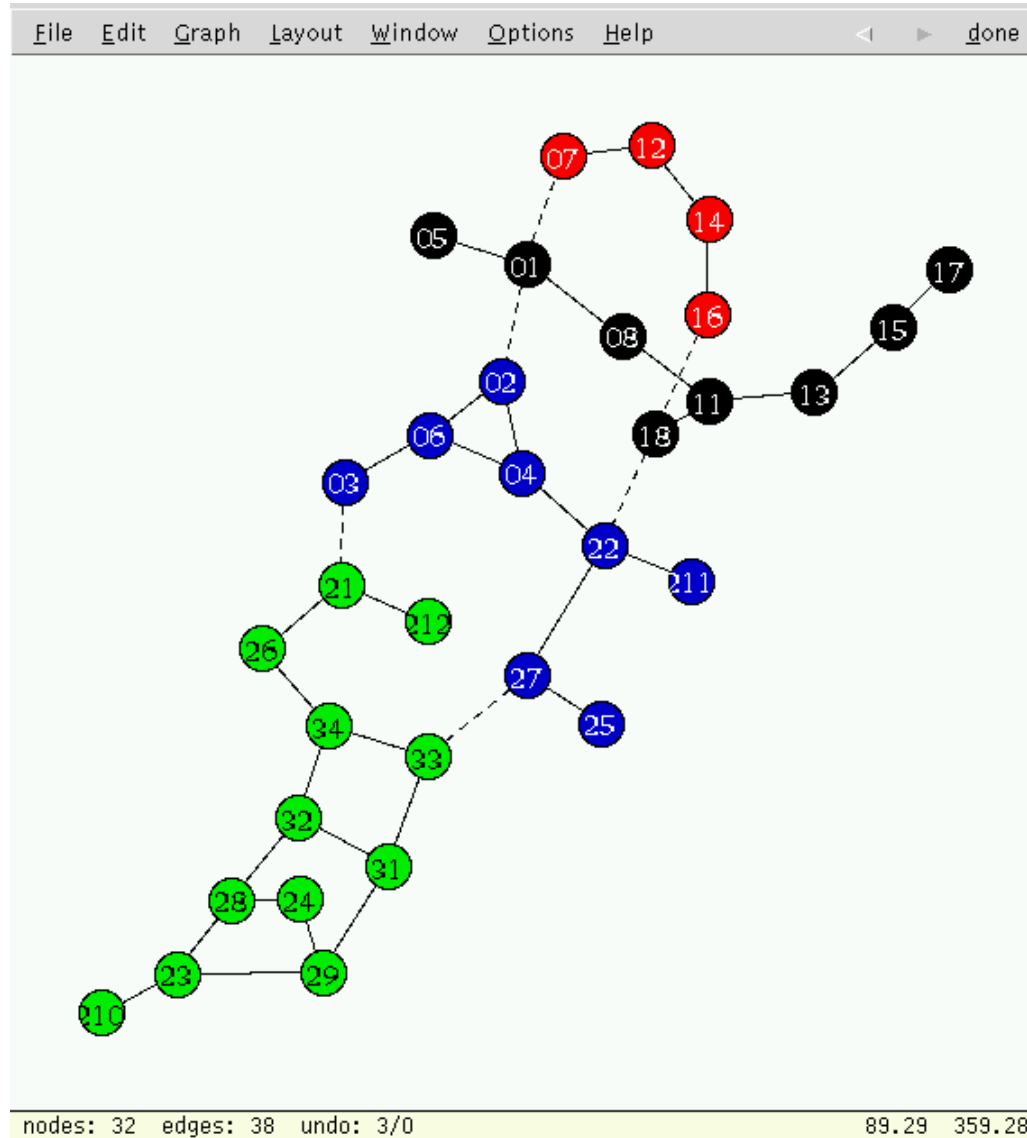
SOs

Reportar

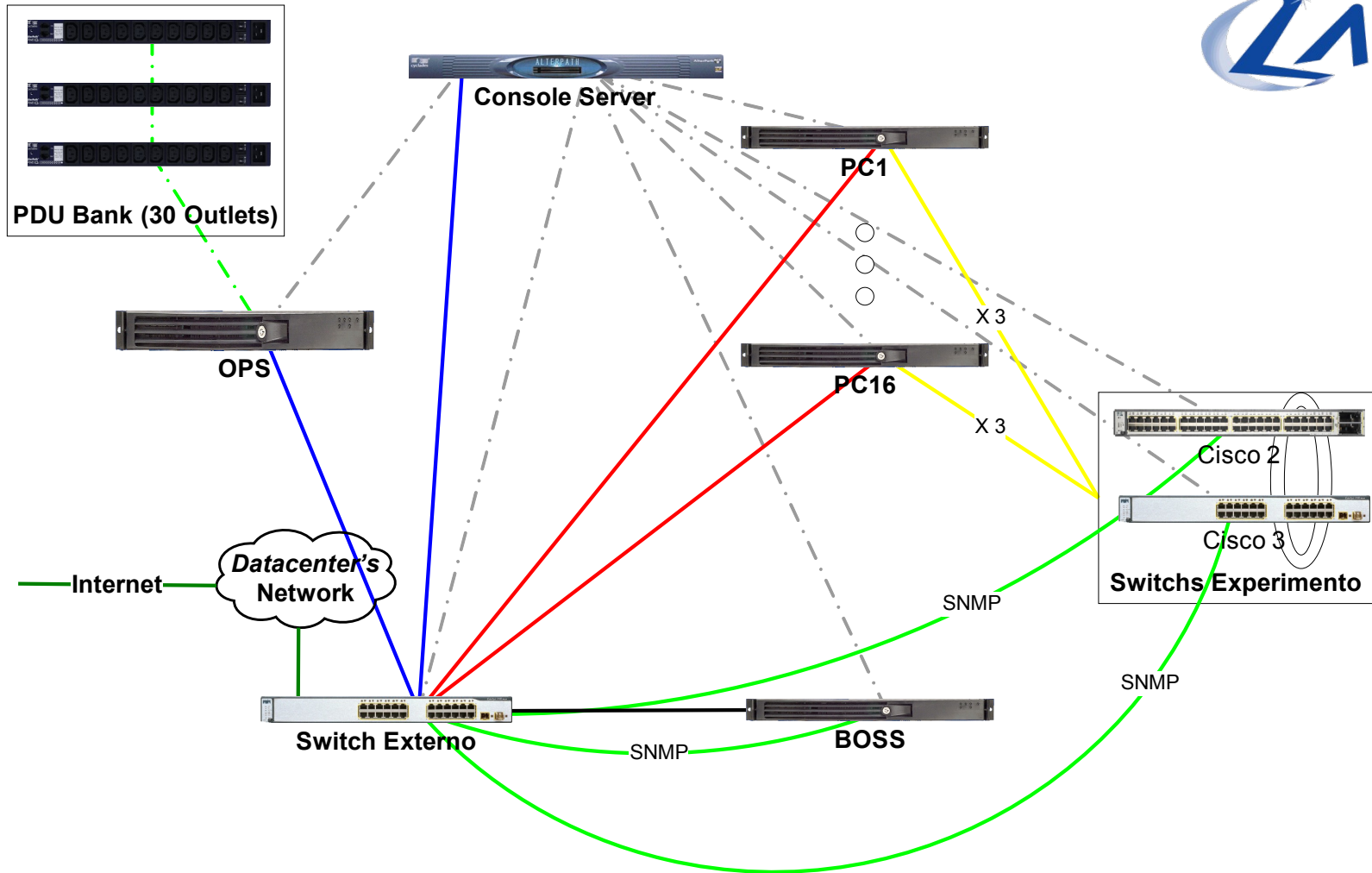
Execução



Exemplo de Experimento



O Ambiente do Emulab no Larc



- Control hardware (vlan4) ·
- External net (vlan2) —
- Control net (vlan5) —
- Public net (vlan3) —
- Experimental net —
- Private net (vlan6) —
- Serial Console —
- Serial PDU Control —

O Ambiente do Emulab no Larc

Disponível em www.emulab.larc.usp.br



Emulab LARC:

- Servidores:

- BOSS: Serviços que suportam o software do Emulab.

- ⊗ Intel Pentium 4 (3GHz), 1GB RAM (DDR2), 580GB HD (Sem RAID – 80GB + 500GB)

- OPS: Servidor de arquivos e ponto de acesso dos usuários.

- ⊗ Intel XEON (3GHz), 1GB RAM (DDR2), 2.08TB HD (80GB Sem RAID + 2TB RAID 5)

- Experimento:

- 16 nós de experimento:

- ⊗ Intel Pentium 4 (3GHz), 1GB RAM (DDR2), 80GB HD (Sem RAID)

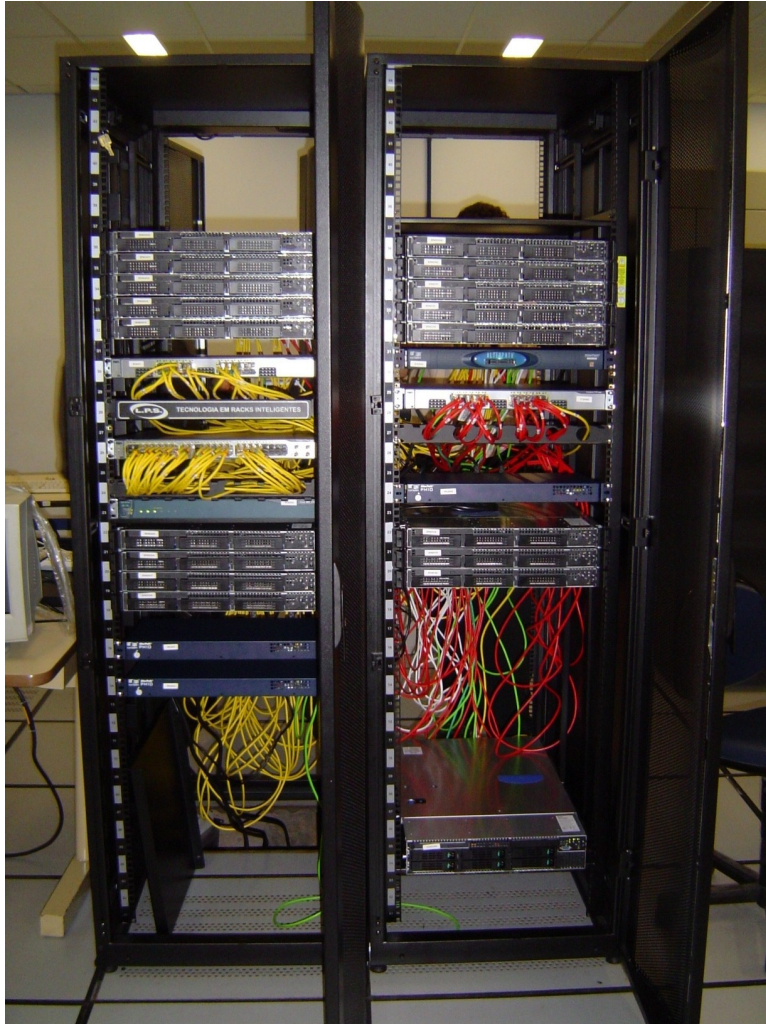
- Suporte:

- Console Server: Avocent ACS 32

- PDU Bank: 3x Power Unit Avocent Alterpath PM10-20A

- Switch: 3x Cisco Catalyst 3750 (2x 24 e 1x48)

Projeto Kyatera



- Instância do Emulab:
 - Realização de Experimentos.



◎ Ambiente Operacional: 08/2007

- Imagens customizadas de sistemas operacionais para os experimentos (p/ monitoração e controle remoto)
 - FreeBSD
 - Linux

◎ Em andamento

- Customização de outras imagens
 - Windows (XP, Vista)
 - Solaris
 - Linux/FreeBSD Appliances (ex.: firewalls, IDS)
- Integração com PlanetLab
 - Atualização de Hardware

Atividades Emulab



- Implantação do software de Emulab para emulação de redes em plataformas física e virtual.
- Utilização do ambiente computacional para realização de *benchmarks* de aplicações distribuídas, considerando diferentes monitores de máquinas virtuais, SOs e tecnologias de SAN (*Storage Area Network*)



PlanetLab Ambiente para Desenvolvimento e Pesquisa de Aplicações

Fernando Frota Redígolo – fernando@larc.usp.br

Márcio Augusto de Lima e Silva – msilva@larc.usp.br

Marcos Tork – mtork@larc.usp.br

Joelle Quaini Sousa – joelle@larc.usp.br

Tereza Cristina Melo de Brito Carvalho – carvalho@larc.usp.br

Técnicas de pesquisa em Sistemas Distribuídos



○ Simulação:

- Implementação de todo o experimento como um descrição funcional para execução por um simulador.
- Ex.: ns-2, Opnet

○ Emulação

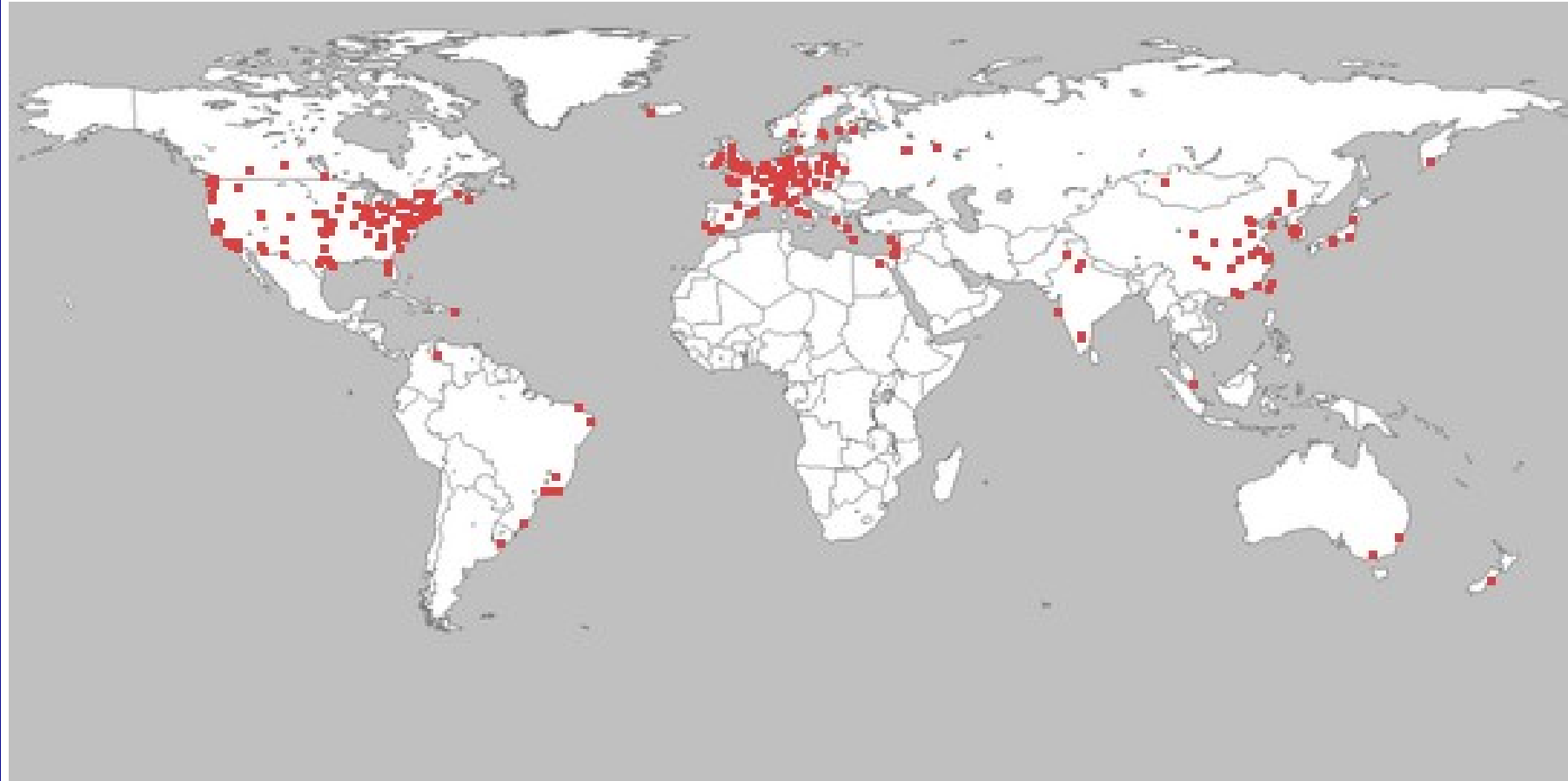
- Uso de nós de borda com aplicações reais, comunicando-se através de uma rede cujos enlaces sejam simulados
- Ex.: Emulab.

○ Sistemas reais

- Execução de todo o experimento em um sistema real equivalente ou próximo do alvo da pesquisa.



- Plataforma para testes de larga escala de aplicações distribuídas em condições reais de uso
 - Nós distribuídos no mundo
 - Aplicações estão sujeitas a condições reais de uso:
 - Disputa por recursos computacionais em nós.
 - Comunicação via Internet, sujeita a atrasos e perdas.
 - Falhas de *hardwares* e enlaces.



- Atualmente 817 nós em 405 sites
- Brasil: LARC-USP, RNP, UFMG, UFCG, IPT-SP

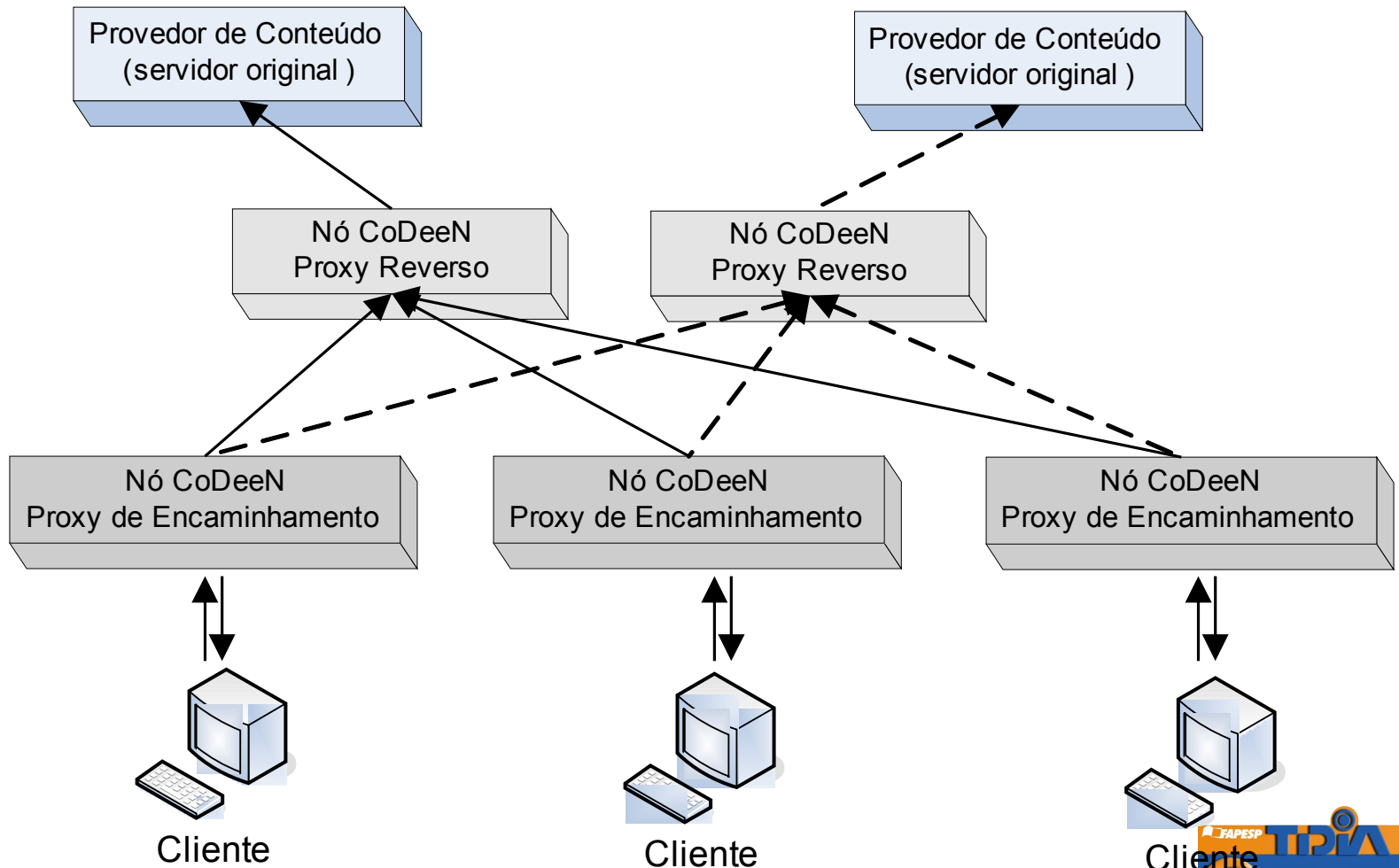


- Plataforma para implementação e disponibilização de novos serviços
 - Content Distribution Networks (CDNs).
 - *Distributed Hash Tables* (DHTs).
 - Armazenamento e *Caching* Distribuído .
 - Mapeamento, Localização e Monitoração distribuída.
 - Roteamento e *Multicast* em nível de Aplicação (exemplo: i3).
 - Aplicações P2P.

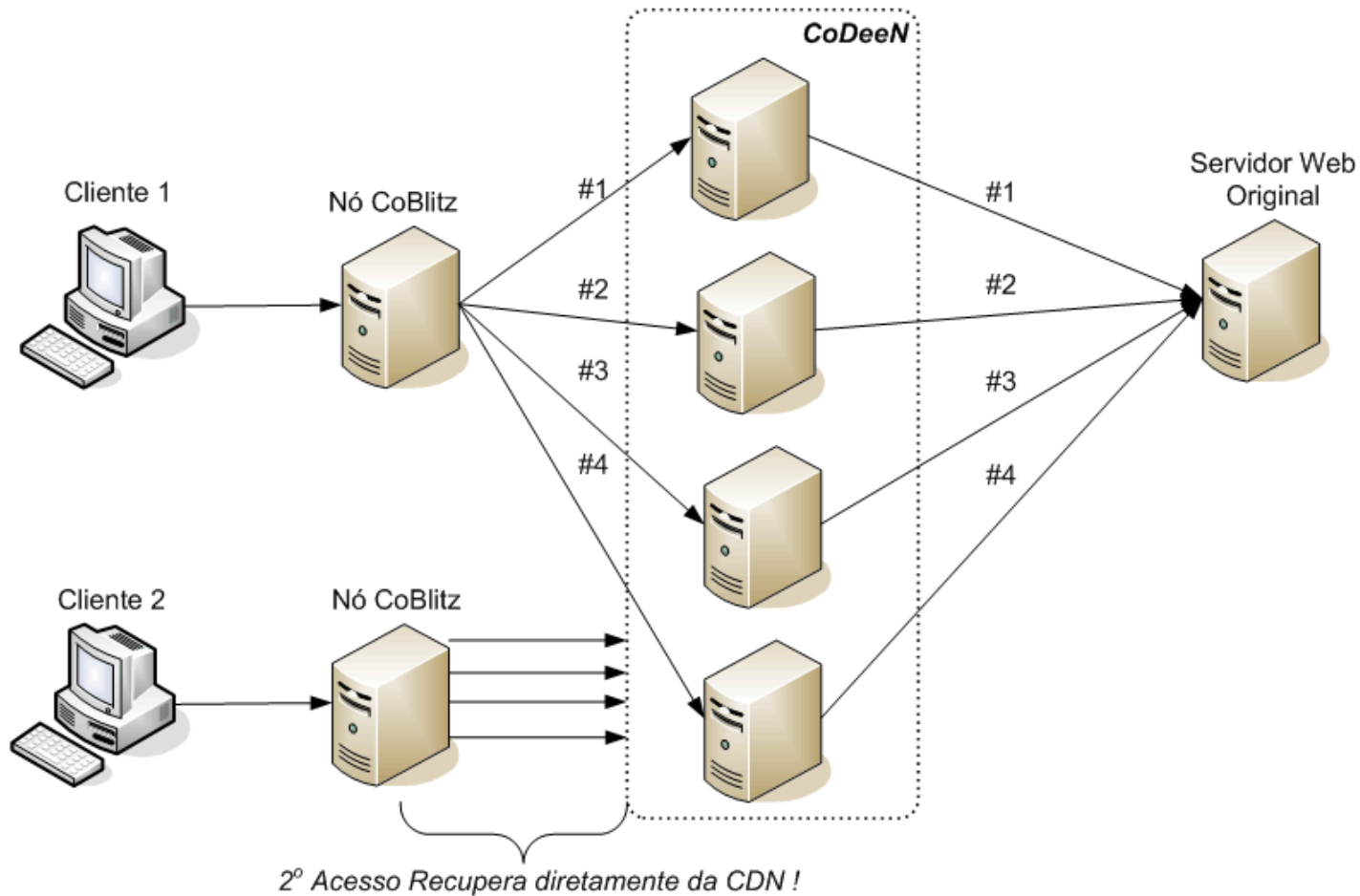
CDNs – Exemplo CoDeeN



Arquitetura CoDeeN (Consulta)



Armazenamento Distribuído – Ex.: CoBlitz



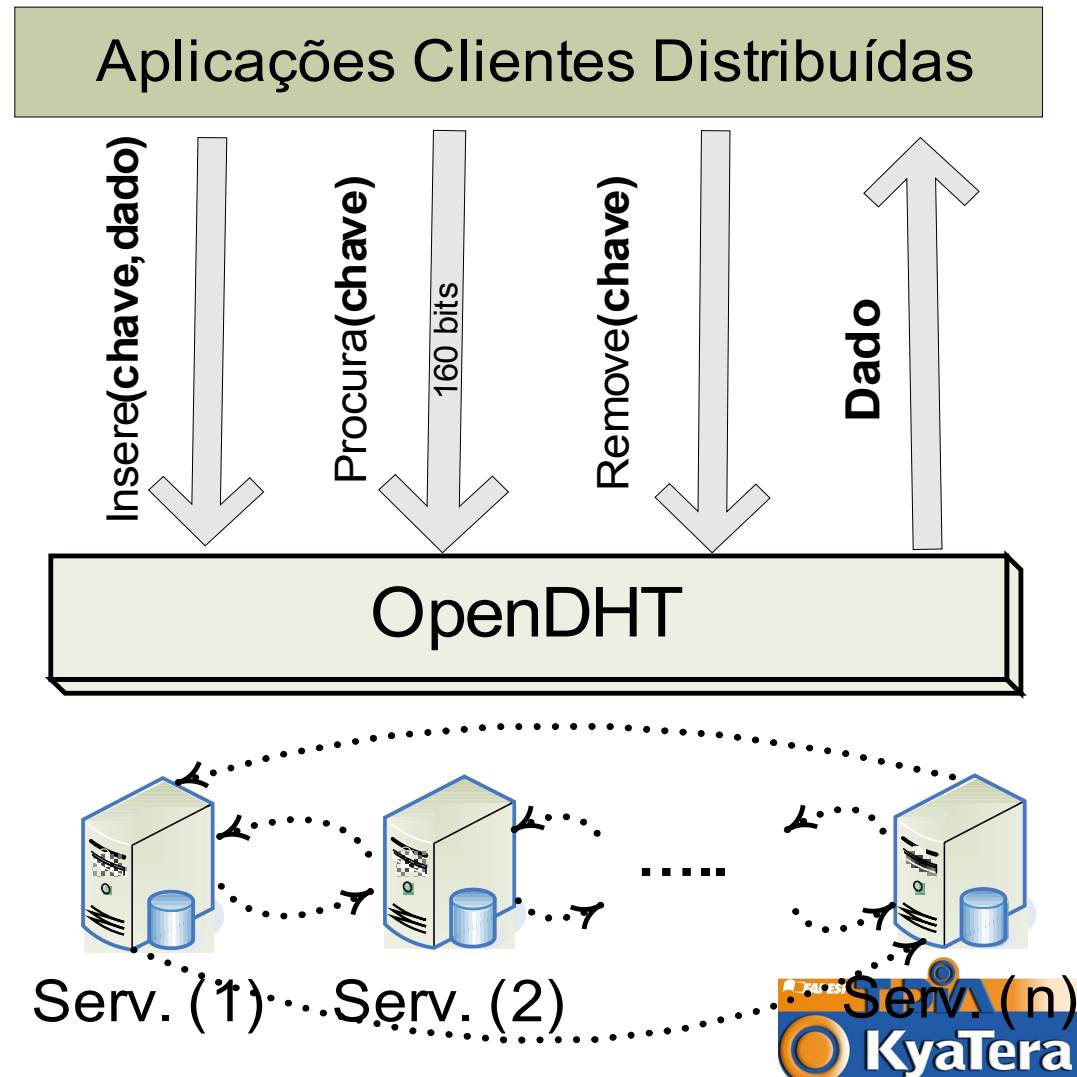
DHT – Exemplo OpenDHT



- Arquitetura OpenDHT

- Lista de servidores:

<http://opendht.org/servers.txt>



Localização e Mapeamento – Ex.: OASIS



Homepage OASIS

- Where am I ?

The Overlay Anycast Service InfraStructure - Mozilla Firefox

File Edit View History Bookmarks Split Tools Help

The Overlay Anycast Service Infr...

"Anycast for Any Service"

OASIS

Overlay Anycast Service InfraStructure

- [Home](#) • Project home page
- [Overview](#) • Brief overview and news
- [Maps](#) • Current deployment
- [Services](#) • Current services
- [Lists](#) • Mailing Lists
- [Pubs](#) • Publications and people
- [Download](#) • Source code

Our Goals

Do you run multiple servers, yet are unable to accurately map clients to the nearest, unloaded instance of your service in an easy manner? Could you benefit by estimating the location of clients accessing your service? If so, OASIS is for you!

Brought to you by the creators of [CoralCDN](#), OASIS is a global anycast service that allows multiple applications to take advantage of its server-selection infrastructure.

As a simple demonstration of OASIS, we use its simple "where am I?" web interface to map your location. (The circle roughly bounds the estimated error of our prediction.)

Map Satellite Hybrid

Brasil

Bolivia

Paraguay

Chile

Argentina

Uruguay

South Atlantic Ocean

POWERED BY Google

Map data ©2007 MapLink/TeleAtlas - Terms of Use

Enter IP / host to locate:

Find: Next Previous Highlight all Match case

Done

Localização e Mapeamento – OASIS

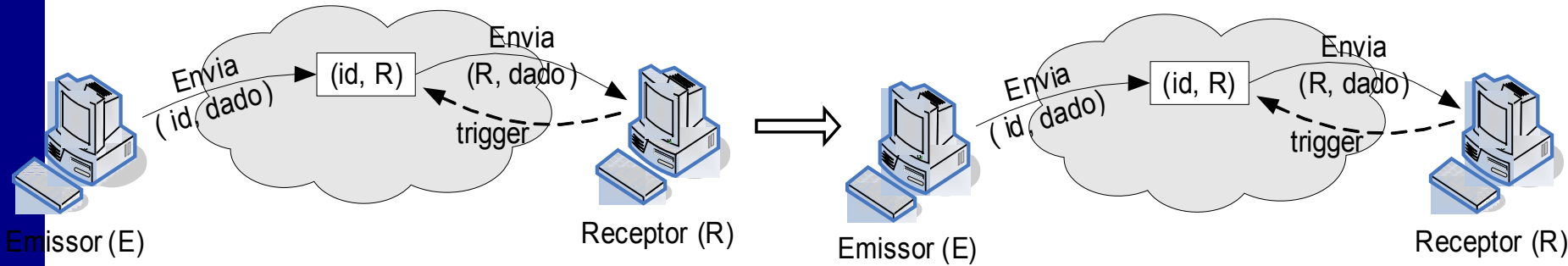


- CoralCDN Nodes configurados no OASIS

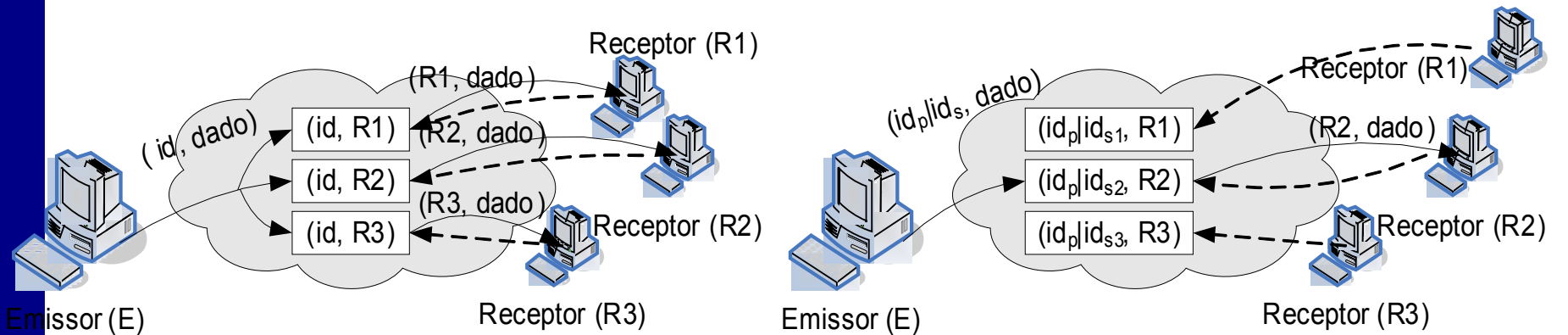
A screenshot of a web browser displaying the OASIS (Overlay Anycast Service Infrastructure) interface. The browser window title is "The Overlay Anycast Service InfraStructure - Mozilla Firefox". The page features the OASIS logo and the text "Overlay Anycast Service InfraStructure". On the right side, there is a navigation menu with links for "Overview", "Maps", "Services", "Lists", "Pubs", and "Download". The main content area is titled "Service deployment" and includes a search box where "coralcdn" has been entered. Below the search box, there are links for "here" and "here". A world map is displayed, showing several red location pins in North America, Europe, and Asia. The map is powered by Google and includes a search bar at the bottom with the text "Find: inp". The browser's status bar shows "Done" and various icons for navigation and utility.

Roteamento - I3

Arquitetura i3 (Comunicação e Mobilidade)



a) Mobilidade



b) Multicast

c) Anycast



- ◎ Ambiente Operacional: 05/2007
- ◎ Curso SBRC 2007 sobre PlanetLab
- ◎ Cooperação com PlanetLab – EUA
 - Workshops sobre overlays - ROADS

- ◎ Próximas etapas
 - Cursos sobre o PlanetLab para Kyatera
 - Integração com Emulab

Considerações Finais



- Arquitetura proposta pelo projeto Kyatera é inovadora em termos de testbeds mundiais:
 - Criação de três redes experimentais com suporte à pesquisa nas camadas física, rede e aplicação.
 - Possibilidade de escalar experimentos com a integração com redes emuladas através do Emulab.
 - Possibilidade de testar novas propostas de arquiteturas de redes – por exemplo, redes overlay, usando PlanetLab.



Perguntas ??