

nic.br egi.br

ceptro.br

Workshop de Medições 2022 da RNP / CT-Mon

São Paulo, SP | 28/10/2022

SIMET e SARA:

Medindo a qualidade do acesso à Internet em provedores e usuários

Henrique de Moraes Holschuh – henrique@nic.br

ceptro.br nic.br egi.br

CEPTRO.br - Medições

Centro de Estudos e Pesquisas em Tecnologias de Redes e Operações (CEPTRO) é responsável por iniciativas e projetos que apoiam ou aperfeiçoam a infraestrutura da Internet no Brasil.

A área de medições de qualidade de Internet cria e opera medidores de qualidade da Internet, realiza e subsidia estudos que possibilitam a melhoria contínua das redes, atua em fóruns multilaterais internacionais para propor mudanças em padrões, e monta cursos de capacitação para as ferramentas de medição, ajudando a fomentar mudanças efetivas na Internet Brasileira.

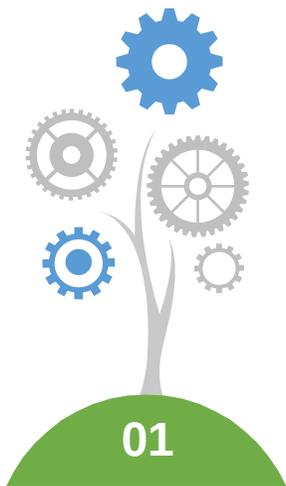
<https://medicoes.nic.br>



Objetivos Principais

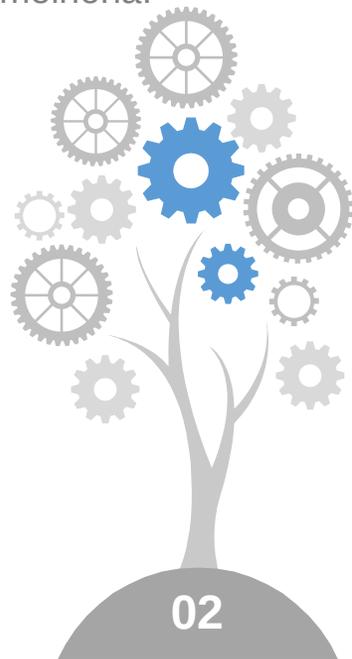
01 Medições

Entender realidades e tecnologias diferentes usando agentes de software e hardware



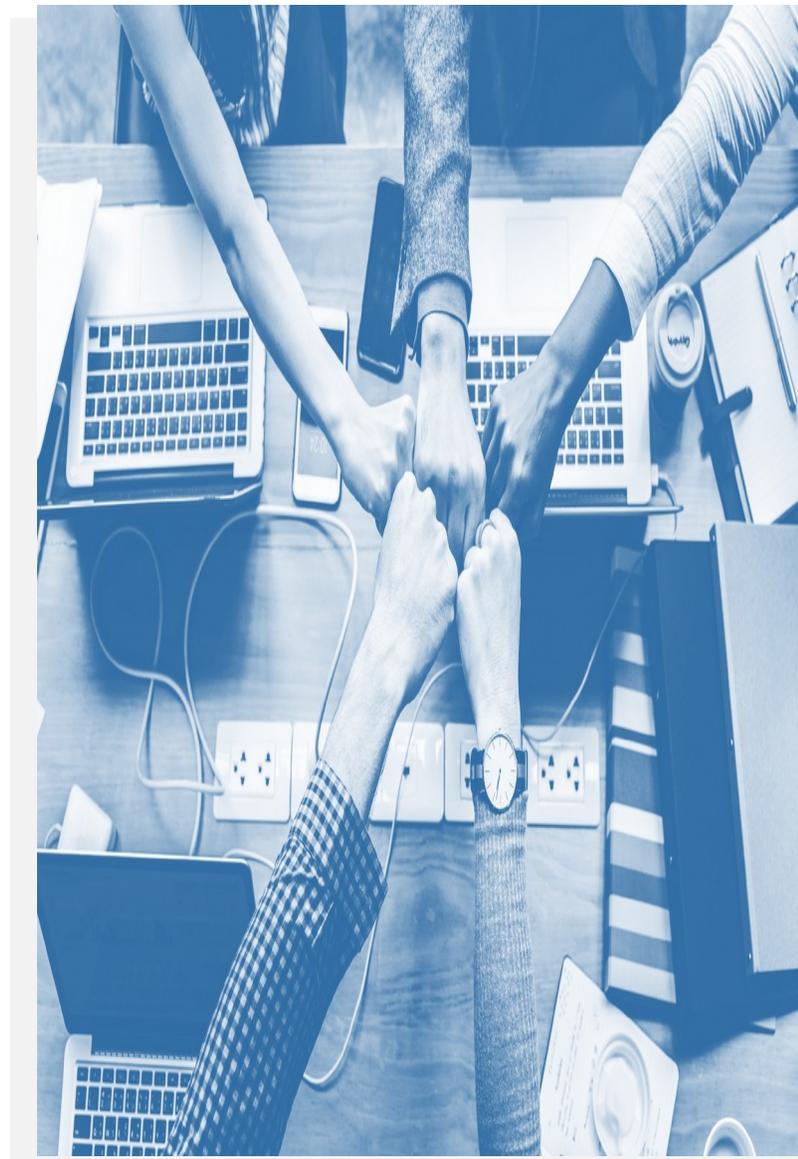
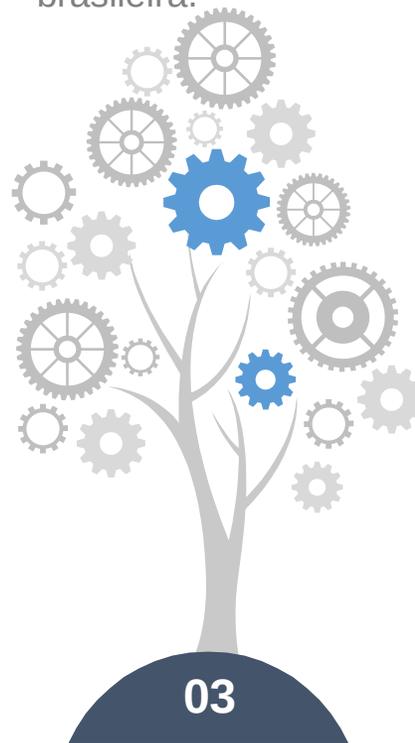
02 Evidências Empíricas

Essas medições servem para gerar correlações com contextos diversos e identificar pontos de melhoria.



03 Melhoria

Fomentar estudos e propor ações que culminem em uma melhoria real na Internet brasileira.



SIMET 2

Sistema de Medição da Qualidade da
Internet Banda larga

- Sistema de Medição de Internet
 - Foco: qualidade do acesso internet banda-larga
- Principais métricas e medições
 - Trem de pacotes: latência (RTT), *jitter*, perda de pacotes, (re-)ordenamento
 - Vazão
- Medições auxiliares
 - Geolocalização
 - DNS, *traceroute*, IPv6, BCP-38...

- Baseado em arquitetura IETF LMAP
- Procura utilizar medições IETF onde possível:
 - TWAMP, no futuro STAMP (e TWAMP light): RTT, *jitter*, perda de pacotes, reordenamento de pacotes
 - RFC 9097 e desdobramentos, no futuro (IP Capacity via UDP)
 - Traceroute TCP e UDP, ping
- Medição de vazão (“IP capacity”) usa um protocolo “proprietário” trivial com 5 streams TCP em paralelo
- Medição web usa técnicas múltiplas, inclusive websockets e WebRTC

- Medidor Web, HTML5 com *WebRTC* e *WebSockets*
 - Usa medições diferentes dos outros medidores
- Motor de medição SIMET-MA (*Linux*) – aceita LMAP schedule
 - SIMETBox / OpenWRT 15+, Debian, Ubuntu (e derivados)
 - 32 ou 64 bits, *little-endian* ou *big-endian*
- Windows (Win7 com absolutamente **todas** as atualizações, Win8+)
- Android (usa o motor de medição SIMET-MA, levemente modificado)
- IOS sem previsão (funcionando com SIMET1)

- Não relevante para o SIMET2-Web, que usa WebRTC e *websockets*
- Latência bidirecional, *jitter*, perda de pacotes unidirecional, reordenamento de pacotes unidirecional
- SIMET2 usa TWAMP em modo simétrico, não-autenticado
- Extensão “proprietária” para incluir um *session-id* no *padding*
 - *Session-id* é necessário para reconhecer fluxo de medição que sofreu NAT com endereços de origem diferentes para TCP e UDP
 - Desnecessário em modo autenticado, será abandonada eventualmente
- TWAMP provavelmente será substituído por STAMP / TWAMP-light para habilitar o uso de refletores em *hardware* ou embarcados em equipamentos de rede.

- SIMET2 Web: *waterfall* de download/upload para servidor HTTP sobre TCP
- SIMET2 outros:
 - Protocolo REST simples para coordenação cliente/servidor
 - 5 *streams* TCP em paralelo, com *session-id* como cabeçalho
- Pipeline descarta *slow-start* e *outliers*
- ***Péssimo para medir long-fat-pipes: buffer TCP do sistema operacional***
- Estamos acompanhando os trabalhos da IETF na parte de IP Capacity em UDP
- Deve ser complementado por uma medição de QUICK/UDP (HTTP/3) no futuro
- Iremos manter alguma forma de medir a vazão TCP enquanto TCP for relevante

- Disponibilidade (via protocolo “proprietário”): sessão “*thin*” TCP com metadados e *keep-alive* bidirecional
 - SIMET-AS utiliza para medir disponibilidade para localidades do IX.br
 - SIMET2 utiliza para medir disponibilidade do trânsito IPv4 e IPv6
- BCP38 – semelhante à medição do “*CAIDA Spoofer*”
 - SIMET-AS apenas, em desenvolvimento final, trânsito e IX.br
- *TOP conexão* (ver <https://top.nic.br>, <https://internet.nl>)
- DNS: Alcance e latência da raiz, e da raiz “.br” (qual nodo *anycast* respondeu, etc), em desenvolvimento final

Mapa no Portal Medidor PIEC

Medições

32 Milhões

Número de medições individuais realizadas

Municípios

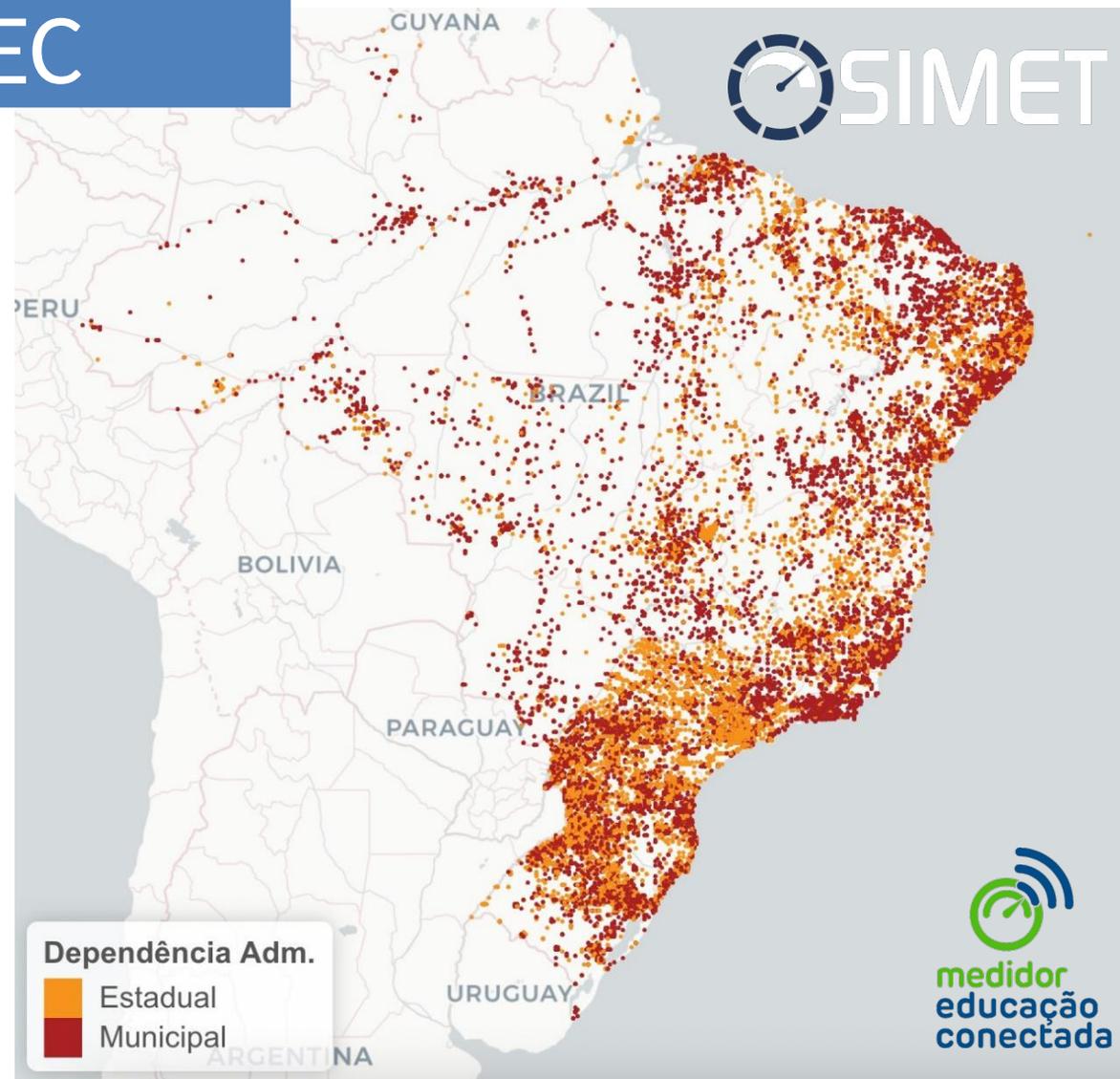
5.055

Número de municípios que possuem ao menos uma escola com o medidor instalado

Escolas

48.904

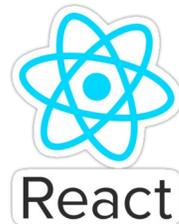
Número de escolas que instalaram o medidor



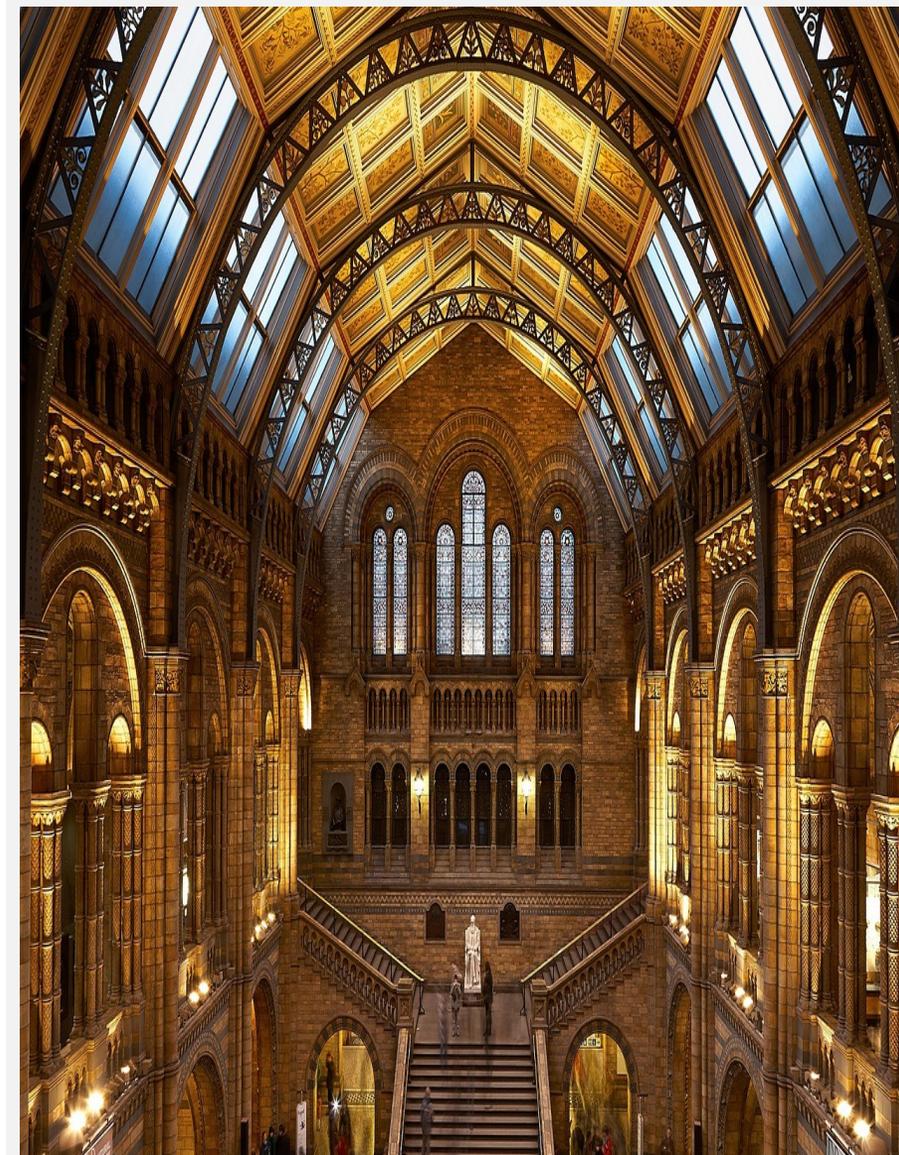
Localização dos Servidores

- Servidores em pontos “neutros” (IX.br quase todas as localidades) – AS14026
- Mas os provedores querem medição **na rede do provedor...** não gostam quando problemas / capacidade de **interconexão** mostra que o serviço prestado pode estar prejudicado...
- Servidores SIMET-ISP (no provedor):
 - Primeira opção dos clientes de medição (segue cone de trânsito BGP)
 - Usuário pode pedir para o cliente usar um servidor neutro do NIC.br (AS14026)
 - Qualquer fração (entre 0 e 100%) das medições pode ser redirecionada pelo NIC.br aos servidores neutros, se necessário
- Medimos a qualidade da interconexão do provedor com os pontos neutros usando o SIMET-AS (mas não há medição volumétrica nesse caso)
- TODAS as medições indicam qual servidor foi utilizado

SIMET2 - Cloudbnative



OpenStreetMap
The Free Wiki World Map



PAS, SIMET-ISP, SIMET-AS

SIMET para Sistemas Autônomos

Motor de medição **SIMET** otimizado para medições entre sistemas autônomos.

Quais as possibilidades?

- Testes qualitativos, **não volumétricos**
- Testes *peer-to-peer* entre os medidores
- Controle centralizado

- Integra com o PAS
- Monitora e calibra o SIMET-MP

- **Métricas:**
 - Latência (RTT), Jitter, Perda de Pacotes
 - Traceroute

- **Features** em implementação
 - Reordenamento de pacotes
 - Teste da filtragem anti-spoofing de origem
 - Alcançabilidade de CDNs nos IX.br
 - TOP.nic.br “conexão”
 - Medições auxiliares de DNS

SIMET-AS - Arquitetura

- Visualização dos resultados através do PAS – Portal do AS
- Medições P2P (*peer-to-peer*) entre os nodos
- Nodo normal:
 - Baixo uso de banda de rede
 - Baixo uso de CPU
 - Sensível à latência (milisegundo)
- Nodo “âncora”
 - Hospedado no AS14026 à princípio (ATM do IX.br)
 - Altíssima capacidade de medição
 - Mesmo *software* que nodo normal
- Firewall restritiva:
 - Nodos comunicam-se apenas entre si, e com NIC.br
 - Firewall do nodo “fecha” completamente se perder comunicação por um período mais prolongado (exceto para NIC.br / gerência)
- Deve ser posicionado na *borda* do AS
- O AS não deve filtrar tráfego do SIMET-AS exceto filtros de borda que ele aplique para **toda** a sua rede (por exemplo: *antispoofing*)

SIMET-MP (servidor de medição SIMET2)

- ✓ SIMETBox (roteador, sensor)
 - ✓ SIMET-MA (Linux nativo, *docker*)
 - ✓ SIMET Lite (*web*)
 - ✓ SIMET Mobile para Android (nova versão, em *beta*)
 - ✓ Portal SIMET para ISPs e SIMET-AS
- ◆ Outros medidores SIMET: em desenvolvimento

Testes dentro do AS, **inclusive de vazão.**

Acesso de medidores ao servidor é limitado às redes do AS e seu cone de clientes de trânsito.



Requisitos:

- ✓ Hypervisor: KVM, Xen, VMware, Hyper-V, Proxmox
- ✓ Ter disponível **para cada VM, são no mínimo 2 VMs (SIMET-AS + SIMET-MP):**
 - ✓ 1 endereço IPv4 e 1 endereço IPv6 **públicos e estáticos** com conectividade com o AS22548
- ✓ Capacidade do servidor físico hospedando a VM do SIMET-MP depende do tamanho do seu provedor (planos dos assinantes, inclusive dos clientes de trânsito)
- ✓ Recursos físicos (aproximados):
 - ✓ 5 vCPUs por interface de 1GbE, 10 GiB RAM (SIMET-AS: 2GiB, SIMET-MP: >= 8GiB)
 - ✓ Se for processador muito antigo, pode precisar mais que isso
- ✓ Escalabilidade horizontal prevista para 2023, mas servidor rejeita novas medições quando está perto de sobrecarregar

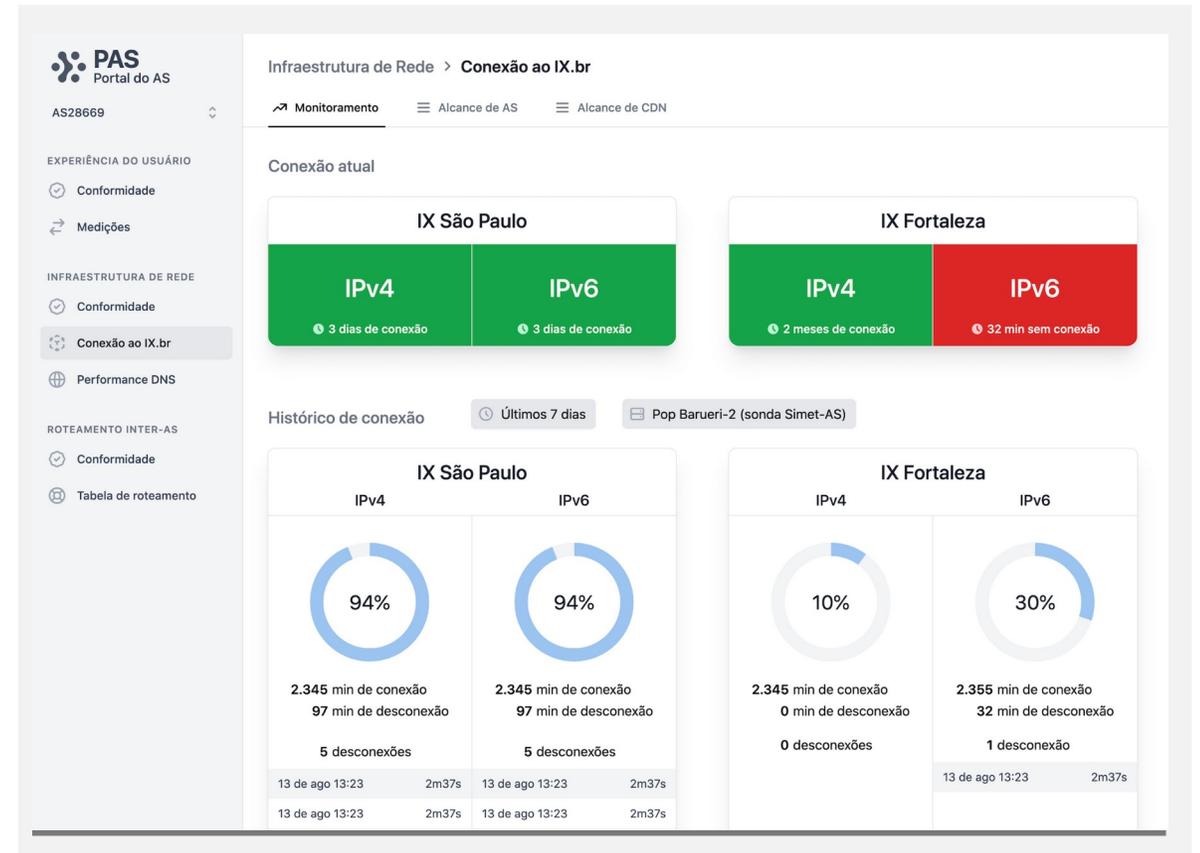


Portal para o AS acessar:

- Medições SIMET-2 realizadas em sua rede
- Medições do SIMET-AS
- Medições do SARA-2
- Anomalias detectadas pelo SIMET-AS e SARA

Futuro:

- Administração das SIMETBox e outros medidores
- Relatórios



SARA v2

Sistema de Análise de Roteamento
entre Sistemas Autônomos

- Objetivos:
 - Coletar as *streams* BGP e tabelas de roteamento
 - Provedores de conexão e conteúdo
 - *Route Servers* do IX.br
 - Manter registro histórico das tabelas de roteamento (“snapshots”) e das *streams* BGP, para análises e estatísticas
 - Realizar análises *offline* e *real-time* e gerar alertas e relatórios
- Semelhante ao RIPE RIS, e ao RouteViews, mas focado na internet Brasileira
- Pode ser alimentado com informações dos coletores RIPE RIS e RouteViews quando necessário (análises).

- Sistema composto por várias camadas
 - Camada coletora: recebe MP-BGP4 (e talvez BMP no futuro)
 - Camada persistidora: *offline* e *online* (*table dumps*, *streams BGP4*)
 - Analisadores e *front-ends*
- As camadas coletora e persistidora já estão em produção, coletando dados de todos os participantes do SARA (inclui os *route-servers* da VLAN multilateral “ATM” do IX.br)
- Disponibilizar as *streams BGP real-time* (estilo RIPE RIS-live) ainda é um assunto não definido. Talvez aconteça, talvez não.

- Coletor
 - Dois servidores Dual-Xeon Silver 4314 com 512GiB RAM e SSDs (RAID1)
 - 3000 sessões MP-BGP4 por servidor.
 - Armazena as *streams* e *table-dumps* em **RAMDISK**
- MultiBird (mesma estratégia dos *route-server* do IX.br/SP):
 - Utiliza BIRD 2.0.9 com *patches* para tornar o MRTDUMP (*stream* BGP) factível, e escalar melhor o *snapshot* das RIBs para arquivos MRT
 - 300 sessões eBGP, 24GiB RAM por instância – coletor não precisa convergir rotas
 - *Nftables* distribui as conexões TCP para as diversas instâncias
 - *Docker* e *docker-compose* para facilitar, mas não é realmente necessário

- Persistidores
 - Rodam no mesmo equipamento que o coletor
 - Transfere os *dumps* da RAMDISK para SSD local, e do SSD local para *storage* CEPH em rede
 - Copia *quasi-real-time* as mensagens das *streams* BGP para um cluster Apache *Kafka*
- *Dataset* em *snapshots* periódicos (das RIB-IN, e das *streams* BGP)
 - Utilizamos MRTv2 **com Extended Timestamps**
 - **Os coletores operam com ADD_PATH (RFC 7911) ativo**

- Consome mensagens BGP do cluster Apache *Kafka* e arquivos MRTv2 da pool de *snapshots* (table dumps)
- Independente das camadas coletora e persistidora: poderia consumir dados de dumps do RouteViews ou RIPE RIS se necessário, ou até mesmo de *feed* RIS-live
- Por enquanto, analisador “para demonstração”, que coleta estatísticas e detecta algumas anomalias e não-conformidades simples
- Área bastante fértil para cooperação com instituições de pesquisa!

- Em planejamento, para o portal público:
 - *Download* da parte pública do *dataset* estilo RIS/RouteViews
 - JSON stream estilo RIS-Live (?) -- Não definido
- Em planejamento, para o PAS:
 - Painel de anomalias e estatísticas
 - Visualização BGPlay
 - *Download* do *dataset* do AS (privado)

- Parte do *dataset* não é público por diversos motivos (inclusive acordos legados da época da primeira versão do SARA), o que causa dificuldades extras em todas as camadas
- BIRD 2: síncrono (processa cada mensagem BGP completamente antes de passar para a próxima), *mono-threaded*, e precisa gravar em disco o mrtdump e os snapshots.

Ao sobrecarregar o BIRD, as sessões BGP caem por falta de envio do *keep-alive*. Uso de RAMDISK e monitoramento do S.O. são obrigatórios.

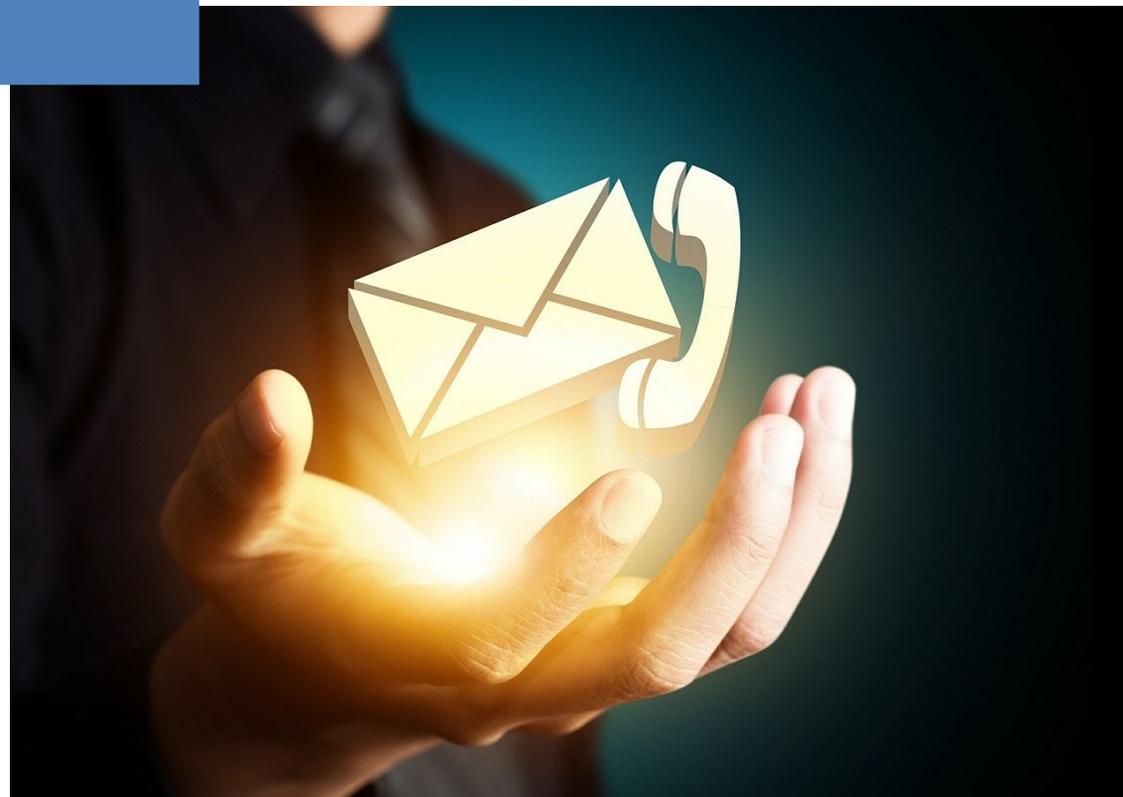
Agradecimento

Estamos abertos a:

- receber feedback e sugestões
- conversar sobre possíveis parcerias
- desenvolver projetos em comum

Novo site:

<https://medicoes.nic.br/>



Contato

medicoes@simet.nic.br

<https://simet.nic.br/projetos>