



Proposta de Piloto

Grupo de Trabalho – Segunda Fase

**GT-ATER: Aceleração do Transporte de Dados
com o Emprego de Redes de Circuitos
Dinâmicos**

Kleber Vieira Cardoso – INF/UFG

05/09/2013

1. Concepção

1.1. Resumo

Na primeira fase, o projeto concluiu o desenvolvimento de um protótipo plenamente funcional. Nesse protótipo, o usuário é capaz de definir regras efetivas ou de monitoramento, as quais permitem encaminhar tráfego através de circuitos dinâmicos ou apenas monitorar o fluxo de dados, respectivamente. Toda a interação do usuário com o protótipo do ATER pode ser realizada através de uma interface *Web* simples que permite, além da criação de regras, remoção de regras e monitoramento de estatísticas. Como administrador, um usuário pode também configurar e gerenciar todo o serviço ATER, realizando atividades como criação de contas, aprovação de regras e verificação de outros componentes do serviço - os RACEs. Além da interface *Web*, o projeto ATER também oferece uma API REST que permite que aplicações utilizem o protótipo de maneira automatizada. Para a segunda fase, a proposta do projeto tem dois objetivos principais: realizar testes com usuários pilotos na rede experimental CIPÓ e evoluir o protótipo para torná-lo mais robusto e escalável. Dentre os parceiros originais do projeto, já há o interesse na realização de testes com sistemas em desenvolvimento e em produção. Além disso, ao longo da primeira fase, foram realizados contatos com outros potenciais usuários que também manifestaram o interesse em utilizar a solução desenvolvida pelo projeto. Para ampliar a robustez do sistema estão planejados métodos para validação de protocolos e testes automatizado de software. A ampliação da escalabilidade está planejada através de uma solução baseada em comutadores (*switches*) para acesso aos circuitos dinâmicos.

1.2. Abstract

The ATER project completed the development of a fully functional prototype in the first phase. In this prototype, the user is able to set effective rules which allows them to forward traffic through dynamic circuits. The user can also set monitoring rules to monitor the flow of data. All user interaction with the ATER prototype can be accomplished through a simple *Web* interface that allows the user, besides creating rules, removing rules and monitoring statistics. As administrator, a user can also configure and manage all ATER services, performing activities such as account creation, rule approval and checking of other service components – e.g. RACEs. In addition to the *Web* interface, the ATER project also offers a REST API that allows applications to use the prototype in an automated manner. The proposed project has two main goals for the second phase: testing with pilot users in the CIPO experimental network and evolve the prototype to make it more robust and scalable. Among the original partners of the project, there is already interest in making tests with systems in development and in production. In addition, during the first phase, contacts were made with other potential users who have also expressed interest in using the solution developed by the project. To enlarge the robustness of the system methods for protocols validation and automated software

testing are being planning. The expansion of scalability is being planned through a solution based on switches for accessing dynamic circuits.

1.3. Descrição do produto/serviço

A proposta do GT-ATER é implantar o sistema desenvolvido como um serviço da RNP, cujo principal objetivo é facilitar o acesso dos usuários à rede experimental de circuitos dinâmicos, conhecida como CIPÓ. Além de uma interface *Web* de fácil uso, o serviço proposto também pode tornar mais simples a interligação entre a rede de um usuário final e a terminação convencional dos circuitos dinâmicos, a qual ocorre geralmente em um ponto de presença da RNP (PoP).

A Figura 1 ilustra como o serviço ATER se acopla à rede experimental CIPÓ. Atualmente, o usuário do serviço experimental CIPÓ (SE-CIPÓ) deve informar quando deseja criar e encerrar os seus circuitos dinâmicos. Além disso, o usuário é responsável por separar o tráfego que deseja enviar através da rede CIPÓ do tráfego que quer enviar pela rede convencional. É importante lembrar que o acesso à Internet continua sendo feito pela rede de produção Ipê, enquanto comunicações entre pares específicos, geralmente dentro da própria RNP, são estabelecidas através da rede experimental CIPÓ. Por exemplo, dois laboratórios de física podem ter interesse em trocar grande volume de informação através de um circuito dinâmico dedicado.

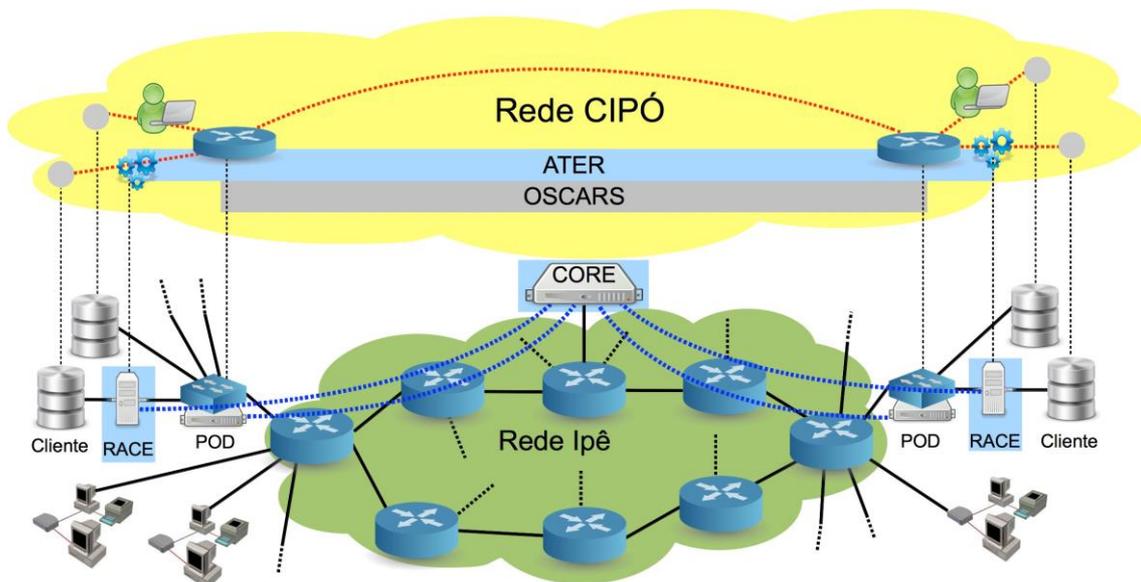


Figura 1: Integração do serviço ATER com o SE-CIPÓ.

No contexto apresentado, o serviço ATER facilita o uso dos circuitos dinâmicos porque não exige que usuário informe quando deseja enviar ou receber o tráfego e nem estimar quanto tempo deve durar a transmissão. O usuário precisa informar o perfil do tráfego enviado através de regras de filtragem conforme ilustrado na Figura 2. Ou seja,

o usuário escolhe qual o perfil do tráfego deseja enviar através de circuitos dinâmicos, fornecendo informações como IP de origem e destino, protocolo de transporte e porta, que caracteriza uma aplicação. Caso o perfil do tráfego não mude, o usuário pode definir a regra uma única vez e utilizá-la, de maneira transparente, quantas vezes desejar. Na prática, isso pode significar que após um acesso inicial e a configuração adequada de suas necessidades, o usuário passará a usufruir do serviço sem nem mesmo lembrar que ele existe.

The screenshot shows the 'GT-ATER' web interface. The top navigation bar includes 'Feedback', 'Sobre', and 'Ajuda'. The left sidebar has 'Regras' selected, with sub-items 'Solicitar Regra', 'Todas Regras', and 'Regras Pendentes'. The main content area is titled 'Nova Regra' and contains the following configuration fields:

- Perfil pré-definido: [Selecione]
- Protocolo de Rede*: IPv4
- Host (selected) / Rede (unselected) radio buttons
- Ip de Origem*: 10 . 1 . 0 . 1
- IP de Destino*: 10 . 2 . 0 . 1
- Configurar Protocolo de Transporte e Portas de Origem/Destino:
- Protocolo de Transporte*: TCP
- Porta de Origem: [Empty]
- Porta de Destino: 2000
- Largura de banda do canal: 1 Gbps
- Tipo da Regra: Efetiva

A green 'Adicionar Regra' button is located at the bottom of the form.

Figura 2: Interface Web para criação de regras.

Cada equipamento RACE (*Rule Applier and Circuit Endpoint*) colocado na borda da rede CIPÓ também se conecta à rede de produção Ipê. Através das regras definidas pelo usuário, o RACE é capaz de separar o tráfego que deve ser enviado através de circuitos dinâmicos do tráfego que deve fluir através da Internet convencional. Isso significa que o usuário não precisa realizar qualquer separação de tráfego em sua rede de origem, basta conectá-la a um RACE. Essa abordagem também facilita a extensão da rede CIPÓ, que geralmente chega até um PoP, até a rede do usuário final, que normalmente está dentro de uma instituição conectada à RNP. Na configuração atual, uma equipe de redes da instituição, coordenada com a equipe de um PoP, precisa identificar quais equipamentos dentro de uma unidade (por exemplo, um laboratório) deverá ter acesso à rede CIPÓ. A equipe da instituição deve então estender o serviço oferecido pela rede experimental CIPÓ, em geral por meio de VLANs, até os equipamentos da unidade e orientar o usuário para enviar apenas tráfego específico através dos circuitos dinâmicos. Caso o usuário tenha demanda por uso da Internet nos mesmos equipamentos que estão sendo utilizados nos circuitos dinâmicos, providências adicionais devem ser tomadas, como acréscimo de interface de rede e/ou configuração adequada dos equipamentos.

Por outro lado, caso o serviço ATER esteja disponível, basta conectar toda a rede da unidade a um RACE, por exemplo, através de VLANs. A partir desse ponto, o usuário

apenas deve acessar a interface *Web*, ilustrada na Figura 2, e definir seu perfil de tráfego. Caso ele não coloque nenhuma regra, todo o tráfego é encaminhado através da Internet convencional, como já era realizado antes da intermediação do RACE. De fato, é possível uma configuração ainda mais simples, conectando toda a rede da instituição diretamente no RACE. Com essa configuração, todo tráfego da instituição que não combinar com alguma regra será simplesmente enviado através da Internet convencional. Adicionalmente, a equipe de rede da instituição pode assumir a responsabilidade de descrição das regras no serviço ATER, consultando o usuário final apenas para identificar sua demanda. Assim, o serviço ATER tem o potencial para facilitar e motivar a adoção do SE-CIPÓ, uma vez que a complexidade na implantação da última milha é sensivelmente reduzida e o envolvimento do usuário final é bastante limitado.

Para muitos usuários, como físicos, biólogos, ecólogos, dentre outros, a disponibilidade de uma infraestrutura de comunicação mais eficiente é muito útil, pois suas atividades fins envolvem comumente transferências de grandes volumes de informação. Atender esse tipo de usuário é a proposta do SE-CIPÓ. O serviço ATER agrega ao CIPÓ a transparência no uso do serviço, tornando a rede CIPÓ praticamente transparente, como é a rede Ipê para esses usuários.

1.4. Identificação do público alvo

O serviço ATER visa atender qualquer usuário que tenha demanda por enviar grandes volumes de dados entre redes interligadas através da RNP. Há usuários com esse perfil em diversas áreas como Física, Meteorologia, Medicina, Computação, dentre outras. Percentualmente, o número desses usuários na rede é pequeno, no entanto, o volume de tráfego que eles movimentam e o impacto do trabalho que eles realizam é muito relevante para academia e para a comunidade em geral.

2. Definição do piloto

2.1. Arquitetura do piloto

A Figura 3 apresenta a proposta de arquitetura para o piloto. Nessa proposta, o intuito é atender usuários de grandes centros de processamento de dados, mais especificamente dois CENAPADs (Centros Nacionais de Processamento de Alto Desempenho), e usuários de sistemas de nuvem para computação científica, através do sistema que vem sendo desenvolvido no GT-MCC.

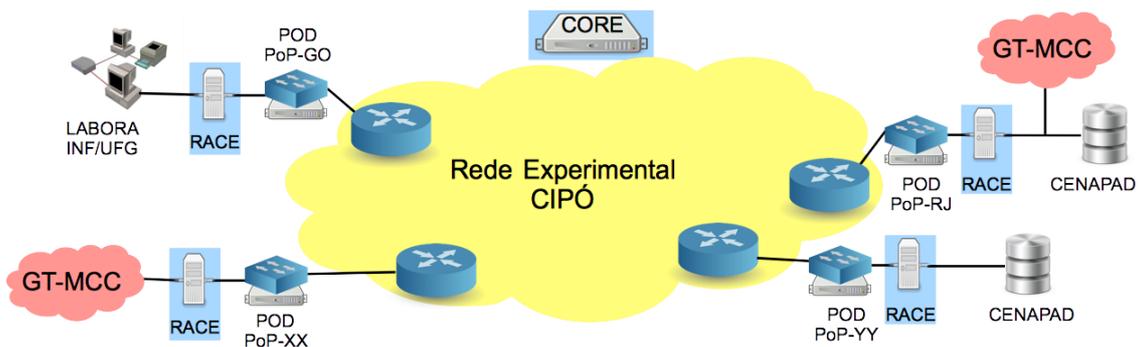


Figura 3: Arquitetura proposta para o piloto do ATER.

A escolha dos CENAPADs e do GT-MCC se baseiam na adequação da demanda desses usuários e da parceria do LNCC no desenvolvimento do serviço ATER. Nos CENAPADs, há demanda por transferências de grande volumes de dados para atender operações como *backup* ou sincronização de *storages*. As transferências são entre equipamentos que possuem restrições quanto a customizações ou modificações, mas que operam através de endereços IP e portas TCP bem conhecidas. Portanto, esses equipamentos representam uma demanda adequada para o serviço ATER. O sistema que vem sendo desenvolvido pelo GT-MCC atende diferentes usuários de computação científica, inclusive aqueles que precisam mover grandes volumes de dados para serem processados ou após o processamento. No entanto, diferentemente da demanda dos CENAPADs, o sistema do GT-MCC precisa de uma interface programática, a qual permita o estabelecimento de regras de encaminhamento de tráfego através de circuitos dinâmicos. Nesse contexto, o usuário final nem sequer percebe que está utilizando o serviço ATER, uma vez que a criação da regra é realizada através do sistema do GT-MCC. Assim, a API REST oferecida pelo serviço ATER também pode ser avaliada dentro do piloto.

Os PoPs GO e RJ foram previamente escolhidos por serem os mais próximos às sedes das equipes envolvidas no projeto e seu envolvimento é fundamental para o desenvolvimento do sistema, testes e avaliações do serviço ATER. Além disso, ainda na primeira fase do projeto, já está sendo preparado um pré-piloto envolvendo esses dois PoPs, os quais receberão, ainda nessa fase, os RACEs. A definição dos demais PoPs é uma das primeiras atividades a serem realizadas na segunda fase do projeto e levará em conta fatores como disponibilidade da rede experimental CIPÓ na localidade e infraestrutura de rede para acesso ao PoP. O local onde o CORE (*Circuit Operation and Rule Establishment*) será implantado também será definido, em conjunto com a RNP, no início da segunda fase. O Instituto de Informática da UFG, onde o projeto vem sendo desenvolvido, pode manter o piloto do serviço com boa qualidade. Por outro lado, a RNP conta com o *Internet Data Center* (IDC) que oferece alta disponibilidade.

Caso o piloto descrito previamente alcance os resultados esperados antes da conclusão da segunda fase deste projeto, foi planejada também uma ampliação desse piloto. Há dois candidatos para participação nessa ampliação do piloto, o LInEA (Laboratório Interinstitucional de e-Astronomia) e o LAVID (Laboratório de Aplicações

de Vídeo Digital). Ambos foram contatados durante a primeira fase do projeto GTATER e manifestaram interesse em utilizar o serviço proposto. No entanto, para atender o LAVID, cuja demanda é enviar vídeo 4K através de circuitos dinâmicos, o serviço ATER e a própria rede experimental CIPÓ precisam aumentar suas capacidades de transmissão, idealmente para 10 Gbps. Nessa segunda fase, a evolução já está prevista no serviço ATER em termos de software, porém o projeto depende da aquisição de comutadores com essa velocidade, assim como a disponibilidade de enlaces com essa taxa de transmissão fim a fim na rede CIPÓ.

Na Figura 4, são mostrados os principais módulos de software do ATER que serão implantados no piloto. Embora não esteja explicitamente ilustrado na figura, todo software desenvolvido é executado sobre a plataforma GNU Linux. A cor verde na figura identifica os componentes desenvolvidos dentro do projeto. Todos os componentes já foram implementados, exceto a aplicação OpenFlow do RACE_OF, ou seja, o RACE baseado em tecnologia OpenFlow. Conforme será descrito em mais detalhes na Seção 2.3 -- Refinamento do protótipo, o RACE_OF é uma evolução do RACE_PC, baseado em PC (*Personal Computer*). Portanto, a implantação do piloto pode ter início com os RACE do tipo PC que foram implementados na primeira fase do projeto.

A cor azul usada na Figura 4 representa o módulo que foi desenvolvido a partir de uma customização de software de terceiros. O principal componente com essa característica é o CORE_OSCARS_Driver (COD) que foi implementado com base em ferramentas clientes distribuídas em conjunto com o OSCARS e também é conceitualmente similar ao software OSCARSBrigde, utilizado pelo MEICAN.

Por fim, a cor amarela usada na Figura 4 está associada aos módulos que são utilizados praticamente sem alteração. Esses componentes são o banco de dados, o OSCARS que representa o componente tecnológico para interação com o SE-CIPÓ e o controlador OpenFlow sobre o qual será desenvolvido uma aplicação para configuração de regras.

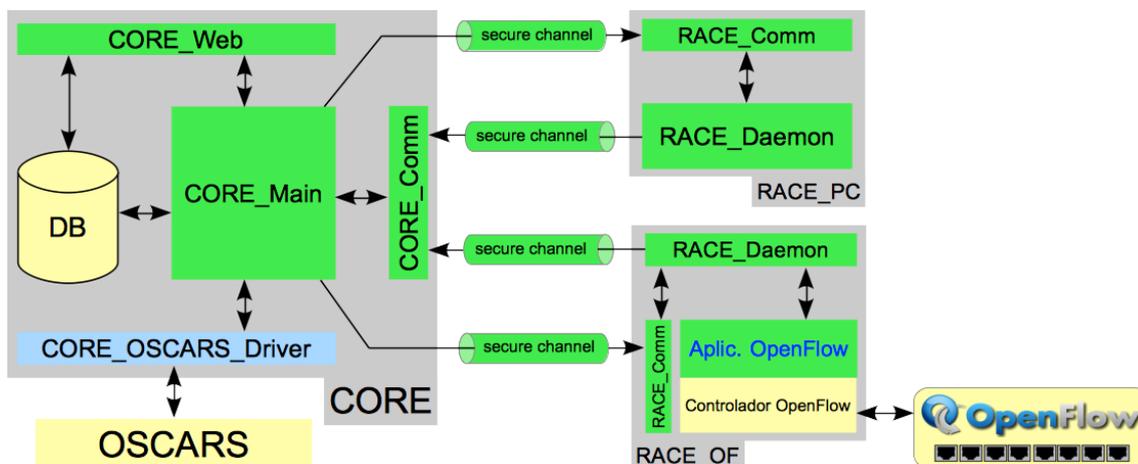


Figura 4: Módulos de software do ATER.

O serviço ATER não exige a instalação de nenhum equipamento nas instituições participantes do piloto. É suficiente que o PoP de acesso da instituição seja coberto pela rede experimental CIPÓ, ou seja, que possua um POD, e que a terminação dos circuitos dinâmicos seja estendida até a rede do usuário final. Em cada PoP participante do piloto é então instalado um RACE. A exigência de estender os circuitos dinâmicos até o cliente existe para o SE-CIPÓ e o ATER apenas a herda. Conforme comentado previamente, o serviço ATER permite que essa exigência seja flexibilizada, simplesmente conectando toda a instituição onde se encontra o usuário final diretamente no RACE que está no PoP. No entanto, na fase piloto, é compreensível que essa abordagem seja preterida, uma vez que o serviço ATER ainda se encontrará em fase pré-experimental.

2.2. Instituições participantes

Para a implantação do piloto principal, o projeto ATER conta com o apoio do Laboratório Nacional de Computação Científica (LNCC) e dos CENAPADs, sob coordenação do SINAPAD (Sistema Nacional de Processamento de Alto Desempenho). O contato é Antônio Tadeu Azevedo Gomes, pesquisador do LNCC e secretário executivo do SINAPAD.

Para a implantação de uma eventual ampliação do piloto, a equipe do projeto ATER estabeleceu contato com o LIneA (Laboratório Interinstitucional de e-Astronomia) através do pesquisador Luiz Nicolaci da Costa. Adicionalmente, a equipe do projeto ATER também estabeleceu diálogo com o LAVID (Laboratório de Aplicações de Vídeo Digital) por meio do pesquisador Lucenildo Lins de Aquino Júnior.

2.3. Refinamento do protótipo

Para que o protótipo desenvolvido na primeira fase do GT atenda adequadamente às necessidades de seus usuários, a equipe do projeto identificou duas questões importantes: 1) aplicação de uma metodologia ampla e ferramentas de teste de software e 2) evolução do RACE para uma arquitetura baseada em OpenFlow.

Na primeira fase, o software foi extensivamente testado, com base em uma metodologia básica e sem o uso de ferramenta específica. Para a segunda fase, é importante ampliar a metodologia e pode ser útil empregar alguma ferramenta para automatização de testes. O intuito é identificar não apenas falhas de programação, mas questões relacionadas a desempenho, manutenção, configuração e outras propriedades relacionadas a sistemas distribuídos. Em especial, o teste de software deve avaliar o protocolo principal desenvolvido para o serviço ATER e sua robustez sob condições variadas de operação e demanda. A aplicação de uma metodologia ampla e ferramenta específica para teste de software tem como principal objetivo tornar o serviço ATER altamente confiável para seus usuários, o tornando pronto para uma eventual adoção em ambiente de produção.

O RACE atual foi projetado e avaliado até uma demanda de 1 Gbps por interface de rede. Como o RACE funciona em modo ponte (*bridge*), interligando redes dos usuários com a rede Ipê ou com a rede CIPÓ, a taxa testada foi efetivamente de 2 Gbps. No entanto, a demanda de múltiplos usuários e o aumento da capacidade dos enlaces de acesso até os PoPs pode superar rapidamente essa oferta de capacidade. Adicionalmente, o uso de equipamento baseado na arquitetura PC tem sido uma crítica recorrente ao projeto. É importante ressaltar que a equipe do projeto não observou nenhum evento de indisponibilidade ou instabilidade nos diversos testes em laboratório realizados ao longo de meses. Por outro lado, a equipe do GT avalia que a solução baseada em arquitetura PC apresenta grandes dificuldades de escala do número de interfaces e da taxa de transmissão das mesmas. Ou seja, utilizar um grande número de interfaces de rede ou interfaces de 10 Gbps exigirão customizações no software desenvolvido e no sistema operacional que demandam longo tempo até se tornarem satisfatórias e estáveis. Além disso, o custo do hardware, baseado em arquitetura PC, para atender a essa demanda tende a ser mais elevado que uma solução baseada em comutador.

O comutador possui diversas características úteis para o contexto de aplicação descrito previamente: grande número de portas, facilidade de expansão para velocidade de 10 Gbps ou superior, arquitetura projetada para encaminhamento eficiente de pacotes e, em geral, alta robustez. A principal dificuldade do comutador é possuir uma interface de programação que seja aberta e flexível para atender às necessidades do projeto. Nesse contexto, a tecnologia OpenFlow se mostra adequada, permitindo a criação de regras de encaminhamento e modificação dos pacotes de maneira simples e eficiente. Além disso, OpenFlow oferece um protocolo aberto que já é implementado por alguns fabricantes tradicionais da indústria de equipamentos de rede. Assim, é possível encontrar flexibilidade para aquisição, substituição e evolução dos RACEs similar à encontrada na arquitetura PC.

Conforme ilustrado pela Figura 4, o serviço ATER prevê a coexistência dos dois tipos de RACE sem qualquer alteração significativa no software desenvolvido. As alterações são realizadas apenas nos próprios equipamentos que estão nos PODs. É possível converter um RACE baseado em PC em um RACE baseado em OpenFlow ou simplesmente utilizar um outro equipamento disponível para essa tarefa, por exemplo, o servidor que faz parte do kit do POD. Na abordagem OpenFlow, todas as ações que exigem alto desempenho são realizadas no comutador e, portanto, não há exigências específicas de hardware para executar o código do RACE, incluindo a aplicação e o controlador OpenFlow. Por fim, vale ressaltar que a equipe proponente do projeto já possui experiência com o desenvolvimento de aplicações OpenFlow e já utiliza essa tecnologia, em ambiente experimental, há mais de um ano.

2.4. Ferramentas de suporte à operação (para propostas de serviço)

Na versão do software desenvolvido na primeira fase do projeto, o ATER já conta com diversos recursos que auxiliam em sua operação. Dentre entre recursos, estão módulos para:

- criação de usuários com diferentes perfis,
- configuração da maior parte do serviço,
- estatísticas de uso das regras e dos circuitos dinâmicos apresentadas através de gráficos e
- monitoramento centralizado dos RACEs no CORE.

Na segunda fase, a equipe do projeto pretende identificar, em conjunto com a RNP, por quais novos recursos há efetivamente demanda. Previamente, a equipe já identificou alguns, comentados a seguir, mas a definição de quais implementar depende de interação com a RNP e com os próprios usuários. Se for identificada a demanda de um grande número de usuários pelo serviço ATER, pode ser útil integrar o sistema de autenticação e autorização do ATER à federação CAFe. Caso haja interesse da equipe de operação da RNP em monitorar alguma característica específica do serviço ATER em uma ferramenta de monitoramento externa, por exemplo o Nagios, haverá a necessidade de escrever um agente customizado (ex.: *plugin*). Outra demanda possível é a disponibilização das estatísticas coletadas pelo ATER para o serviço de monitoramento baseado em perfSONAR, o qual é utilizado na RNP. Como as estatísticas do ATER já são persistidas em banco de dados, basta exportá-las através de uma interface *Web Service* seguindo o esquema NMWG (*Network Measurement Working Group*).

3. Cronograma

O cronograma descrito a seguir é baseado no cronograma resumo das entregas descrito na chamada de propostas para a segunda fase. Além das atividades previstas na chamada de propostas, foram acrescentadas as seguintes atividades:

- **Imp_CORE**: implantação do CORE.
- **Imp_RACEs_PC**: implantação dos RACEs do tipo PC.
- **Melhorias**: correção de falhas, ajustes de desempenho e pequenas customizações.
- **Des_RACE_OF**: customização do código do RACE para interagir com aplicação OpenFlow e desenvolvimento da aplicação OpenFlow.
- **Imp_RACEs_OF**: implantação dos RACEs do tipo OpenFlow.
- **Avaliação**: identificação das demandas efetivas do serviço.
- **Ajustes**: com base nas demandas efetivas, realização de ajustes no serviço.

	2013		2014											
	Nov	Dez	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
RP4	■													
RP5		■												
RA5			■											
Site do GT			■											
RT4			■											
RT5				■										
Imp_CORE	■	■												
Imp_RACEs_PC			■	■										
Início piloto					■									
Código-fonte					■									
Documentação					■									
Testes do piloto					■	■	■	■	■	■	■	■	■	■
Melhorias	■	■	■	■	■	■	■	■	■	■	■	■	■	

Des_RACE_OF														
Imp_RACEs_OF														
Avaliação														
Ajustes														
RP6														
RA6														
WRNP														
RWRNP														
RP7														
RA7														
RT6														
Workshop														
RT7														
Apresentação														
RA8														
RP8														
RP9														

4. Recursos financeiros

4.1. Equipamentos e softwares

Descrição	Quantidade
Computador com 3 interfaces de rede	2
Comutador (<i>switch</i>) OpenFlow com suporte completo ao padrão 1.1 ou superior, versão 1.3 é a recomendada	2

Adicionalmente, solicitamos as passagens e diárias dentro do limite previsto.

