

Just in Time Clouds: Uma Abordagem Baseada em Recursos Terceirizados para a Ampliação da Elasticidade de Provedores de Computação na Nuvem

Rostand Costa (Autor), Francisco Brasileiro (Orientador)

¹ Programa de Pós-Graduação em Ciências da Computação (PPGCC)
Universidade Federal de Campina Grande (UFCG)
{rostand.costa, fubica}@lsd.ufcg.edu.br

1. Contextualização

Computação na nuvem (do inglês *cloud computing*) é um paradigma em evolução que permite o fornecimento de Tecnologia da Informação (TI) como um serviço que pode ser adquirido interativamente, *on line* e sob demanda pelos clientes. Os recursos utilizados para prover serviço aos clientes podem ser rapidamente provisionados e liberados pelos provedores do serviço. Quando o serviço é cobrado dos clientes, os provedores utilizam um modelo de tarifação onde o cliente paga apenas pelo que foi efetivamente consumido. Este paradigma pode ser usado em diferentes níveis da pilha de TI [Stanoevska-Slabeva and Wozniak 2010]. Por exemplo, no nível mais alto, clientes podem adquirir serviços que provêm uma funcionalidade particular de *software*. Este tipo de fornecimento de TI é normalmente chamado de SaaS (do inglês, *software-as-a-service*) [Stanoevska-Slabeva and Wozniak 2010]. Por outro lado, no nível mais baixo da pilha, clientes podem adquirir máquinas virtuais totalmente funcionais executando um determinado sistema operacional, sobre o qual eles podem instalar e executar as suas próprias aplicações. Este tipo de serviço recebeu o nome de IaaS (do inglês, *infrastructure-as-a-service*) [Stanoevska-Slabeva and Wozniak 2010] e é nele que este trabalho está focado.

Ao adquirir recursos de TI de um provedor de computação na nuvem, os clientes podem desfrutar da elasticidade oferecida, podendo aumentar e diminuir o seu consumo de serviços de uma forma virtualmente ilimitada, sem qualquer custo adicional. Em teoria, essa elasticidade ilimitada permitiria aos usuários decidirem livremente, por exemplo, se desejam usar 1 recurso por 1.000 horas ou 1.000 recursos por 1 hora, pagando o mesmo preço em ambos os casos. Essa propriedade singular de computação na nuvem é chamada de *associatividade de custos* (*cost associativity*) [Fox 2011].

Ao traduzir infraestrutura de TI em serviços elásticos e ilimitados, utilizados sob demanda e pagos de acordo com a quantidade de serviço consumida, o paradigma de computação na nuvem oferece inúmeras possibilidades novas para o planejamento de capacidade das instituições que utilizam TI de forma intensiva. Em particular, a capacidade de instanciar concomitantemente um grande número de recursos por um período de tempo relativamente curto é um requisito fundamental para um modelo de programação de aplicações paralelas cada vez mais popular, chamado computação de alta vazão (HTC, do inglês *High-Throughput Computing*) [Litzkow et al. 1988]. Essas aplicações têm cargas de trabalho altamente paralelizáveis e quanto mais cedo a sua execução possa ser concluída, melhor. Assim, idealmente, elas poderiam ser executadas simultaneamente pela totalidade dos recursos necessários para terminar o mais rapidamente possível e, ainda, com um custo que só dependeria da carga de trabalho que tiver sido realmente processada. Desta forma, muitas aplicações HTC, científicas ou comerciais, poderiam potencialmente obter um enorme benefício a partir da elasticidade dos fornecedores de computação na nuvem.

Infelizmente, os provedores públicos atuais de IaaS precisam limitar o número máximo de máquinas virtuais que podem ser adquiridas simultaneamente por um dado cliente. Por exemplo, durante todo o tempo de desenvolvimento desta pesquisa, o serviço EC2 (*Elastic Compute Cloud*) da *Amazon Web Services* (AWS), um dos principais provedores comerciais em atividade, limitava em 20 o número de máquinas virtuais que podiam ser instanciadas de forma dedicada (*on-demand instances*) e em 100 o número de máquinas virtuais que podiam ser instanciadas segundo um modelo “*best-effort*” (*spot instances*) [Amazon 2011]. Para este provedor em particular, clientes podem usar um canal paralelo de negociação para tentar aumentar este limite de forma *ad hoc*, mas como as condições sob as quais uma negociação é bem sucedida não são documentadas, nós consideramos neste trabalho apenas o canal de comunicação automático.

Embora os limites atualmente impostos pelos provedores de IaaS não impeçam que a maioria dos clientes enxerguem o serviço provido como uma fonte infinita de recursos, este não é o caso para a maioria das aplicações HTC. Estas aplicações podem requerer a instanciação de um sistema com milhares de máquinas virtuais. Além disso, quanto mais máquinas elas puderem usar, mais curto será o tempo de utilização das mesmas. O projeto Belle II Monte Carlo [Sevior et al. 2010], por exemplo, requer de 20.000 a 120.000 máquinas virtuais para o processamento, em tempo aceitável, dos dados produzidos em três meses de experimentos. Ou seja, eles têm uma altíssima demanda por recursos de forma bastante esporádica. Esse padrão de consumo é muito comum entre os usuários que executam aplicações HTC e, possivelmente, também para outras classes de aplicações.

Como já existem serviços de alta demanda hospedados em provedores de IaaS públicos e privados (ex. Gmail, Twitter, Bing etc.) e também a possibilidade de se negociar alocações superiores com provedores públicos, é possível inferir que o limite serve como um regulador do uso intensivo de recursos por períodos curtos, ou seja, o alvo do limite não é o volume da requisição em si, mas o exercício extremo da elasticidade através de grandes alocações com liberações logo em seguida. Desta forma, embora as infraestruturas de computação em nuvem sejam muito flexíveis e fáceis de configurar, não é fácil atingir computação de vazão extremamente alta nelas, considerando as implementações disponíveis.

A baixa amplitude da elasticidade dos provedores de IaaS atuais reflete duas realidades diferentes. Da perspectiva do cliente, o modelo de computação em nuvem permite que este aplique aos seus investimentos em TI os mesmos princípios do *Toyota Production System* (TPS) [Toyota Motor Co 2011]. Criada pela Toyota nos anos 50, a filosofia de sistema de produção “Just in Time” (JiT) é baseada em uma idéia muito simples: “o que é necessário, quando necessário e na quantidade necessária”. Os provedores de IaaS, por sua vez, não possuem as mesmas facilidades quando estão montando a infraestrutura sobre a qual eles irão prover os seus serviços, tendo que lidar com a complexidade e riscos associados com o planejamento de capacidade de longa duração.

Para lidar com esta limitação e como contribuição principal desta pesquisa, nós propomos o conceito de *Just in Time Clouds* (*JiT Clouds*) [Costa et al. 011f], uma abordagem na qual os provedores de serviço apenas incorrem em custos de provisionamento quando os recursos que eles usam para fornecer os seus serviços são demandados pelos seus clientes e apenas durante o período que eles são necessários. Isto alivia os riscos e custos do planejamento de capacidade envolvidos tanto com sub-provisionamento quanto com excesso de provisionamento de recursos. Para tal, provedores de *JiT Clouds* utilizam apenas o poder de processamento ocioso de recursos pertencentes a terceiros.

Do ponto de vista da escala, os detentores de recursos computacionais ociosos considerados aqui podem ser classificados em duas categorias principais: a) os que possuem capacidade excedente suficiente para atuar como provedores públicos de IaaS, oferecendo os seus recursos ociosos diretamente para os usuários, como fez a Amazon Bookstore, por exemplo, dando origem à AWS; e b) os que não possuem, sozinhos, recursos ociosos suficientes para uma atuação solo no mercado de IaaS.

A última categoria, que chamamos de *recursos terceirizados de pequena escala*, envolve todo o espectro de escala imediatamente inferior ao nível esperado para a primeira categoria, incluindo desde as empresas de grande porte, passando por *data centers* de pequeno porte e chegando até servidores e recursos individuais, convencionais ou não convencionais, pertencentes a instituições ou a indivíduos. Explorando tais recursos terceirizados ociosos, um fornecedor de *JiT Cloud* pode aumentar a sua capacidade de oferecer IaaS de uma forma mais escalável e com uma elasticidade virtualmente ilimitada, uma vez que é baseada na descoberta, federação e revenda de recursos ociosos cujos custos de montagem e operação são pagos por terceiros.

2. Justificativa e Relevância

A comunidade científica não está indiferente ao fenômeno da computação na nuvem e inúmeras iniciativas em todo o mundo já investigam a aplicabilidade desse novo ambiente para computação científica ou e-ciência (do inglês *e-science*) [Evangelinos and Hill 2008, Juve et al. 2009, Keahey 2010, de Oliveira et al. 2011, Iosup et al. 2008, Walker 2008]. É reconhecido que muitos dos avanços recentes em pesquisas científicas somente foram possíveis devido à habilidade dos cientistas em usar eficientemente computadores para gerar e processar grandes quantidades de dados.

Neste contexto, a elasticidade do modelo de computação na nuvem é particularmente interessante para uma classe importante de aplicações de e-ciência que são caracterizadas por cargas de trabalho que requerem computação de alta vazão. Muitas destas aplicações podem ser paralelizadas trivialmente, através da quebra do trabalho a ser realizado em várias tarefas menores que podem ser processadas independentemente. Esta classe de aplicação é referenciada na literatura como aplicações “embaraçosamente paralelas” (*embarrassing parallel*) ou simplesmente “saco-de-tarefas” (BoT, do inglês *bag-of-tasks*) [Cirne et al. 2003]. Por exemplo, as simulações de Monte Carlo, que podem envolver a execução de milhares de cenários diferentes, podem ser paralelizadas simplesmente pela execução de cada cenário em uma unidade de processamento diferente. Aplicações que processam enormes quantidades de dados podem usualmente ser paralelizadas através da divisão dos dados entre um número de processos idênticos que executam a computação sobre cada bloco de dados independentemente; no final, pode ser necessário realizar algum tipo de consolidação dos processamentos individuais [Dean and Ghemawat 2008]. A renderização de imagens complexas e vídeos se encaixa bem nesta descrição. A lista de aplicações BoT é vasta e engloba não apenas usuários da academia, mas também da indústria e do governo. Além disso, a quantidade crescente de dados gerada e consumida pela sociedade moderna deve aumentar a pressão para executar eficientemente estas aplicações [Hey and Trefethen 2003].

Se o cliente que necessita executar uma aplicação BoT fosse capaz de requisitar de um provedor de computação na nuvem tantas máquinas virtuais quanto as necessárias para maximizar o nível de paralelização da execução da aplicação, isto lhe permitiria executar esta aplicação no menor tempo possível, sem que isso implicasse em um gasto extra com os recursos computacionais usados. A elasticidade do serviço oferecido por um provedor de IaaS é, obviamente, limitada pela quantidade física de recursos que ele dispõe. Acontece que, atualmente, esse limite é muito mais restritivo, uma vez que os provedores de computação na nuvem em operação restringem a quantidade de recursos que cada cliente pode demandar de cada vez a um número relativamente muito baixo, comparado com a capacidade dos provedores.

Lidar com as demandas por elasticidade extremamente alta de aplicações HTC, BoT ou mesmo com *slashdot effects* ou *flash crowds* [Jung et al. 2002], quando um grande número de usuários acessa simultaneamente um sítio Web que adquire uma popularidade instantânea, não é uma tarefa trivial. Proporcionar tal nível de flexibilidade traz desafios enormes para o planejamento de capacidade que precisa ser realizado pelos provedores de IaaS. Para dar suporte a este tipo de utilização, esses provedores provavelmente teriam que enfrentar níveis de ociosidade de suas estruturas maiores do que os que são observados hoje, com forte impacto em sua lucratividade. Dessa forma, é pouco provável que os provedores de IaaS atualmente em operação possam vir a oferecer um serviço mais adequado para os usuários que precisam executar aplicações que demandem uma elasticidade mais extrema. O resultado desta limitação é que existe uma faixa inteira de aplicações que ainda não está sendo bem atendida pelos serviços oferecidos atualmente pelos provedores de computação na nuvem.

Contando com modelos alternativos de provisionamento que permitam custos menores ou irrelevantes para a disponibilidade de recursos, os provedores de *JiT Clouds* podem proporcionar aos clientes com aplicações HTC, em geral, e BoT, em particular, os benefícios de uma maior amplitude na elasticidade da alocação de recursos: obter o menor tempo de processamento possível sem incorrer em aumento de custos.

3. Resultados

Neste trabalho foram analisadas as razões que levam os fornecedores atuais de IaaS a imporem limites muito estritos sobre o número de recursos que qualquer cliente pode adquirir simultaneamente. Nossa avaliação utilizou um modelo de simulação para um provedor de IaaS, que é alimentado com uma carga de trabalho sintética, o que permitiu a simulação de uma ampla variedade de cenários. A utilização de um modelo mais próximo da realidade nos pareceu a escolha mais adequada para este estudo. Para minimizar a complexidade do modelo e da falta de dados de campo, foram utilizadas técnicas como projeto de experimentos (DoE), para identificar as variáveis independentes mais importantes, e a varredura de parâmetros, permitindo a instanciação de uma grande variedade de configurações distintas. Foram obtidos resultados consistentes em todos os cenários simulados.

A análise mostrou que é obrigatória a atribuição de um limite para a quantidade de recursos que podem ser alocados simultaneamente por qualquer usuário, a fim de manter a disponibilidade do serviço

suficientemente elevada e a um custo razoável para o prestador. O valor real para esse limite vai variar de provedor para provedor dependendo de sua própria avaliação de onde situa-se o seu equilíbrio, mas os nossos resultados indicam que ele tende a não ser muito maior do que os valores atualmente praticados e que se enquadram no intervalo de algumas dezenas. Observou-se também que os usuários com padrão de uso eventual ou com aplicações *BoT* pressionam a capacidade mínima necessária e aumentam a ociosidade do sistema, aumentando os custos operacionais do provedor. Além disso, mantidos o mesmo perfil da população e o mesmo valor de limite, a dinâmica do sistema independe da quantidade de usuários e, aparentemente, não constitui um contexto onde a economia de escala possa trazer melhorias substanciais.

Nosso estudo evidenciou que quando a demanda dos usuários regulares é permanente e previsível, seu crescimento é benéfico para a lucratividade do provedor, posto que não impõe um risco de super provisionamento da infraestrutura. Desta forma, o lucro do provedor é negativamente afetado somente pela parcela da demanda que vem dos usuários eventuais, a qual pode resultar no crescimento da inatividade da infraestrutura, se não for controlada. Tal aspecto é especialmente ampliado quando os usuários eventuais são ávidos consumidores de recursos e fazem requisições pontuais muito grandes.

Os resultados ajudam a entender a necessidade do uso de um limite e como o seu impacto na lucratividade do provedor está diretamente relacionado com o padrão de utilização da população de usuários, nos fazendo concluir que algumas categorias de usuários/aplicações que se beneficiariam de uma elasticidade mais ampla, tendem a continuar sendo mal servidas se um modelo alternativo de provisionamento de recursos para provedores públicos de IaaS não emergir.

Neste sentido, os passos seguintes deste trabalho foram dedicados à investigação de formas alternativas para minimizar os custos envolvidos com o aumento da capacidade dos provedores públicos de computação na nuvem para lidar apropriadamente com a demanda de usuários eventuais ávidos por recursos, tais como aqueles que precisam executar grandes aplicações científicas *BoT*. Os custos associados com a ociosidade da infraestrutura são um dos principais obstáculos para a oferta de elasticidade em condições mais flexíveis, mesmo que ainda limitada, mas que permitam que classes de aplicações de uso intenso possam se beneficiar das vantagens do modelo de computação na nuvem. A descoberta, federação e revenda de recursos terceirizados pode representar um caminho promissor, pois se baseia no aproveitamento, sob demanda, de capacidade ociosa existente em contextos onde os custos de instalação e disponibilidade são absorvidos por terceiros.

Inspirados na filosofia “*Just in Time*” (JiT) da Toyota, nós propusemos as *Just in Time Clouds* ou *JiT Clouds*, uma abordagem alternativa para a construção de provedores de IaaS baseada na utilização de recursos terceirizados, onde os provedores apenas incorrem em custos quando os recursos usados para prover a sua infraestrutura são demandados pelos seus clientes, permitindo uma ampliação de algumas ordens de magnitude no limite que precisa ser imposto aos clientes. Dessa forma, as *JiT Clouds* podem se apresentar como uma infraestrutura adequada para a execução de aplicações *BoT* de larga escala.

As *JiT Clouds* podem ser montadas sobre recursos que estejam distribuídos por todo o espectro de recursos terceirizados de baixa escala. Uma das missões do *JiT Provider* é descobrir e explorar o potencial dos recursos disponíveis alinhando-os com as necessidades das aplicações de clientes. Dependendo de suas características, os recursos terceirizados podem fornecer diferentes níveis de qualidade de serviço, elasticidade e escalabilidade. O nível de qualidade de serviço oferecido por um *JiT Data Center* (ou *JiT DC*) é totalmente dependente do nível de qualidade de serviço suportado pelos recursos usados para montá-lo, o qual está relacionado ao padrão de granularidade, volatilidade e dispersão dos mesmos.

Quando os recursos estão concentrados em centros de dados e sua capacidade está localizada mais próxima do topo da magnitude que limita a baixa escala de recursos tercerizados, os níveis de serviço oferecidos são consistentes com os praticados pelos provedores tradicionais de computação na nuvem. Dessa forma, *JiT Clouds* baseadas em recursos de baixa granularidade, baixa volatilidade e baixa dispersão podem ser usadas para hospedar aplicações tipicamente suportadas por computação na nuvem. No outro extremo do espectro da escala, quando os recursos terceirizados são de grão pequeno e distribuídos, eles precisam ser agrupados e coordenados pelo *JiT Provider* para a sua exploração.

Para demonstrar a sua viabilidade, nós analisamos o potencial das *JiT Clouds* no seu cenário mais desafiador: considerando o uso de recursos computacionais de alta granularidade, alta volatilidade e alta

dispersão para a composição de *JiT DCs* de alta vazão. Neste sentido e usando o conceito de redes de *broadcast*, foi proposta uma nova arquitetura, chamada de *Infraestrutura Computacional Distribuída Sob Demanda* (ou *OddCI*), para construção de *JiT DCs* dinâmicos baseados em tais recursos computacionais através do uso de mecanismos específicos para a sua descoberta, alocação e coordenação. Nossos resultados de simulação mostram que, mesmo em cenários de altíssima volatilidade de nós autônomos e distribuídos geograficamente e sem o uso de algoritmos compensatórios ótimos, foi possível obter disponibilidade coletiva de dispositivos isolados para entregar vazão computacional com perdas máximas de 10% sob regimes de até 40% de volatilidade de nós, causada por falhas ou abandonos voluntários.

O uso da capacidade ociosa de processamento de muitos recursos computacionais distribuídos, tais como os dos receptores de TV digital já havia sido demonstrada antes. Com a generalização feita com a arquitetura *OddCI* e a construção de uma prova de conceito com a implementação da sistema *OddCI-DTV* sobre uma rede de TV Digital, a montagem de um testbed real e uma avaliação do seu desempenho foi possível constatar não apenas a viabilidade dessa abordagem como também o fato de que a mesma é promissora.

Tirando partido das funcionalidades já disponibilizadas sobre os dispositivos que implementam tais tecnologias ou complementando e/ou adaptando estas funcionalidades, é possível projetar implementações de Sistemas *OddCI* para diversos contextos.

Embora o foco da pesquisa tenha sido a investigação da viabilidade técnica da abordagem proposta, há algumas evidências que apontam para a sua viabilidade do ponto de vista econômico.

Pela ótica dos proprietários dos recursos, um dos aspectos importantes a serem considerados é que a recompensa percebida pelo fornecimento dos recursos excedentes seja superior aos custos envolvidos na própria cessão e permitam também um alívio nos custos que ocorrem independentemente dela. Ou seja, devem cobrir os custos de utilização e permitir a amortização, em algum grau, dos custos de disponibilidade associados com a manutenção de recursos excedentes, que continuam sendo de sua responsabilidade.

Do ponto de vista do provedor da *JiT Cloud*, a vazão computacional ofertada deve ser atrativa e equilibrar preço e qualidade de serviço com o custo de operação da federação. Como o serviço prestado pode ser, potencialmente, muito mais elástico que os serviços ofertados pelos provedores atuais de computação na nuvem, o preço praticado por um *JiT Provider* pode ser balizado, no mínimo, com o preço cobrado pelos provedores de *IaaS* por recursos de capacidade similar. Note que, mesmo no caso de recursos não convencionais, dispositivos mais modernos já apresentam este tipo de equivalência com algumas classes de máquinas virtuais comercializadas.

Como o ônus do custo de disponibilidade dos recursos permanece como uma responsabilidade dos seus proprietários e o custo de utilização somente ocorre quando os recursos são efetivamente utilizados, o custo de coordenação da federação é o insumo mais relevante para o provedor da *JiT Cloud*. Considerando que o custo de coordenação é uma função do tamanho da infraestrutura a ser gerenciada e não da forma com a mesma foi montada, possivelmente o custo de coordenação de uma *JiT Cloud* se manterá nos mesmos patamares apresentados por serviços baseados em infraestruturas próprias com a mesma categoria e tamanho. Entretanto, a coordenação da federação pode ser impactada pelo nível de serviço suportado pelos recursos envolvidos. Em especial, cenários de alta volatilidade podem apresentar níveis de falha que causem reflexos tanto nos custos operacionais da federação, pelo aumento do nível de redundância praticado, quanto na reputação do provedor da *JiT Cloud*, que pode ser afetada por quedas na vazão entregue e por outras violações em SLAs.

Para algumas classes de aplicação, as *JiT Clouds* podem se apresentar como uma alternativa de maior valor agregado. É o caso em que a capacidade de prover grandes DCIs em regime de elasticidade extrema se torna um diferencial competitivo. Neste sentido, a escolha adequada pelo provedor da *JiT Cloud* dos recursos terceirizados a serem federados em cada situação é fundamental. Por exemplo, no caso de recursos de uma rede de TV Digital, além da capacidade computacional requerida para os recursos, a observância de outros aspectos como audiência e horário de alocação, podem permitir o controle sobre a escala a ser atingida e a volatilidade a ser evitada.

4. Contribuições

Os principais resultados e contribuições deste trabalho, considerando as três questões de pesquisa que foram abordadas, são os seguintes:

Por que os provedores de nuvens públicas impõem limites que restringem a utilidade de seus serviços para clientes com aplicações BoT?

- Investigação das causas que levam os provedores públicos de computação na nuvem a impor um limite estrito na quantidade de recursos que um único usuário pode adquirir concomitantemente e análise de qual o impacto que eventuais aumentos no limite imposto apresentam sobre a lucratividade do provedor. Este resultado foi publicado no *XXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos* (SBRC 2011): “Sobre a Amplitude da Elasticidade dos Atuais Provedores de Computação na Nuvem” [Costa et al. 011a]. Este trabalho ganhou um prêmio de distinção, sendo considerado um dos cinco melhores trabalhos do evento. Uma versão estendida do mesmo foi aceita para publicação na *Revista Brasileira de Redes e Sistemas Distribuídos* (RB-RESO), edição 2011/02: “Uma Análise do Impacto da Elasticidade no Lucro de Provedores de Computação na Nuvem” [Costa et al. 011e]. Uma versão em inglês ampliada da análise foi apresentada no *2nd International Workshop on Cloud Computing and Scientific Applications (CCSA 2012)* [Costa et al. 012c], realizado em conjunto com o *12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2012)* e, posteriormente, uma versão estendida foi publicada no periódico *Elsevier Future Generation Computer Systems*: “Analyzing the Impact of Elasticity on the Profit of Cloud Computing Providers” [Costa et al. 012e];

Como podemos servir adequadamente os usuários BoT em um cenário de IaaS?

- Uma proposta de abordagem alternativa para montagem da infraestrutura computacional de um fornecedor de computação na nuvem com recursos de terceiros. A proposta introduz o conceito de *Just in Time Clouds*, cujos provedores apenas alocam os recursos quando eles são exigidos e somente durante o período que eles são necessários para os seus clientes. Isso elimina a necessidade de antecipar o planejamento de capacidade e exclui os custos associados ao excesso de provisionamento de recursos. Este resultado foi apresentado como poster na *3rd IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2011)*: “Just in Time Clouds: Enabling Highly-Elastic Public Clouds over Low Scale Amortized Resources” [Costa et al. 011f]. Esta mesma abordagem foi submetida em 2010 na forma de um projeto para um edital da RNP/CTIC na área de Computação na Nuvem e foi aceito. Este projeto nomeou o consórcio *JiT Clouds*, uma das duas redes de pesquisa do CTIC na área de computação na nuvem, a qual foi coordenada pela UFCG e congregou 17 instituições nacionais e internacionais em oito subgrupos de pesquisa;

É possível construir JiT DCs nos cenários mais desafiadores, que envolvem recursos terceirizados de alta granularidade, alta volatilidade e alta dispersão?

- Uma proposta de uma nova arquitetura para computação distribuída que é ao mesmo tempo flexível e altamente escalável. Chamada de *OddCI*, ela é suportada pela existência de um grande contingente de dispositivos que podem ser acessados simultaneamente através de uma rede de transmissão em *broadcast*. Este resultado foi publicado no *2nd Workshop on Many-Task Computing on Grids and Supercomputers*, realizado em conjunto com o *Supercomputing 2009*: “OddCI: On-demand Distributed Computing Infrastructure” [Costa et al. 2009];
- Implementação de um protótipo de sistema *OddCI* em um ambiente real de TV Digital para validação do conceito e obtenção de medições de campo. Um artigo com os resultados obtidos foi publicado na *13th IEEE/ACM International Conference on Grid Computing*: “OddCI-Ginga: A Platform for High Throughput Computing Using Digital TV Receivers” [Costa et al. 012c];
- Um artigo consolidando esses e os outros resultados relacionados com a arquitetura *OddCI* foi publicado no periódico *Journal of Grid Computing* em 2012: “Using Broadcast Networks to Create On-demand Extremely Large Scale High-throughput Computing Infrastructures” [Costa et al. 012d] e um pedido de patente de invenção foi protocolado no INPI sob o N° ”BR 10 2013 027255-8”.

5. Referências

As referências citadas podem ser encontradas na Lista de Referências da tese de doutorado resumida aqui, cujo texto está disponível no sítio http://docs.computacao.ufcg.edu.br/posgraduacao/teses/2013/Tese_RostandCosta.pdf.