Redes Centradas na Informação: Uma Comparação de Abordagens

Bruno Magalhães Martins & Antônio Marcos Alberti
Instituto Nacional de Telecomunicações - Inatel
Santa Rita do Sapucaí - MG - Brazil

brunom@mtel.inatel.br, alberti@inatel.br

Abstract. *The current Internet was designed using the host-centric principle, which put host connectivity on the center of its original design. As a result, the network evolved to provide hosts communication through IP routing. Complex functions were moved from network core to end hosts. Recent surveys suggest that a significant portion of the current Internet traffic is related to content exchange, instead of original Internet applications. New communication models, such as peer-to-peer, appeared to improve content distribution and sharing on the Internet. However, in recent years the host-centric model began to be questioned and a new paradigm emerged: the information-centric. Since the Internet today is basically a content exchanging network, why not to focus on this aspect of its evolution? In this paper, we discuss NetInf, CCN e PSIRP information-centric approaches, presenting their characteristics and providing a comparing of them. The aim is to critically review some of the existing proposals and determine opportunities and challenges for future research.*

Resumo. *A Internet atual foi projetada utilizando o paradigma host-centric, que colocou os terminais da rede no centro do projeto original. Como resultado, a rede evoluiu para conectar hosts através do roteamento IP. Funções complexas foram retiradas do núcleo da rede e movidas para os terminais. Levantamentos recentes sugerem que parte significativa do tráfego atual da Internet é voltada à transferência de conteúdos, e não mais para as aplicações originais da rede. Novos modelos de comunicação, dentre eles o peer-to-peer, surgiram para melhorar a distribuição e a troca de conteúdos na Internet. Entretanto, nos últimos anos o modelo host-centric começou a ser questionado e um novo paradigma apareceu: o information-centric (ou redes centradas na informação). Já que a Internet hoje basicamente é uma rede de transferência de conteúdos e informações, porque não centrar sua evolução neste aspecto, ao invés da conectividade de terminais? Neste artigo, discutiremos as propostas de redes centradas na informação NetInf, CCN e PSIRP, apresentado suas características e comparado-as. O objetivo é analisar criticamente algumas das propostas existentes e determinar oportunidades e desafios para pesquisa futura.*

1. Introdução

Diante da disseminação do uso da Internet e o aumento da acessibilidade aos dispositivos de criação e modificação de mídia e informação, hoje as redes de comunicação passam por uma fase em que grande parte dos usuários conectados está deixando de ser um mero consumidor de conteúdos, para se tornar produtor. A maioria destes conteúdos (fotos, vídeos, *posts*, etc.) é disponibilizada na rede e o grande problema é que na arquitetura atual, toda a preocupação é centrada nos *hosts*, não nos conteúdos gerados e armazenados por estes *hosts*. Em outras palavras, os usuários se

importam com o conteúdo que a rede possui, enquanto a Internet atual é focada na localização do armazenamento deste conteúdo. Há uma série de problemas oriundos desta incompatibilidade de modelos, tais como disponibilidade de conteúdo, segurança e dependência de localização. Para Van Jacobson[1] [Jacobson et al. 2009], a forma direta e unificada de resolver estes problemas é substituir o "onde" pelo "o que" em relação ao conteúdo.

Nas redes atuais, o endereço IP é sobrecarregado ao se responsabilizar pela dupla funcionalidade de identificação e localização dos *hosts*. Ou seja, a mobilidade na rede muitas vezes leva a troca do endereço IP, o que pode levar indiretamente a troca de identificação na rede. É como se você trocasse de identificação ao se mover. Por isto é tão difícil identificar o originador de um ataque em uma rede IP. O desacoplamento da identificação e localização existente no endereço IP permite a melhoria do suporte a segurança, mobilidade, *multihoming*[2]. Além disto, a amarração das informações na *web* ao URL (*Uniform Resource Locator*) e ao endereço do domínio e do *host* onde elas se encontram, exige o conhecimento prévio da localização da informação para que o acesso a mesma possa ser feito.

Nas redes orientadas/centradas em informação/conteúdo o fluxo de mensagens é dirigido para os nós que manifestam o seu interesse através de identificadores de nomes da informação, em vez de nomes de interfaces de *hosts* [Rothenberg et al. 2008]. Elas possuem como funcionalidade principal a interconexão dos produtores de conteúdo aos consumidores, independente das localizações dos *hosts* envolvidos na comunicação. Desta forma, consegue-se um aumento da eficiência na disponibilidade e disseminação da informação, já que a recuperação dos conteúdos pode ser beneficiada ao utilizar cópias ao longo da rede. Assim, do ponto de vista do usuário, toda informação terá um identificador único e esta informação poderá ser recuperada de forma mais eficiente já que a informação pode ser encontrada em um local mais próximo ou até mesmo no seu ambiente local, além do aumento do desempenho e facilidade no acesso a estas informações. Ou seja, o grande desafio destas redes é, portanto, estabelecer um padrão para nomeação da informação de forma consistente e eficiente, além de fornecer soluções para roteamento das informações baseadas em informações nomeadas.

O restante deste artigo é dividido da seguinte forma. Na Seção 2 serão apresentadas três importantes abordagens de redes centradas na informação: NetInf [Niebert et al. 2008], CCN (*Content-Centric Network*) [Palo Alto Research Center 2010] e PSIRP (*Publish/Subscribe Internetworking Routing Paradigm*) [Ain et al. 2008]. Na Seção 3, estas abordagens são comparadas. Isto também é feito por [Ahlgren et al. 2010]. Finalmente, na Seção 4 são apresentadas as conclusões do artigo.

2. Abordagens de Redes Centradas na Informação

Nas seções seguintes apresentaremos uma visão geral das abordagens NetInf, CCN e PSIRP, bem como algumas comparações de características mais relevantes destas abordagens. É importante salientar que apesar das particularidades, os três projetos a

[1] Van Jacobson é o autor do cabeçalho TCP/IP e trabalha atualmente como pesquisador do PARC (*Palo Alto Research Center Incorporated*) no programa de investigação de redes centrados em conteúdo. Seu ponto de vista é preservar as decisões de *design* do TCP/IP: simplicidade, robustez e escalabilidade [Jacobson et al. 2009][Palo Alto Research Center 2010].

[2] O termo *multihoming* significa que existem múltiplas possibilidades de conexões para o acesso à Internet através de uma rede ou de uma estação. Ou seja, significa ter múltiplas interfaces com múltiplos localizadores para um mesmo *host*, ao mesmo tempo.

serem citados adotam o modelo de comunicação publica/assina (*publish/subscribe*), na qual o interessado em uma determinada informação deve solicitá-la a rede.

Publicar significa disponibilizar qualquer conjunto de dados na rede, enquanto a assinatura de um conteúdo corresponde a uma porção de dados solicitada por algum interessado na informação. O modelo de comunicação publica/assina (*publish/subscribe*) visa trazer ao receptor o controle do fluxo de comunicação, saindo, portanto do modelo atual "receptor aceita tudo". Os termos "centrado no conteúdo" e "orientado à informação" são aplicados essencialmente da mesma forma para indicar redes nas quais os objetos de informação são em si o foco principal do projeto. A seguir, portanto, os termos informação e conteúdo serão utilizados de forma equivalente.

2.1. 4WARD NetInf

O 4WARD (*Architecture and Design for the Future Internet*) é um projeto de nova Internet que tem o apoio do FP7 (*EU Framework Programme 7*) e visa criar uma nova arquitetura de rede baseada na abordagem *clean slate*. O projeto 4WARD é dividido em seis grupos de trabalho incluindo o NetInf [Niebert et al. 2008]. A ideia principal do NetInf [Dannewitz 2008][Ahlgren et al. 2010][Dannewitz et al. 2008][Ahlgren et al. 2008] é obter um sistema para recuperação de informação baseado em um identificador único desta informação. Para tal, é necessário um mecanismo de resolução de nomes que faça o mapeamento de um identificador para um localizador desta informação. A proposta considera que este mecanismo deve ser persistente e independente de cópia ou localização. Em outras palavras, a identificação da informação não pode mudar ao se relacionar com uma cópia ou com uma localização diferente da informação.

Além disso, o mecanismo de resolução de nomes para identificadores de informação do NetInf é concebido de forma plana, ou seja, não hierárquico. Assim, este mecanismo plano é incompatível com a concepção hierárquica do DNS (*Domain Name System*) atual. Apesar do mecanismo plano não ser legível a humanos, este mecanismo suporta melhores mecanismos de segurança e privacidade e não exige entidades centralizadas de gerenciamento de nomes. Para [Dannewitz 2008], o NetInf estende o conceito de desacoplamento da identificação e localização com outro nível de indireção, visando desacoplar os objetos de informação de seus locais de armazenamento, ou seja, criar acesso à informação independente do local de armazenamento e se beneficiar de cópias distribuídas ao longo da rede.

O modelo de informação do NetInf é baseado em dois objetos: o objeto de informação (IO – *Information Object*) e o objeto de dados (DO – *Data Object*). Os IOs são objetos que representam a informação propriamente dita. Eles podem representar uma imagem, um arquivo de som, um conteúdo de vídeo, páginas *Web* e assim por diante. Já os DOs são um tipo especial de IO que representam um objeto no seu nível de *bit*. Um DO aponta, por exemplo, para um arquivo, uma seqüência binária resultante da codificação de mídia, tal como um arquivo com codificação MP3, ou para um texto. A partir deste modelo de informação do NetInf, a pesquisa e recuperação da informação pode se desenvolver de forma mais eficiente. Uma pesquisa baseada em um IO, por exemplo, pode referir-se uma música específica sem especificar a codificação desta música ou a orquestra que a executa, dado que muitas vezes estas informações não são relevantes ao usuário. Os IOs, portanto, possibilitam que os usuários localizem o conteúdo interessado independentemente de sua representação específica ou outras características, que a princípio, não são interessantes a este usuário.

Segundo [Dannewitz 2008], a pesquisa de objetos de informação do NetInf é feita através de um mecanismo de busca integrado na arquitetura NetInf, ou alternativamente, através de um mecanismos de busca externa. Ambos, porém, devem incluir novos tipos de funcionalidades baseadas em atributos do mundo real de entidades físicas, tais como posição GPS (*Global Positioning System*) ou etiquetas de RFID (*Radio Frequency Identification*).

O paradigma publica/assina (*publish/subscribe*) do NetInf é implementado a partir da publicação de um objeto de informação. Esta publicação é feita ao registrar o nome de determinado objeto de informação em um sistema de resolução de nomes, e a assinatura por sua vez, é feita quando algum interessado solicita ao sistema de resolução de nomes algum objeto publicado. A localização e assinatura de um objeto de informação no NetInf só é possível, portanto, depois que este for publicado.

Para que se tenha segurança e confiança no conteúdo a ser disseminado na rede, o NetInf propõe que a informação seja auto-certificável, ou seja, a confiança deve ser nativa do conteúdo, ao contrário das redes de hoje em que a confiança na informação é baseada no *host* que enviou esta informação. A integridade do conteúdo é garantida através de uma ligação criptográfica entre o objeto de informação e o nome deste objeto e a autenticidade é garantida a partir da aplicação de uma função *hash* de chave pública como parte do nome auto-certificável. A verificação da autenticidade, portanto, necessita de uma relação com uma entidade externa.

Os nomes da arquitetura NetInf são constituídos de três campos: um campo Tipo responsável por definir o formato do nome, que especifica por exemplo, o algoritmo *hash* utilizado no processo de auto-certificação; o campo Autenticador, responsável por fazer a ligação entre o identificador do IO à chave pública; e finalmente, o campo Rótulo, responsável por fazer a identificação do objeto individual. A combinação do Autenticador com o Rótulo necessita ser único globalmente.

A arquitetura NetInf faz o uso de DHTs (*Distributed Hash Tables*) para o roteamento dos conteúdos. Segundo [Cirani e Veltri 2007], as DHTs são uma classe de sistemas distribuídos descentralizados que prestam um serviço de pesquisa semelhante a uma tabela *hash*. Pares <*chave, valor*> são armazenados na DHT, e qualquer nó pode participar recuperando o valor associado a uma dada chave. A responsabilidade por manter o mapeamento entre as chaves e valores é distribuída entre os nós participantes da topologia virtual DHT. Isso permite as DHTs escalarem um número extremamente grande de nós e lidar com as chegadas e partidas contínuas de nós. As DHT formam uma infra-estrutura que pode ser usada para criar serviços mais complexos, tais como sistemas de arquivos distribuídos, compartilhamento de arquivos *peer-to-peer* e sistemas de distribuição de conteúdo, *web caching* cooperativo, *multicast*, *anycast*, serviços de nome de domínio e mensagens instantâneas.

A recuperação e entrega do conteúdo em NetInf se resume na combinação de dois processos: o de resolução de nomes, que se responsabiliza por prover a localização para um determinado identificador de objeto através de consulta em uma DHT; e o processo de roteamento, que se responsabiliza por encaminhar a requisição para a localização do armazenamento e retornar o conteúdo associado.

No NetInf existem duas abordagens para lidar com estes dois processos citados: a primeira alternativa separa a fase de resolução de nomes da fase de roteamento, como usado tradicionalmente nos esquemas de roteamento baseado em topologia, tais como o

OSPF (*Open Shortest Path First*) e o BGP (*Border Gateway Protocol*); a segunda alternativa seria um esquema integrado de roteamento baseado em nomes de conteúdos (LLC – *Late Locator Construction*), que combina o caminho de resolução e o caminho de recuperação em um processo único. Segundo [Ahlgren et al. 2010], este último esquema executa o roteamento através do mapeamento direto entre o identificador do objeto de dados e a rota, o que diminui a latência e resulta, provavelmente, em melhor desempenho. Ou seja, no roteamento LLC o localizador de um objeto é construído dinamicamente baseado no caminho seguido pelos primeiros pacotes de sinalização enviados. O assinante consulta o localizador do nó publicador no sistema de resolução de nomes e envia pacotes de controle para o publicador, a fim de levantar informações atualizadas da topologia da rede. A localização do objeto é construída no último momento antes de enviar os pacotes de dados, daí a origem do termo *Late Locator Construction*.

2.2. CCN – Content-Centric Network

CCN é um projeto do PARC (*Palo Alto Research Center*) que visa transformar a forma como as redes de comunicação de dados atuais operam, através de uma mudança na arquitetura de rede que faça a recuperação do conteúdo pelo seu nome, e não pela sua localização. Neste contexto, a principal motivação das CCNs é melhorar a eficiência da utilização da Internet atual, direcionando o foco da rede para a distribuição de conteúdo.

Desta forma, as CCNs propõem um novo mecanismo de distribuição de conteúdo com os mesmos princípios de engenharia do IP, porém, utiliza a nomeação de conteúdo, em vez de identificadores de *hosts* como foco principal. Segundo Van Jacobson em [Palo Alto Research Center 2010], a visão das CCNs é reutilizar os elementos bem sucedidos do TCP/IP, e construir uma nova rede, substituindo o modelo IP centrado em *hosts* por um modelo orientado a conteúdo, para ser utilizado como o protocolo central de interconexão de redes.

Conforme [Jacobson et al. 2009], nas CCNs existem dois tipos de pacotes: o de interesse e o de dados. Um consumidor envia por *broadcast* a requisição do conteúdo interessado através de um pacote de interesse. Esta informação fica armazenada em uma tabela de interesses pendentes - PIT (*Pending Interest Table*). Uma base de informação de encaminhamento - FIB (*Forwarding Information Base*) é utilizada para encaminhar pacotes de interesse da tabela PIT, em busca de uma potencial fonte (nó publicador ou cache) para a informação desejada. Qualquer nó que tenha o conteúdo de mesmo nome do pacote de interesse enviado pelo consumidor pode responder com um pacote de dados. Para que uma potencial fonte consiga identificar e localizar o assinante do conteúdo, os pacotes de interesse deixam, figurativamente falando, “migalhas de pão” ao longo do caminho, para que os pacotes de informação retornem pelo sentido reverso do caminho. Desta forma, é possível maximizar o compartilhamento de informações na rede, já que cada pacote enviado pode ser recuperado por vários consumidores interessados. Esta característica reduz os custos de operação, pois uma única cópia pode ser compartilhada por vários assinantes [Jacobson et al. 2009].

O controle de fluxo nas CCNs é baseado numa regra simples na qual um pacote de interesse pode recuperar no máximo um pacote de dados. Em outras palavras, pacotes de dados e de interesse utilizam uma razão de um para um, o que mantém o controle de fluxo balanceado. Um controle de fluxo semelhante é utilizado no TCP, entre o pacote de dados e o pacote de reconhecimento (*Ack - Acknowledge*). Os pacotes

de interesse têm o mesmo papel do anúncio da janela deslizante do TCP, ou seja, o tamanho da janela do transmissor pode variar a partir da quantidade de pacotes de interesse enviados [Jacobson et al. 2009]. Esta regra básica garante que o controle de fluxo seja mantido e permite a comunicação de forma eficiente entre várias máquinas sobre diversas redes heterogêneas.

A segurança nas CCNs é nativa do conteúdo, ou seja, a proteção e a confiança "viajam" com o próprio conteúdo, evitando assim, várias das vulnerabilidades das redes IP atuais. Ao contrário do NetInf, nas CCNs os nomes são tipicamente hierárquicos. No contexto das redes centradas na informação, os mecanismos para resolução de nomes de forma hierárquica podem criar nomes legíveis a humanos, além de fornecerem "dicas" para a resolução de localizadores.

Ainda segundo Jacobson [Jacobson et al., 2009], nas CCNs, a relação de confiança com o publicador é feita através do prefixo do nome do conteúdo, já que cada pacote enviado contém a assinatura do publicador explícita no seu nome. A assinatura é feita em cada pacote através do seu respectivo nome e estas assinaturas são padrões de chaves públicas fornecidas por uma PKI (*Public Key Infrastructure*). A verificação do nome com o conteúdo é feita através da respectiva chave privada, pertencente ao requisitante do conteúdo. A confiança na chave de assinatura e a integridade dos dados devem ser fornecidas através de mecanismos adicionais. Por exemplo, através de um certificado baseado em alguma PKI. O paradigma publica/assina (*publish*/*subscribe*) nas CCNs se dá quando um nó anuncia o nome de determinado conteúdo ao roteador. É assim que a informação é publicada. A assinatura é feita quando um nó interessado no conteúdo envia um pacote de interesse em busca de um potencial publicador. Os pacotes de interesse são enviados por *broadcast* e a escolha de um dado publicador é feita a partir da análise do maior prefixo correspondente ao nome especificado no pacote de interesse. Os pacotes de dados seguem o caminho reverso até alcançar o emitente do pacote de interesse (nó assinante).

2.3. PSIRP – Publish/Subscribe Internetworking Routing Paradigm

O PSIRP [Ain et al. 2008] é um projeto Europeu financiado pelo FP7 que começou em janeiro de 2008 e tem previsão de término em 2010. O paradigma de roteamento publica/assina (*publish/subscribe*) visa a troca do atual modelo de comunicação da Internet "receptor aceita tudo que lhe for enviado" por um paradigma de comunicação consensual ou autorizada e controlada totalmente pelo receptor (ou assinante). O PSIRP propõe o projeto e análise de protocolos criptográficos focados apenas em publicar e assinar o conteúdo, em vez do paradigma enviar e receber da Internet atual. Publicar significa tornar um conjunto de dados disponível na rede, enquanto a assinatura de um conteúdo corresponde a uma porção de dados solicitada por algum interessado na informação publicada.

Para [Wong et al. 2008], o paradigma de roteamento publica/assina (*publish/subscribe*) é um mecanismo atraente para a localização e recuperação de conteúdo, uma vez que tal mecanismo trata de forma independente os publicadores e os assinantes de conteúdo. No PSIRP, as mensagens não são dirigidas a qualquer nome de receptor, já que não existem nomes de receptor ou transmissor fornecidos por redes. Desta forma, o roteamento da requisição e do próprio conteúdo no PSIRP é dividido em dois processos: um para o encontro do solicitante (assinante) com uma potencial fonte

(publicadora) do conteúdo, e outro destinado à entrega dos pacotes de conteúdo ao assinante.

Quando um nó publica algum dado, a transferência da informação propriamente dita não ocorre. O que ocorre é a publicação da disponibilidade desta informação em um sistema chamado de *rendezvous* ou ponto de encontro. O rendezvous é responsável pelo encontro entre o publicador e o assinante de determinada informação. Em outras palavras, quando um nó se inscreve para determinada informação, a rede encontra a publicação de interesse e cria o caminho de entrega entre o publicador e o assinante desta informação.

A arquitetura RTFM (*Rendezvous, Topology, Forwarding and Mediation*) do PSIRP é composta por quatro funções [Rothenberg et al. 2008][Zahemzky et al. 2010]: ponto de encontro (*rendezvous*), gerenciamento de topologia, encaminhamento e mediação. O *rendezvous* é responsável por fazer a ligação de um interessado em um conteúdo com uma ou mais potenciais fontes para este conteúdo. O gerenciamento de topologia constrói e mantém caminhos entre diferentes redes de transmissão baseadas no paradigma publica/assina. O encaminhamento realiza a entrega de pacotes baseada em técnicas de comutação por rótulos. E finalmente, a mediação refere-se à transmissão física de dados entre dois nós.

Conforme [Ain et al. 2008], a arquitetura do PSIRP possui ainda um mecanismo de escopo, utilizado para limitar o alcance de uma informação publicada. Com base em políticas de governança, um usuário pode construir uma relação de confiança e privacidade entre as entidades participantes na comunicação (publicadores, assinantes e a própria informação). Uma determinada publicação pode estar disponível em diversos escopos e pode ser inserida em um escopo a partir do interesse de um publicador ou de um assinante.

Ainda segundo [Ain et al. 2008], os identificadores de aplicação (AIds – *Application Identifiers*) são utilizados por publicadores e assinantes para classificar, distinguir e rastrear entidades como dispositivos, produtos, pessoas, títulos de conteúdos (músicas, endereços de páginas *web*), serviços (endereço de serviços *web* ou de email, número de telefone), etc. Estes identificadores, diferente dos outros identificadores a serem apresentados, são definidos usando a semântica da aplicação e contextos específicos.

Os identificadores de *Rendezvous* (RIds – *Rendezvous Identifiers*) são utilizados para ligar os identificadores de nível superior (identificadores de aplicações) com os identificadores das camadas inferiores (identificadores de encaminhamento). Os RIds podem ainda incluir operações como autenticação de rede, controle de acesso de publicadores e de assinantes interessados em participar de um determinado encontro, além de permitir o encaminhamento de solicitações de conteúdo para um determinado publicador [Zahemzky et al. 2010].

Os identificadores de encontro podem ser criados por um emissor ou um receptor de dados a fim de estabelecer uma conexão. Em outras palavras, o publicador pode adquirir um identificador e enviar para o assinante, ou um assinante pode selecionar um identificador para o encontro e informar a um potencial publicador. Existe ainda uma terceira opção para a escolha de identificadores, onde uma aplicação ou um serviço necessita de um identificador de encontro para estabelecer a comunicação.

Os identificadores de escopo (SIds – *Scope Identifiers*) são utilizados para delimitar a acessibilidade do conteúdo publicado. Os SIds são implementados pelos RIds para limitar a distribuição de conteúdo a regiões específicas e definir a rota para a entrega dos pacotes de dados, ao definir o FId (*Forwarding Identifiers*). Um dos desafios do PSIRP é gerenciar a escalabilidade relacionada às estruturas de escopos, já que para estabelecer um determinado ponto de encontro, precisa-se dimensionar corretamente o escopo relacionado.

Ainda segundo [Ain et al. 2008], um par (RId, SId) é usado pelo sistema de encontro para determinar o conjunto apropriado de identificadores de encaminhamento para a entrega dos dados para um conjunto de assinantes. Um único RId pode ser associado a um conjunto de SIds. Em outras palavras, uma publicação de dados pode ser associada a um ponto de encontro estabelecido pelo publicador, e ao mesmo tempo, estar ligada a diferentes escopos de acessibilidade. Resumidamente, existe pelo menos um ponto de encontro por escopo. O publicador de uma informação deve publicá-la dentro de um escopo específico, identificando-o durante a operação de publicação.

Os identificadores de encaminhamento (Fids) são utilizados para o transporte das publicações através das redes. Cada par ativo (RId, SId) possui um mapeamento de pelo menos um identificador de encaminhamento. As árvores de transmissão identificadas por um FId podem ser compartilhadas para finalidades de escalabilidade [Ain et al. 2008].

Para questões de segurança, é esperado que os identificadores sejam concebidos de maneira semelhante à utilizada pelo HIP, ou seja, utilizando o identificador baseado em uma chave pública cuja a correspondente chave privada é de posse do publicador. Isso permite que o *host* publicador escolha um identificador que possa ser usado por assinantes para autenticar os dados e verificar se eles foram enviados de fato pelo publicador que possui a associada chave privada.

Em PSIRP, as propriedades de segurança para mensagens de controle mais importantes são a proteção da integridade e a autenticação. A proteção da integridade pode verificar, por exemplo, se o pacote foi modificado, e a autenticação garante que o publicador possui a permissão do escopo para publicar uma informação com um determinado par (RId, SId) [Zahemzky et al. 2010].

Conforme [Zahemzky et al. 2010], o PSIRP possui ainda uma autenticação a nível de pacote (PLA - *Packet Level Authentication*). A PLA é uma função para prover segurança aos *hosts* finais através de uma assinatura criptográfica feita individualmente em cada pacote. Ela permite ainda, que os nós ao longo do caminho de entrega verifiquem os pacotes sem que existam associações de confiança pré-estabelecidas com o publicador. Em PSIRP, a PLA é utilizada geralmente para segurança de mensagens de controle (mensagens publica/assina), podendo ser ampliada para a sua utilização em todo o tráfego.

O PSIRP possui ainda um mecanismo anexado aos pacotes para melhorar o desempenho do encaminhamento de dados, denominado de filtros de Bloom [Ahlgren et al. 2010]. Os filtros de Bloom são uma simples estrutura de dados aleatória com o propósito de consultar e selecionar elementos de forma eficiente em um dado conjunto de participantes, como por exemplo, a escolha de uma rota em um conjunto de rotas. A desvantagem dos filtros de Bloom é que estes permitem falsos positivos, em contrapartida, a economia de espaço muitas vezes compensa essa desvantagem quando

a probabilidade de um erro é controlada ou é suportada por alguma aplicação [Ahlgren et al. 2010].

3. Comparações entre NetInf, CCN e PSIRP

Nesta seção são comparadas e discutidas características e opções de projeto mais relevantes entre as arquiteturas apresentadas.

3.1. Nomeação da Informação

Em qualquer projeto de redes centradas na informação, deve ser possível se referir à própria informação, independentemente das redes em que ela se encontra. Um esquema de nomeação para objetos de informação é, portanto, talvez a parte mais importante do projeto [Ahlgren et al. 2010].

A princípio, os nomes NetInf e PSIRP são semelhantes por usarem um esquema de nomeação plano, e divergem pelo fato do NetInf utilizar um único *namespace*, enquanto o PSIRP utiliza dois *namespaces*, um para o encontro e outro para o encaminhamento. O *namespace* especificado pelo NetInf é semelhante ao do PSIRP destinado ao encontro. Por outro lado, o NetInf não possui o *namespace* destinado ao encaminhamento, pois este *namespace* é parte do protocolo particular da camada de rede. Já o esquema de nomeação da CCN utiliza a concepção hierárquica. A estrutura de nomes para a informação é concebida em forma de árvore, sendo que a raiz da estrutura representa o prefixo do nome da informação. Isto torna este prefixo único e exclusivo a um determinado publicador.

Os esquemas de nomeação hierárquicos têm a vantagem de estabelecer nomes legíveis a humanos e aptos à interação através de um *browser*. Outra vantagem dos sistemas hierárquicos é que estes permitem a agregação de estado de roteamento – *routing state* – além do fato que, o conteúdo CCN pode ser criado dinamicamente em resposta a uma solicitação. Esta peculiaridade das CCNs só é possível devido à sua estrutura de nomes hierárquicos concebidos de forma significativa. Em contrapartida, os nomes planos possuem melhores propriedades de segurança, já que não dependem de entidades centralizadas de resolução de nomes para desenvolver esta atividade, permitindo a auto-identificação das informações. Esta última peculiaridade é bastante atraente às aplicações que necessitam gerar informações com identificadores de forma autônoma. A questão mais importante, portanto, seria como fazer a ligação entre o esquema de nomeação de forma plana e um sistema de resolução de nomes para endereços como o DNS atual, já que este último é baseado em identificadores hierárquicos e, portanto, inapropriado para identificadores planos. Ou ainda, no caso de soluções *clean-slate*, como relacionar os nomes planos a localização e posterior encaminhamento.

Existem outras propostas de encaminhamento de conteúdo e consulta a sistemas de *cache* baseado em nomes planos na literatura como o SPSwitch baseado em filtros de Bloom [Rothenberg et al. 2008]; consulta para roteamento de conteúdo baseado em DHTs [Ahlgren et al. 2010][Liu et al. 2008]; além do roteamento baseado em rótulos planos (ROFL – *Routing on Flat Labels*) [Campista 2010].

3.2. Questões de Segurança

O fato de que nas redes centradas em conteúdos, as mensagens enviadas contêm controles relacionados apenas ao conteúdo cria um isolamento natural entre informação e os *hosts* da rede, tornando estes menos vulneráveis, já que usuários mal-intencionados terão mais dificuldade em acessar estes *hosts* usando apenas informação fraudulenta. As abordagens baseadas em nomes planos estão fortemente ligadas às relações de confiança, já que os nomes possuem ligações com o conteúdo através de criptografia de chave pública.

A integridade dos objetos de dados do NetInf é feita através do campo autenticador, que faz a ligação direta com o publicador dos dados. "Se o campo de rótulo contiver um valor de *hash* criptográfico calculado sobre o objeto de dados, o nome também conterá uma ligação direta com o conteúdo, que fornece, assim, a auto-certificação da integridade do objeto de dados" [Ahlgren et al. 2010]. Nas CCNs, a relação com o publicador é feita através do seu prefixo único explícito em cada pacote de dados enviado. A assinatura é feita em cada pacote através do respectivo nome e estas assinaturas são padrões de chave pública. A verificação do nome com o conteúdo é feita através da respectiva chave privada. A confiança na chave de assinatura e a integridade dos dados devem ser fornecidas através de mecanismos adicionais. Por exemplo, um certificado baseado em PKI. Para o PSIRP a segurança é garantida através dos identificadores de encontro e escopo construídos a partir do *hash* do conteúdo, além de possuir um mecanismo de segurança indireto através do *hash* de chave pública adicionado ao rótulo. Mais ainda, o PSIRP dispõe da autenticação em nível de pacote (PLA – *Packet-Level Authentication*) adicionada para o encaminhamento de pacotes, a fim de combater ataques DoS [Ahlgren et al. 2010].

3.3. Roteamento

Nas CCNs o pacote de interesse é roteado por *broadcast* para uma potencial fonte que contenha os dados desejados. O publicador escolhido será o que possuir o maior prefixo correspondente ao nome especificado no pacote de interesse, daí mais uma peculiaridade dos nomes hierárquicos nesta abordagem. Os pacotes de dados seguem o caminho inverso até alcançar o solicitante. Além disso, segundo [Jacobson et al. 2009], as CCNs podem utilizar os protocolos de roteamento convencionais, já que elas são compatíveis com as redes IP. Já no NetInf e no PSIRP o roteamento é dividido em dois processos: um para o encontro do solicitante com uma potencial fonte, e outro destinado à entrega dos pacotes propriamente dita. O PSIRP possui ainda um mecanismo de encaminhamento de dados baseados em filtros de Bloom anexado aos pacotes. Já o NetInf possui o mecanismo de roteamento baseado em nomes (LLC), que integra os caminhos de resolução e recuperação, e que segundo [Ahlgren et al. 2008] pode resultar em melhor desempenho.

3.4. Aspectos de *Caching*

Um dos princípios das redes centradas na informação é que os usuários podem se beneficiar de cópias de conteúdos espalhados ao longo da rede, permitindo otimizar a disseminação global da informação. O *cache* de objetos de informação, portanto, é parte crucial de todas as três abordagens de redes centrada em informações. Na arquitetura NetInf, o conteúdo em *cache* também pode ser encontrado através de busca local ou ser encontrado através do sistema de resolução de nomes se, neste caso, existir uma cópia em *cache* neste

local. Há também a possibilidade do transporte NetInf encontrar cópias em *cache* no caminho para a localização de uma cópia do objeto. Já no PSIRP, o *cache* de conteúdo se limita ao escopo e ao ponto de encontro. Nas CCNs, o conteúdo em *cache* pode ser localizado através de um mecanismo de pesquisa local ou ao longo do caminho do pacote de interesse até a potencial fonte. Questões de *caching* são indispensáveis até quando se fala nos seus aspectos legais e contratuais. Por exemplo, quem se responsabiliza pelos conteúdos em *cache* ao longo da rede, os desenvolvedores do conteúdo ou os proprietários do *cache*? Quais são os direitos e obrigações dos usuários que fornecem o *cache* para operadoras de serviços?

3.5. Relação entre o Publicador e o Assinante

Como citado anteriormente, nas CCNs, apenas quando um nó anuncia um prefixo de determinado conteúdo ao roteador a informação é publicada. A assinatura é feita quando um nó interessado no conteúdo envia um pacote de interesse em busca de um potencial publicador. Já no NetInf, a publicação é feita ao registrar o nome de determinado objeto de informação em um sistema de resolução de nomes e a assinatura é feita a partir da solicitação do objeto publicado ao sistema de resolução de nomes. No PSIRP, o contato entre o publicador e o assinante é feito através do processo de *rendezvous*.

3.6. Encontro entre o Emissor e o Receptor

No NetInf e no PSIRP, o ponto de encontro entre o emissor e o receptor é estabelecido apenas depois que os identificadores são registrados em um sistema de resolução de nomes. Já nas CCNs, o publicador pode criar o conteúdo dinamicamente a partir de uma solicitação. Em outras palavras, um interessado pode construir um nome para determinado conteúdo que ainda não exista, e o remetente cria dados em resposta, completando o encontro.

3.7. Escalabilidade

Um novo projeto de arquitetura de rede que aspira tornar-se uma verdadeira rede mundial deve ser escalável ao extremo, permitindo que trilhões de nós e terminais sejam conectados a rede. Além disso, os projetos devem ser capazes de transportar os *exabytes* de informação mensal Internet atual e futura, e ainda ser economicamente viável [MINTS 2010]. Os roteadores das CCNs têm dois desafios ao lidarem com a escala: o gerenciamento do número de prefixos de nomes e o armazenamento de informações de estado por pacote (*Per-Packet State*). As informações de estado são necessárias ao longo do caminho fim a fim, o que representa uma desvantagem do ponto de vista da escalabilidade. Já para o PSIRP, o desafio é gerenciar a escalabilidade relacionada às estruturas de escopos, já que para estabelecer um determinado ponto de encontro, precisa-se dimensionar corretamente o escopo relacionado. Outro desafio é o gerenciamento do encaminhamento baseado no filtro de Bloom, que se torna um problema quando usado com grandes grupos de destinatários. O desafio de escalabilidade enfrentado pela arquitetura NetInf é que cada objeto de informação deve ser representado por uma entrada no sistema de resolução de nomes global.

3.8. Outros Aspectos

As abordagens de redes centradas no conteúdo devem suportar a mobilidade tanto dos objetos de informação, quanto de *hosts* e redes [Martins e Alberti 2011]. Para o NetInf e o PSIRP, a mobilidade da informação é garantida pela persistência de nomes e pelo

desacoplamento entre a identificação e a localização da informação. Enquanto a mobilidade de entidades físicas (nós e redes) é suportada por mecanismos adicionais de desacoplamento da identificação e localização de *hosts*, semelhantes ao *Mobile* IP [Perkins 2002] (acrescentando o *care-of-address*) e ao HIP [Moskowitz e Nikander 2006] (conceito de *rendezvous*), para o NetInf e PSIRP, respectivamente. Dado que a localização do objeto de informação influência semanticamente nos nomes CCNs, estas redes não desacoplam a identificação e a localização da informação em um dado domínio. O desacoplamento não acontece também para entidades físicas, como *hosts* e redes. Segundo [Jacobson et al. 2009], as CCNs endereçam conteúdo e não *hosts*, não existindo a necessidade de mapear um endereço a partir de um identificador. Neste caso, quando uma conexão é interrompida devido a um evento de mobilidade, esta conexão pode ser restabelecida logo que exista uma conectividade disponível. A Tabela 1 é uma versão modificada e ampliada da tabela contida em [Ahlgren et al. 2010] e resume as comparações feitas entre as abordagens de redes centradas no conteúdo.

Tabela 1. Resumo das comparações em Redes Centradas na Informação

	NetInf	CCN	PSIRP
Esquema de Nomeação	Um namespace plano.	Hierárquico.	Dois namespaces planos. Um para o rendezvous e outro para o transporte.
Nomes	Nomes persistentes, opacos e não agregáveis.	Nomes persistentes, legíveis e agregáveis.	Nomes persistentes, opacos e não agregáveis.
Segurança	Ligação direta do nome com a chave pública.	Ligação com o publicador através do prefixo do nome.	Hash de conteúdo + hash chave pública + autenticação a nível de pacote.
Roteamento	Dois roteamentos: um para encontro, outro para transporte + LLC.	Broadcast do pacote de interesse para fonte, dados caminho reverso.	Dois roteamentos: Um para encontro, outro para transporte + filtro de Bloom.
Caching	Localizado pelo mecanismo de resolução de nomes; pesquisa local ou transp.	Localizado por pesquisa local ou no caminho para o publicador.	Limitado ao escopo do ponto de encontro.
Relação entre o Publicador e o Assinante	Pub: Objeto de Informação As: Consulta NRS. Encontro necessita do registro prévio do nome junto ao NRS.	Pub: Prefixo de nome As: Envia o interesse. O encontro pode ser criado dinamicamente em resposta a um interesse.	Pub: Estabelece o ponto de encontro As: Estabelece contato. Encontro desenvolvido no rendezvous.
Escalabilidade	Uma entrada no NRS por objeto; agregação de publicação.	Nº de prefixos; estado de encam. por pacote.	Confusa – depende das estruturas de escopos.
Desacoplamento ID/Loc Hosts	Semelhante ao Mobile IP – Home Address e Care-of Address.	Mapeamento ID/LOC considerado desnecessário.	Semelhante ao HIP - Rendezvous.
Desacoplamento ID/LOC Informação	Desacopla – Persistência de nomes.	Não desacopla Localização influencia semanticamente no nome.	Desacopla – Persistência de nomes.

4. Conclusões

As abordagens de redes centradas na informação propõem o desenvolvimento de uma rede com o foco na informação propriamente dita em vez de *hosts*, se preocupando, sobretudo com a segurança, escalabilidade, eficiência, qualidade e suporte a comunicação em tempo real. Tais abordagens podem ser classificadas como *clean slate*, uma vez que redesenham do zero a forma como trocamos informação na Internet. Estas redes têm como principal função a interligação em grande escala dos desenvolvedores e consumidores de conteúdo aumentando a eficiência no acesso e na disseminação da informação.

O grande desafio das redes centradas na informação é estabelecer um padrão para nomeação da informação de forma consistente e eficiente, além de soluções para localização e roteamento das informações baseadas nas informações nomeadas. Assim sendo, a principal diferença entre as três abordagens é o sistema de resolução de nomes,

já que os demais desafios como roteamento, segurança e escalabilidade estão fortemente ligados ao esquema de nomeação escolhido. A arquitetura NetInf e o PSIRP utilizam o esquema de resolução de nomes na forma plana e a CCN utilizam o esquema de nomeação de forma hierárquica. Ainda, os sistemas com nomeação plana sofrem com a incompatibilidade com o DNS atual e, portanto, necessitam de um mecanismo que interopere com o DNS ou a concepção de um sistema de resolução de nomes para nomes planos que seja escalável. Em contrapartida, estes sistemas planos possuem uma ligação mais próxima com os requisitos de segurança.

Outro desafio para a utilização de nomes planos é que estes nomes são opacos e livres de semântica. Em outras palavras, um nome plano é representado por uma seqüência de *bits*, a qual não traz informações legíveis e interpretáveis a usuários. A arquitetura NetInf faz o uso de DHTs. Para nomes planos, sistemas baseados em DHTs são uma abordagem promissora [Ahlgren et al. 2010]. Entretanto, ao contrário das CCNs, estes sistemas requerem que o conteúdo seja publicado explicitamente para informar à DHT a sua localização antes que a informação possa ser recuperada.

Referências

Jacobson, V., Smetters, D. K., Thornton, J. D., Plass, M., Briggs, N., Braynard, R., Networking Named Content, ACM CoNEXT 2009, Italy.

Palo Alto Research Center Incorporated. Focus Area. Disponível em 24/06/2010 no endereço http://www.parc.com/work/focus-area/networking/.

Rothenberg, C., Verdi, L., Magalhães, F., Towards a New Generation of Information-Oriented Internetworking Architectures, ACM CoNEXT 2008, EUA.

Niebert, N., Lundgren, L e Abramowicz, H., 4WARD Deliverable 0.1, Dissemination and Exploitation Plan, 2008.

Dannewitz, C., NetInf: An Information-Centric Design for the Future Internet, 2008.

Ahlgren, B., D'ambrosio, M, Dannewitz, C., et al., Second NetInf Architecture Description. Deliverable 6.2. 2010. Último acesso em 20/04/2010 no endereço http://tools.ietf.org/html/draft-ietf-lisp-06/.

Dannewitz, C., Pentikousis, K., Rembarz, R., Renault, É., Strandberg, O., Ubillos, J., Scenarios and Research Issues for a Network of Information, MobiMedia 2008, Finland.

Ahlgren, B., D'ambrosio, M, Marsh, I., Marchisio, M., Strandberg, O., Design Considerations for a Network of Information, ACM CoNEXT 2008, EUA.

Cirani, S., Veltri, L., Implementation of a Framework for a DHT-based Distributed Location Service, SoftCOM 2008, Itália.

Zahemzky, A., Borislava, G., Rothenberg., C.E., Experimentally Driven Research in Publish/Subscribe Information Centric Internetworking, Tridentcom 2010, Alemanha.

Nikander, P., Marias, G., Towards Understanding Pure Publish/Subscribe Cryptographic Protocols, Cambridge Security Protocols Workshop (SPW 2008), 2008.

Liu, S., Bi, J., Wang, Y., A DHTs-Based Mapping System for Identifier and Locator Separation Network, First International Conference on Advances in Future Internet, 2009, EUA.

Campista, M. E., et al., Interconexão de Redes na Internet do Futuro: Desafios e Soluções. 2010. Disponível em 15 set. 2010 no endereço http://www.gta.ufrj.br/ftp/gta/TechReports/CFM10.pdf.

Minnesota Internet Traffic Studies (MINTS), The Exabyte Era, Acessado em 20 de setembro de 2010 no endereço http://www.dtc.umn.edu/mints/references.html.

Perkins, C., IP Mobility Support for IPv4, RFC3344, 2002.

Moskowitz, R., Nikander, P., Host Identity Protocol (HIP) Architecture, RFC4423, 2006.

Ain, M., et al., PSIRP Publish-Subscribe Internet Routing Paradigm Deliverable 2.2: Conceptual Architecture of PSIRP Including Subcomponent Descriptions, 2008.

Wong, W., Verdi, F., Magalhães, M. F., A Security Plane for Pub-Sub Based Content Oriented Networks, ACM CoNEXT Conference, 2008.

Martins, B. M., Alberti, A. M., Host Identification and Location Decoupling: A Comparison of Approaches, International Workshop on Telecommunications (IWT 2011), 2011, Brasil.