Information Centric Networking

André Diegues - 201206858
Fábio Teixeira - 201305725
Tópicos Avançados em Redes - CC4037
Departamento de Ciencia de Computadores
Faculdade de Ciencias da Universidade do Porto

***Abstract*—Hoje em dia, o enorme aumento de tráfego de conteúdo e dados entre utilizadores na *Internet* motivou o desenvolvimento de novas ideias de arquitetura de *Internet* que resolvam eficazmente este problema. Uma delas é a que vamos abordar neste artigo, a *Information Centric Networking* (ICN), que através de uma abordagem de pesquisa de informação nas redes permite fornecer à rede um serviço mais resiliente a falhas que cumpre as exigências atuais de distribuição de conteúdo [1]. Vamos abordar o seu funcionamento, as arquiteturas que nasceram desta ideia e estudar se a mudança de arquitetura de *Internet* é ou não viável.**

I. Introdução

O paradigma ICN foi baseado numa primeira abordagem de arquitetura *Internet* denominada de TRIAD [2], cujo principal objectivo seria facilitar e aliviar cerca de 80% do tráfego de *Internet* que servia apenas para entrega de conteúdo através de uma arquitetura *data-centric*, isto é, um utilizador pede o conteúdo ao servidor em vez de pedir ao *host* que detém esse mesmo conteúdo [1]. A TRIAD define uma nova camada de conteúdo que está implementada por *content routers* que encaminham os pedidos aos *content servers* que, de seguida, fornecem o conteúdo .

O ICN procura substituir a arquitetura atual, que é um modelo de comunicação *host-to-host*, por uma arquitetura baseada num modelo *data-centric*, tratando o conteúdo como entidade principal na arquitetura das redes como podemos observar na figura 1. Uma rede com este tipo de arquitetura ganha inúmeras vantagens em relação ao modelo *host-to-host*, nomeadamente, na distribuição de conteúdo, segurança e desenvolvimento de aplicações [3].

Existem várias arquiteturas baseadas neste paradigma. Neste artigo vamos abordar algumas destas arquiteturas, nomeadamente as que ganharam mais apoio ao longo do tempo:

- *Data-Object Network Architecture* (DONA)
- *Publish-Subscribe Internet Technology* (PURSUIT)
- *Named Data Networking* (NDN) baseada em *Content-Centric Networking* (CCN)

II. Como funciona o ICN?

O ICN utiliza o modelo de comunicação *Publish/Subscribe*. Neste modelo, os emissores, chamados *Publishers*, em vez de enviarem as mensagens diretamente aos recetores específicos,

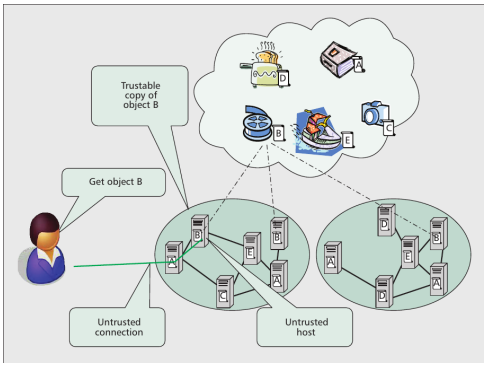

Fig. 1. Modelo de comunicação ICN. Retirado de [1]

caracterizam as mensagens publicadas em classes sem conhecimento de quem vão ser os subscritores. Por outro lado, os subscritores, chamados de *Subscribers*, demonstram interesse numa ou mais classes e apenas recebem mensagens da(s) classe(s) que subscreveram, também sem conhecimento de quem as enviou.

Neste novo paradigma, os dados tornam-se independentes de localização, memória e meio de transporte, permitindo o armazenamento e replicação em *cache*. Algo que impulsiona uma melhoria na eficiência, na escalabilidade e na robustez de comunicação em cenários difíceis. Existem várias abordagens do ICN, todas focadas num novo desenho da arquitetura da Internet atual, com o objetivo de substituir o atual modelo centrado nos *hosts*, para implementar um modelo mais orientado a dados e centrado nos conteúdos [4]. A saber:

A. DONA

A resolução de nomes no DONA [5] é da responsabilidade de uns servidores especializados chamados *Resolution Handlers* (RHs). Existe pelo menos um RH em cada sistema autónomo. Estes *Handlers* estão interconectados, formando um serviço hierárquico de resolução de nomes por cima das relações interdomínio de *routing* existentes, para que seja possível conciliar resolução de nomes com *routing* de informação no mesmo sistema. O *Publisher* envia uma mensagem *REGISTER* com o nome do objeto para o seu RH local, que guarda um apontador para o *Publisher*. O RH depois propaga o *REGISTER* para os RHs que tem ligação,

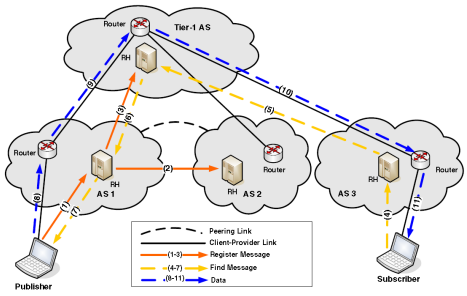

Fig. 2. Arquitetura DONA. Retirado de [4]

seguindo as rotas estabelecidas, guardando estes RHs o mapeamento entre o nome do objeto e o endereço do RH que encaminhou o registo.

Com isto, os *REGISTERs* são replicados nos RHs até aos *tier-1*, e já que estes estão conectados com os outros todos, o RH que está localizado nesse *tier* está ciente do que se passa em toda a rede.Para localizar um item, o *Subscriber* envia uma mensagem *FIND* para o RH local, que também propaga esta mensagem aos RHs parentes, até que é feito um *match* com o *tier-1*. Depois seguem-se os apontadores para encontrar o *Publisher*, já que o *tier-1* é conhecedor de toda a estrutura da rede. A figura 2 ilustra como funciona esta arquitetura.

B. PURSUIT

Esta arquitetura substitui completamente a *stack* do protocolo IP por uma *stack* do protocolo *publish-subscribe*. Consiste em três funções distintas: *rendezvous*, *topology management* e *forwarding*. Quando o *rendezvous* encontra a subscrição de uma publicação, direciona a função de gestão da topologia para criar uma rota entre o *Publisher* e o *Subscriber*. Esta rota é usada pela função de encaminhamento para realizar a transferência de dados.

A resolução de nomes no *PURSUIT* é tratada pela função *rendezvous*, a qual é implementada por uma coleção de *Rendezvous Nodes* (RNs), pertencentes à *Rendezvous Network* (RENE). Quando um *Publisher* quer anunciar um objeto de informação, emite uma mensagem *PUBLISH* para o RN local, o qual é encaminhado por tabelas de *hash* distribuídas para o RN atribuído com o ID do *scope* correspondente. Quando o *Subscriber* emite um *SUBSCRIBE* para o mesmo objeto de informação ao seu RN local, é encaminhado pela tabela de *hash* para o mesmo RN. Depois, o RN informa um nó da *Topology Management* (TM) para criar uma rota ligando o *Publisher* ao *Subscriber* para entrega de dados. A gestão da topologia envia a rota ao *Publisher* numa mensagem *START PUBLISH*, que já utiliza a rota para enviar o objeto de informação via *Forwarding Nodes* (FNs), vejamos o examplo da figura 3.

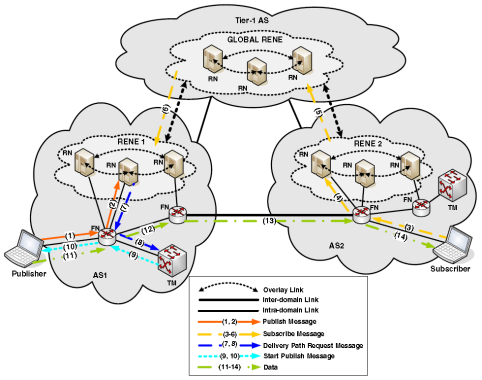

Fig. 3. Arquitetura PURSUIT. Retirado de [4]

C. NDN

Os *Subscribers* emitem mensagens *INTEREST* para pedir informação sobre objetos que chegam na forma de dados. As mensagens são encaminhadas por *Content Routers* (CRs), e cada um dos CRs mantém três estruturas de dados: *Forwarding Information Base* (FIB), *Pending Interest Table* (PIT) e *Content Store* (CS).

O FIB mapeia informação para as interfaces de saída que devem ser usadas para encaminhar as mensagens *INTEREST*. O PIT segue as interfaces de entrada nas quais as mensagens de *INTEREST* chegou. Já o CS funciona como *cache* local para objetos de informação que passaram pelo CR.Quando um *INTEREST* chega, o CR extrai a informação do nome e procurar por um objeto nesse CS cujo nome coincida com o prefixo pretendido. Se for bem-sucedido, é imediatamente enviado através da interface de entrada numa mensagem *DATA* e o *INTEREST* é descartado. Senão, o *router* executa uma procura do prefixo mais longo no seu FIB, para decidir a direção do encaminhamento. Se a entrada for encontrada no FIB, o router regista a interface de entrada do *INTEREST* no PIT e empurra o *INTEREST* para o CR indicado pelo FIB. O exemplo da figura 4 permite observar como é feita a transmissão de conteúdo.

D. Tráfego da Internet e métricas usadas no ICN

Analisar o tráfego da *Internet* é importante para perceber que modelo de funcionamento deve ser adotado. Vamos abordar o Labovitz10[1], um estudo realizado a larga escala, com o propósito de estudar o tráfego dos *links* na rede da *Internet*. O resultado indica um domínio claro do tráfego *Web*, correspondente a mais de 30% do tráfego total. No entanto, técnicas de *Deep Packet Inspection* (DPI) revelam que entre 13 a 21 porcento do tráfego através do protocolo HTTP se deve a tráfego de vídeos. Já a percentagem de tráfego *Peer-to-Peer* (P2P) situa-se entre os 17% e os 19% como podemos ver na figura 5.

[1]https://tools.ietf.org/html/rfc7945#ref-Labovitz10

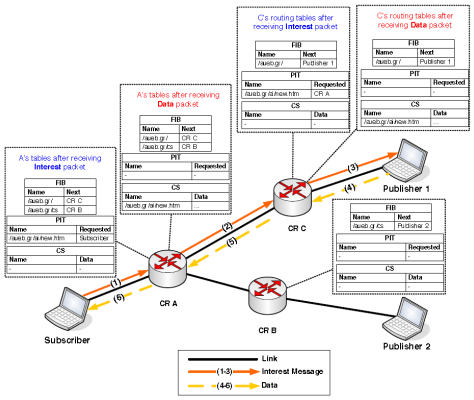

Fig. 4. Arquitetura NDN. Retirado de [4]

Traffic Type	Ratio
Web	31-39%
P2P	17-19%
Video	13-21%
Other	29-31%

Fig. 5. Percentagem de cada tipo de tráfego na *Internet*

A *Internet Engineering Task Force*[2] (IETF) tem trabalhado há mais de uma década em desenvolver métricas e métodos para medir a *performance* das redes IP. O trabalho foi realizado em grande parte dentro das métricas do *IP Performance Metrics* (IPPM).

As métricas do IPPM incluem atrasos, a variação dos mesmos, perdas, reordenações e duplicação. Enquanto o trabalho do IPPM for baseado em redes IP *packet-switched*, isto é, um método de comunicação de rede digital que agrupa todos os dados transmitidos em blocos de tamanho adequado, chamados de pacotes, que são transmitidos através de um meio que pode ser compartilhado por várias sessões de comunicação simultneas, é concebível que possa ser modificado e estendido para também cobrir as redes do ICN.

No entanto são necessários mais estudos para transformar essa afirmação em certeza. Muitos especialistas têm trabalhado na reinventação e refinação das métricas e métodos IPPM, algo que pode ser benéfico para medir o desempenho do ICN. Como estas métricas já trabalham em redes *host-centric*, em comparação com o ICN implicaria apenas adicionar a extensão do mesmo ao quadro IPPM. Um benefício importante de medir o desempenho de transporte de uma rede no seu *output*, usando métricas baseadas na qualidade de serviço (QoS), é que tal pode ser conseguido, em grande parte, sem qualquer dependência de aplicações. Outra opção para medir a *performance* de transporte seria usar métricas baseadas em QoS, não ao nível do *output*, mas sim ao nível do *input* da aplicação.

Para um *streaming* de um vídeo ao vivo, as métricas utilizadas seriam a latência de início, o atraso de *playout* e a eficiência na continuidade. Os benefícios desta abordagem é a abstração dos detalhes do transporte de rede, o que pode ser benéfico quando comparado com diferentes conceitos e tipos de rede. A desvantagem desta abordagem é a sua dependência em relação às aplicações, sendo por isso provável que os diferentes tipos de aplicações exijam diferentes métricas. Pode ser possível identificar métricas-padrão para cada tipo de aplicação, porém é mais claro utilizar métricas IPPM.

III. Quais os obstáculos e custos de implementação o ICN?

A. Vantagens vs Custo de Implementação

Nesta secção vamos abordar se as vantagens que o ICN traz são fortes o suficiente para motivar uma mudança de arquitetura na *Internet*.

Como sabemos, o tráfego para entrega de conteúdo tem vindo a aumentar significativamente ao longo dos últimos 5 a 10 anos, de tal modo que foram desenvolvidas soluções, a curto prazo, para lidar com esta questão, nomeadamente, as *Content Distribution Networks* (CDNs) [6] e o *Peer-to-Peer Networking* (P2P) [7], que representam um novo modelo de comunicação baseado no conteúdo que é transportado.

Assim sendo, haverá a necessidade de alterar toda a arquitetura da *Internet* por uma arquitetura baseada em ICN? A resposta não é imediatamente sim, porém não só o crescimento de tráfego é um dado que vai pesar bastante na resposta, pois significa que a atual arquitetura não representa corretamente a sua utilização, como as CDNs e o P2P não são soluções perfeitas ao problema, o que poderá levar a que se considere de forma mais séria a abordagem ICN.

O que podemos ganhar com uma arquitetura baseada em ICN [1]:

1) **Distribuição de conteúdo eficiente e escalável**

   Segundo [9], esta vantagem por si só não é suficiente para motivar a mudança de arquitetura mas, como já referimos acima, se a quantidade de tráfego e replicação de conteúdo continuar a aumentar o ICN é a melhor

[2]https://tools.ietf.org/html/rfc2330

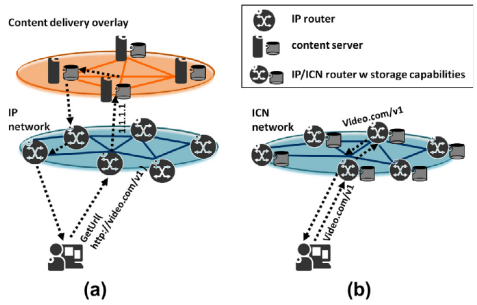

Fig. 6. (a) CDN vs (b) ICN. Retirado de [8]

estratégia a adotar.

2) ***Naming* único e persistente**

Os *Uniform Resource Identifications* (URIs) são localizadores de objetos que exibem o endereço IP de um servidor *web* que responde a pedidos resolvendo a parte local de um URI, logo, se um objeto for movido o seu *site* muda de domínio ou o *site* fica *unreachable* quebrando o *naming* do objeto.
Outro problema é se existirem cópias do mesmo objeto em diferentes servidores *web* essas cópias vão ter diferentes URIs, o que é um desperdício visto que se trata do mesmo objeto.

Com o ICN, isto já não acontece pois é possível atribuir um nome único e persistente aos NDOs e com o modelo de serviço que separa produtores de consumidores.

3) **Modelo de Segurança**

A segurança nas redes, hoje em dia, protege a comunicação entre dois utilizadores, na maior parte das vezes entre cliente e servidor, usando protocolos criptográficos. Este modelo exige que o cliente confie a entrega da sua informação ao servidor.

O modelo de segurança do ICN fornece integridade de nome-dados e verificação da origem dos NDOs. Garante também integridade e autenticidade através *caching* ubíquo.

4) **Mobilidade e *Multihoming***

As redes, como as conhecemos hoje em dia, têm problemas de gestão das ligações *end-to-end* e na escolha da rota destas ligações.

O ICN não tem estes problemas pois não necessita desta gestão de ligações. O que acontece é que o cliente continua a fazer *requests* de um novo acesso que possivelmente é servido por outro servidor em vez de ter de manter uma ligação do servidor anterior. Da mesma forma, um cliente *multi-homed* pode escolher enviar *requests* a um ou mais acessos.

5) **Tolerância a ruturas**

Comunicação *end-to-end* com transporte de sessões para servidores de origem é muito dificultada em redes de conetividade esparsa, mobilidade veloz e com ruturas.

Se o principal objectivo é aceder a conteúdo, o ICN oferece armazenamento com *in-network caching* para transporte *hop-by-hop*. Este mecanismo dá garantias de *performance* e fiabilidade.

Migrar da atual estrutura da *Internet* para o ICN só será possível utilizando uma topologia de rede realista. Estas podem ser inferidas através de rastreios de *Internet*, como a CAIDA Macroscopic Internet Topology Data Kit[3] ou a Rocketfuel[4].

No entanto, existem problemas associados como o tamanho da topologia (aproximadamente 45 mil sistemas autónomos e perto de 200 mil *links*), o que pode limitar a escalabilidade da ferramenta de avaliação. Tal como nas topologias *host-centric*, definir apenas um grafo de nós não vai ser suficiente. preciso também definir e listar as respetivas matrizes a que correspondem a rede e a *storage* e as capacidades computacionais disponíveis para cada nó, tal como as características (e atrasos) de cada *link*[5]. Valores aproximados podem ser estimados a partir de plataformas como o iPlane [6].

B. Desafios

O ICN ainda está longe de estar pronto a ser implementado. Os *researchers* têm lidado com vários desafios para tornarem este novo modelo disponível para poder ser colocado ao serviço dos utilizadores. Entre os desafios que têm encontrado [10] destacam-se:

1) ***Naming* dos dados**

Dar nomes aos dados é tão importante para o ICN como dar nomes aos *hosts* é para a Internet de hoje. Sendo assim, o ICN requer nomes únicos para os *Named Data Objects* (NDOs), já que os nomes são utilizados para identificar objetos independentemente da sua localização ou conteúdo. Usar tabelas de *hash* é uma forma possível de resolver este imbróglio, já que permite depois que se possa comparar isso com o próprio nome do componente.

[3]http://www.caida.org/data/active/internet-topology-data-kit
[4]http://www.cs.washington.edu/research/networking/rocketfuel
[5]https://tools.ietf.org/html/rfc7945#ref-Montage
[6]http://iplane.cs.washington.edu/

2) **Proteção da privacidade**

Como a rede pode ver quem faz o pedido pela informação e já que a tendência do ICN é guardar a história dos utilizadores, torna-se um problema para o utilizador não ter garantias de privacidade, já que como os nomes devem ter um longo tempo de uso, seria uma limitação desperdiçar nomes.

3) **Atualizações**

Se um NDO pode ser replicado e guardado na rede para futuros tratamentos, os nomes têm de ter um prazo de validade longa e o conteúdo do nome não deve ser alterado, o que impossibilita a atualização de objetos.

4) **Integridade dos dados**

A verificação da integridade dos dados é um passo importante para a consolidação do ICN. O facto de os NDOs não só serem recuperados a partir da copa original como também a partir de qualquer ponto da rede em que estejam guardados em cache e de poderem ser modificados faz com que não se possa confiar a 100% na integridade dos dados. Utilizar uma *hash* como parte do nome de objeto é também uma possível solução deste problema, embora a utilização de chaves criptográficas seja melhor aplicada nestes casos.

5) ***Encryption***

É possível cifrar NDOs no ICN e apenas os consumidores que tiverem as chaves podem aceder a esse conteúdo privado. No entanto distribuir e gerir estas chaves, tal como fornecer as interfaces para os utilizadores é ainda uma matéria de estudo.

6) **Agregação e filtragem de tráfego**

Uma mensagem de um pedido *request* para receber um objeto de dados pode agregar vários pedidos de vários consumidores. Esta agregação reduz o tráfego na rede, mas torna a filtragem mais difícil. O desafio neste caso é fornecer um mecanismo que pretenda agregação, mas ao mesmo tempo uma pré-filtragem dos *request* dos utilizadores. Uma possível solução é indicar o conjunto de utilizadores que fizeram o *request* nesse *request* agregado, permitindo assim gerar numa resposta apenas o subconjunto dos utilizadores que fizeram *request* e têm acesso aos dados. No entanto, esta solução requer a utilização de outros nós na rede e não permite fazer *caching*.

7) **Roteamento pelo nome**

Uma vez que o número de objetos de dados tem tendência a aumentar, o tamanho das tabelas de encaminhamento é um problema a pensar, pois pode ser proporcional ao número de objetos de dados, a menos que seja introduzido um mecanismo de agregação. Por outro lado, o *Route-By-Name Routing* (RBNR) reduz a latência e simplifica o processo de roteamento devido à omissão do processo de resolução.

C. Testes de trabalhos anteriores

Na secção anterior vimos que o *caching* poderia manter os dados continuamente disponíveis na rede. Alguns *researchers* investigaram o assunto realizando testes de *performance* sobre as várias formas de fazer *caching*.

IV. ICN: O FUTURO DA *Internet*

Os utilizadores estão cada vez mais interessados em receber conteúdos, seja qual for a sua origem, do que ter de aceder a um servidor para receber essa informação. E o facto de a *Internet* ainda ser centrada nos *hosts* implica que o utilizador tenha de especificar em cada pedido não só a informação que deseja receber, como também especificar o servidor do qual a informação pode ser retirada. Com o ICN isso já não acontece [4].

O pressuposto básico por trás do ICN é que a informação é nomeada, endereçada e encontra o conteúdo independentemente da sua localização. Uma implicação indireta da implementação do ICN é que a informação se torna orientada para o recetor, em contraste com a atual realidade da Internet em que os emissores têm controlo total sobre os dados trocados [11]. No ICN só são recebidos os dados que o recetor tenha pedido. Depois de ser enviado o pedido, a rede é responsável por localizar a melhor origem para fornecer a informação desejada ao recetor.

Quando um elemento da rede receber um pedido por conteúdo, pode ter duas ações: se estiver em *cache*, responde imediatamente com o conteúdo; se não estiver, faz um pedido aos elementos com os quais tem ligação e depois guardar em *cache* o conteúdo quando for encontrado [4].

Nesse sentido, e já que os conteúdos chegam de elementos da rede ao invés da origem, o desenho do ICN tem de garantir a segurança dos conteúdos, contrariamente à estrutura atual da Internet que se foca no caminho. Para isso, quem fornece os dados assina um modelo de segurança para que os elementos da rede e os consumidores apenas tenham de verificar essa assinatura para garantir a sua fiabilidade [9].

V. CONCLUSÃO

The conclusion goes here.

REFERENCES

[1] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 26–36, July 2012.

[2] D. Cheriton and M. Gritter, "Triad: A new next-generation internet architecture," 2000.

[3] "Guest editorial [information centric networking]," *China Communications*, vol. 12, no. 7, pp. iii–iv, July 2015.

[4] G. Xylomenos, C. N. Ververidis, V. A. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. V. Katsaros, and G. C. Polyzos, "A survey of information-centric networking research," *IEEE Communications Surveys Tutorials*, vol. 16, no. 2, pp. 1024–1049, Second 2014.

[5] *A Data-Oriented (and Beyond) Network Architecture*, vol. 37. Kyoto, Japan: ACM, 08/2007 2007.

[6] D. Farber, R. Greer, A. Swart, and J. Balter, "Internet content delivery network," Nov. 25 2003, uS Patent 6,654,807. [Online]. Available: https://www.google.com/patents/US6654807

[7] O. A. Skaflestad and N. Kaurel, "Peer-to-peer networking."

[8] G. Carofiglio, G. Morabito, L. Muscariello, I. Solis, and M. Varvello, "From content delivery today to information centric networking," *Comput. Netw.*, vol. 57, no. 16, pp. 3116–3127, Nov. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2013.07.002

[9] A. Ghodsi, S. Shenker, T. Koponen, A. Singla, B. Raghavan, and J. Wilcox, "Information-centric networking: Seeing the forest for the trees," in *Proceedings of the 10th ACM Workshop on Hot Topics in Networks*, ser. HotNets-X. New York, NY, USA: ACM, 2011, pp. 1:1–1:6. [Online]. Available: http://doi.acm.org/10.1145/2070562.2070563

[10] D. Saucez, T. Schmidt, D. Kutscher, S. Eum, I. Psaras, D. Corujo, and K. Pentikousis, "Information-centric networking (icn) research challenges," 2016.

[11] S. Arianfar, P. Nikander, and J. Ott, "On content-centric router design and implications," in *Proceedings of the Re-Architecting the Internet Workshop*. ACM, 2010, p. 5.