Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 18, 2014

D. Kutscher, Ed.
NEC
S. Eum
NICT
K. Pentikousis
EICT
I. Psaras
UCL
D. Corujo
Universidade de Aveiro
D. Saucez
INRIA
T. Schmidt
HAW HAmburg
M. Waehlisch
FU Berlin
February 14, 2014

ICN Research Challenges
draft-kutscher-icnrg-challenges-02

Abstract

This memo describes research challenges for Information-Centric Networking. Information-Centric Networking is an approach to evolve the Internet infrastructure to directly support information distribution by introducing uniquely named data as a core Internet principle. Data becomes independent from location, application, storage, and means of transportation, enabling in-network caching and replication. Challenges include naming, security, routing, system scalability, mobility management, wireless networking, transport services, in-network caching, and network management.

This Internet-Draft will expire on August 18, 2014.

Table of Contents

1. Introduction

Distributing and manipulating named information is a major application in the Internet today. In addition to web-based content distribution, other distribution technologies (such as P2P and CDN) have emerged and are promoting a communication model of accessing data by name, regardless of origin server location.

In order to respond to increasing traffic volume in the current Internet for applications such as mobile video and cloud computing, a set of disparate technologies and distribution services are applied that employ caching, replication and content distribution in different specific ways. These approaches are currently deployed in separate silos -- different CDN providers and P2P applications rely on their own, often proprietary, distribution technologies. It is not possible to uniquely and securely identify named information independently of the distribution channel; and the different distribution approaches are typically implemented as an overlay, potentially leading to unnecessary inefficiency.

For example, creating and sharing multimedia content in a social networking application today, typically requires uploading data objects to centralized service provider platforms, from where it can be accessed individually by other users. Even if content sharing is intended to happen locally, e.g., in a local network or local area, the actual communication will require interactions from any interested user with the service provider. CDNs can alleviate the situation only partly, because, due to organizational and economic reasons, it is not common to deploy CDN gear ubiquitously. Moreover, since CDNs and the respective communication sessions form overlays, the actual communication, i.e., the requests for named content and the actual responses, are largely invisible to the network, i.e., it is not easily possible to optimize efficiency and performance. For example in a wireless access network, it is not possible to leverage the inherent broadcast nature of the medium (and avoid duplicate transmission of the same content) due to limitations from point-to-point and overlay communication.

Information-centric networking (ICN) is an approach to evolve the Internet infrastructure to directly support this use by introducing named data as a core network-layer primitive. Data objects become independent of location, application, storage, and means of transportation, allowing for inexpensive and ubiquitous in-network caching and replication. The expected benefits are improved efficiency, better support for provenance verification and name-content binding validation, better scalability with respect to information/bandwidth demand and better robustness in challenging communication scenarios.

ICN concepts can be deployed by retooling the protocol stack: name-based data access can be implemented on top of the existing IP infrastructure, e.g., by providing resource naming, ubiquitous caching and corresponding transport services, or it can be seen as a packet-level internetworking technology that would cause fundamental changes to Internet routing and forwarding. In summary, ICN is expected to evolve the Internet architecture towards a network model that is more suitable for the current and future use patterns.

This document presents the ICN research challenges that need to be addressed in order to achieve these goals. These research challenges are seen from a technical perspective, although business relationships between Internet players will also influence developments in this area. We leave business challenges for a separate document, however. The objective of this note is to document the technical challenges and corresponding current approaches and to expose requirements that should be addressed by future research work.

2. Problems with Information Distribution Today

The best current practice to manage the above-mentioned growth in terms of data volume and number of devices is to increase infrastructure investment, employ application-layer overlays such as CDNs, P2P applications, and M2M application platforms that cache content, provide location-independent access to data, and optimize its delivery. In principle, such platforms provide a service model of accessing named data objects (NDOs) (e.g., replicated web resources, M2M data in data centers) instead of a host-to-host packet delivery service model.

However, since this functionality resides in overlays only, the full potential of content distribution and M2M application platforms cannot be leveraged as the network is not aware of data requests and data transmissions. This has the following impact:

- data traffic typically follows sub-optimal paths as it is effectively routed depending on the overlay topology instead of the Internet layer topology;

- network capabilities, such as multicast and broadcast, are largely underutilized or not employed at all. As a result, request and delivery for the same object have to be made multiple times;

- overlays typically require significant infrastructure support, e.g., authentication portals, content storage, and applications servers, making it often impossible to establish local, direct

communication;

- since the network is not aware of the nature of data objects, it is unable to manage access and transmission (without layer violations);

- provenance validation uses host authentication today. As such, even if there are locally cached copies available, it is normally not easily possible to validate their authenticity; and

- many applications follow their own approach to caching, replication, transport, authenticity validation (if at all), although they all share similar models for accessing named data objects in the network.

3. ICN Terminology and Concepts

3.1. Terminology

Information-Centric Networking (ICN) A concept for communicating in a network that provides accessing named data objects as a first order service. See Section 3.2 for details.

Named Data Object (NDO): Addressable data unit in an ICN that can represent a collection of bytes or a piece of information. In ICN, each data object has a name bound to it, and there are typically mechanisms to secure (and validate) this binding. Different ICN approaches have different concepts for how to map NDOs to individual units of transport. Within the context of this document, an NDO is any named object that can be requested from the network.

Requestor: Entitiy in an ICN network that is sending a request for a Named Data Object to the network.

3.2. Concepts

Fundamentally, ICN is providing access to named data as a first-order network service, i.e., the network is able to serve requests to named data natively. That means, network nodes can receive requests for named data and act as necessary, for example, by forwarding the request to a suitable next-hop. Consequently, the network processes requests for named data objects (and corresponding responses) natively. Every network node on a path is enabled to perform forwarding decisions, to cache objects etc. This enables the network to forward such requests on optimal paths, employing the best transmission technologies at every node, e.g., broadcast/multicast

transmission in wireless networks to avoid duplicate transmission of both requests and responses.

In ICN there is a set of common concepts and node requirements beyond this basic service model. Naming data objects is a key concept. In general, ICN names represent neither network nodes nor interfaces -- they represent NDOs independently of their location. Names do play a key role in forwarding decisions and are used for matching requests to responses: In order to provide better support for accessing copies of NDOs regardless of their location, it is important to be able to validate that a response actually delivers the bits that correspond to an original request for named data.

Name-content binding validation is a fundamental security service in ICN, and this is often achieved by establishing a verifiable binding between the object name and the actual object or an identity that has created the object. ICN can support other security services, such as provenance validation, encryption -- depending on the details of naming schemes, object models and assumptions on infrastructure support. Security services such as name-content binding validation are available to any node, i.e., not just the actual receivers. This is an important feature, for enabling ingress gateways to check object authenticity to prevent denial-of-service attacks.

Based on these fundamental properties it is possible to leverage network storage ubiquitously: every node and every device can cache data objects and respond to requests for such objects -- it is not required to validate the authenticity of the node itself since name-content bindings can be validated. Ubiquitous in-network storage can be used for different purposes: it can enable sharing, i.e., the same object copy can be delivered to multiple users/nodes as in today's proxy caches and CDNs. It can also be used to make communication more robust (and perform better) by enabling the network to answer requests from local caches (instead of from origin servers). In case of disruption (message not delivered), a node can re-send the request, and it could be answered by an on-path cache, i.e., on the other side of the disrupted link. The network itself would thus support retransmissions -- enabling shorter round-trip times and offloading origin servers and other parts of the network.

The request/response model and ubiquitous in-network storage also enables new options for implementing transport services, i.e., reliable transmission, flow control, etc. First of all, a request/response model can enable receiver-driven transport regimes, i.e., receivers (the requestors of NDOs) can control message sending rates by regulating the request sending rate (assuming that every response message has to be triggered by a request message). Retransmission would be achieved by re-sending requests, e.g., after a timeout.

Because objects can be replicated, object transmission and transport sessions would not necessarily have end-to-end semantics: requests can be answered by caches, and a node can select one or multiple next-hop destination for a particular request -- depending on configuration, observed performance or other criteria.

This receiver-driven communication model potentially enables new interconnection and business models: a request for named data can be linked to an interest of a requestor (or requesting network) in data from another peer, which could suggest modeling peering agreements and charging accordingly.

4. ICN Research Challenges

4.1. Naming and data authenticity

Naming data objects is as important for ICN as naming hosts is for today's Internet. Fundamentally, ICN requires unique names for individual NDOs, since names are used for identifying objects independently of their location or container. In addition, since NDOs can be cached anywhere, the origin cannot be trusted anymore hence the importance to establish a verifiable binding between the object and its name (name-data integrity) so that a receiver can be sure that the received bits do correspond to the NDO originally requested (object authenticity). Information about an object's provenance, i.e., who generated or published it, is also useful to associate to the name.

The above functions are fundamentally required for the information-centric network to work reliably, otherwise neither network elements nor receivers can trust object authenticity. Lack of this trust enables several attacks including DoS attacks by injecting spoofed content into the network. There are different ways to use names and cryptography to achieve the desired functions [ICNNAMING] [ICNSURVEY], and there are different ways to manage namespaces correspondingly.

Two types of naming schemes have been proposed in the ICN literature: hierarchical and flat namespaces. For example, a hierarchical scheme may have a structure similar to current URIs, where the hierarchy is rooted in a publisher prefix. Such hierarchy enables aggregation of routing information, improving scalability of the routing system. In some cases, names are human-readable, which makes it possible for users to manually type in names, reuse, and, to some extent, map the name to user intent.

The second general class of naming schemes follows a "self-

certifying" approach, meaning that the object's name-data integrity can be verified without requiring a public key infrastructure (PKI) or other third party to first establish trust in the key. Self-certification is achieved, e.g., by binding the hash of the NDO content to the object's name. For instance, this can be done by directly embedding the hash of the content in the name. Another option is an indirect binding, which embeds the public key of the publisher in the name and signs the hash of the content with the corresponding secret key. The resulting names are typically non-hierarchical, or flat, although the publisher field could be employed to create a structure which could facilitate route aggregation. There are several design trade-offs for ICN naming, which affect routing and security. Self-certifying names are not human readable nor hierarchical. They can however provide some structure for aggregation, for instance, a name part corresponding to a publisher.

Research challenges specific to naming include:

- naming static data objects can be performed by using content hashes as part of object names, so that publishers calculate the hash over existing data objects and receivers (or any ICN node) can validate the name-content binding by re-calculating the hash and comparing it to the name (component). [RFC6920] specifies a concrete naming format for this.

- naming dynamic objects refers to use cases where the name has to be generated before the object is created. For example, this could be the case for live streaming, when a publisher wants to make the stream available by registering stream chunk names in the network. One approach to this can be self-certified names as described above.

- requestor privacy protection can be a challenge in ICN as a direct consequence of the accessing-named-data-objects paradigm: if the network can "see" requests and responses, it can also log request history for network segments or individual users, which can be undesirable, especially since names are typically expected to be long-lived. That is, even if the name itself does not reveal much information, the assumption is that the name can be used to retrieve the corresponding data objects in the future.

- Updating and versioning NDO can be challenging because it can contradict fundamental ICN assumptions: if an NDO can be replicated and stored in in-network storage for later retrieval, names have to be long-lived -- and the name-content binding must not change: updating an object (i.e., changing the content without generating a new name) is impossible. Versioning is one possible solution, but requires a naming scheme that supports it (and a way

for requestors to learn about versions).

- Managing accessibility: whereas in ICN the general assumption is to enable ubiquitous access to NDOs, there can be relevant use cases where access to objects should be restricted, for example to a specific user group. There are different approaches for this, such as object encryption (requiring key distribution and related mechanisms) or the concept of scopes, e.g., based on names that can only be used/resolved under some constraints.

4.2. Security

Security can take many different forms in ICN and instead of discussing specific attacks or technical details, we propose here the most important security challenges that come from the shift to information-centric communications. Some challenges are well-understood, and there are (sometimes multiple different) approaches to address them, whereas other challenges are active research and engineering topics.

4.2.1. Data Object Authentication

As mentioned in section Section 4.1, data object authentication is an important ICN feature, since ICN data objects are retrieved not only from an original copy holder but also from any caching point. Hence, the communication channel endpoints to retrieve NDOs are not trustable anymore and solutions widely used today such as TLS [RFC5246] cannot be used as a general solution. Since data objects can be maliciously modified ICN should provide users with a security mechanism to verify the origin and integrity of the data object, and there are different ways to achieve this.

An efficient approach for static NDOs is providing a name-content-binding by hashing an NDO and using the hash as a part of the object's name. [RFC6920] provides a mechanism and a format for representing a digest algorithm and the actual digest in a name (amongst other information).

For dynamic objects (where it is desirable to refer to an NDO by name before the object has been created), public-key cryptography is often applied, i.e., every NDO would be authenticated by means of a signature performed by the data object publisher so that any data object consumer can verify the validity of the data object based on the signature. However, in order to verify the signature of an object, the consumer must know the public key of the entity that signed the object.

One research challenge is then to support a mechanism to distribute

the publisher's public keys to the consumers of data objects. There are two main approaches to achieve this; one is based on an external third party authority such as hierarchical Public Key Infrastructure (PKI) [RFC5280] and the other is to adapt a self-certifying scheme. The former, as the name implies, depends on an external third party authority to distribute the public key of the publisher for the consumers. In a self-certifying scheme, the public key (or a hash of it) would be used as part of the name -- which is sufficient to validate the object's authenticity.

In cases where information about the origin of a data object is not available by other means, the object itself would have to incorporate the necessary information to determine the object publisher, for example with a certificate, that can be validated through the PKI. Once the certificate is authenticated, its public key can be used to authenticate the signed data object itself.

4.2.2. Binding NDOs to Real-World Identities

In addition to validating NDO authenticity, it is still important to bind real-world identities, e.g., a publisher identity, to objects, so that a requestor can verify that a received object was actually published by a certain source.

With hash-based and self-certifying names, real-world-identity bindings are not intrinsically established: the name provides the hash of the NDO or of the public key that has been used to sign the NDO. There needs to be another binding to a real-world-identity if that feature is requested.

If the object name directly provides the publisher name and if that name is protected by a certificate that links to PKI-like trust chain, the object name itself can provide an intrinsic binding to a real-world identity.

Binding between NDOs and Real-World Identities is essential but there is no universal way to achieve it as it is all intrinsic to the particular ICN approach.

4.2.3. Traffic aggregation and filtering

One request message to retrieve a data object can actually aggregate requests coming from several consumers. This aggregation of requests reduces the overall traffic but makes per-requestor filtering harder. The challenge in this case is to provide a mechanism that allows requests aggregation and per-requestor filtering. A possible solution is to indicate the set of requestors in the aggregated request such that the response can indicate the subset of requestors

allowed to access the data object. However, this solution requires collaboration from other nodes in the network and is not suitable for caching. Another possible solution is the encrypt data objects and ensure that only authorised consumers can decrypt them. This solution does not preclude caching and does not require collaboration from the network. However, it implies a mechanism to generate group keys (e.g., different private keys can be used to decrypt the same encrypted data object) [Chaum].

4.2.4. State overloading

ICN solutions that implement state on intermediate routers for request routing or forwarding (e.g., CCN [CCN]) are subject to denial of service attacks from overloading or superseding the internal state of a router (e.g., 'interest flooding' [BACKSCATTER]). Additionally, stateful forwarding can enable attack vectors such as resource exhaustion or complexity attacks to the routing infrastructure. The challenge is then to provision routers and construct internal state in a way that alleviates sensibility to such attacks. The problem becomes even harder, if the protocol does not provide information about the origin of messages. Without origin, it is a particular challenge to distinguish between regular (intense) use and misuse of the infrastructure.

4.2.5. Delivering data objects from replicas

A common capability of ICN solutions is data replication and in-network storage. Delivering replicated data objects from caches decouples content consumption from data sources which leads to a loss of control on (1) content access, and (2) content dissemination. In a widely distributed, decentralized environment like the Internet, this raises several challenges.

One group of challenges is related to content management. Without access control, a content provider looses the means to count and survey content consumption, to limit access scopes, to control or know about the number of copies of its data in the network, or to withdraw publication reliably. Any non-cooperative or desynchronized data cache may hinder an effective content management policy.

Another group of challenges arises from potential traffic amplifications in the decoupled environment. ICN solutions that attempt to retrieve content from several replicas in parallel, or decorrelated network routing states, but also distributed attackers may simultaneously initiate the transmission of content from multiple replicas towards the same destination (e.g., 'initiated overloads' or 'blockades' [BACKSCATTER]). Methods for mitigating such threats need rigorous forwarding checks that require alignment with caching

procedures (e.g., on-path or off-path).

4.2.6. Cryptographic robustness

Content producers sign their content to ensure the integrity of data and to allow for data object authentication. This is a fundamental requirement in ICN due to distributed caching. Publishers, who (a) massively sign content, which is (b) long-lived, offer time and data to an attacker for comprising cryptographic credentials. Signing large amount of data eases common attacks that try to breach the key of the publisher. Based on this observation, the following research challenges appear. To which extent does the content publication model conflict with cryptographic limitations? How can we achieve a transparent re-signing without introducing additional cryptographical weaknesses or key management overhead?

4.2.7. Routing and forwarding information bases

Section 4.3 shows that routing and forwarding information bases are subject to scalability issue when routing by name is used. While the system is designed it is not only important to ensure that it cannot suffer from targeted attacks aiming at increasing routing and forwarding information bases. As the routing system is tightly bound to the ICN solution itself, there is no universal way to avoid such attacks. However, a possible approache is to combine routing information authenticity validation with filtering (e.g., maximum deaggregation level whenever applicable, black lists, etc.).

4.3. Routing and Resolution System Scalability

ICN routing locates a data object based on its name which is initially provided by a requestor. ICN routing may comprise three steps: a name resolution step, a discovery step, and a delivery step. The name resolution step translates the name of the requested data object into its locator. The discovery step routes the request to data object based on its name or locator. The last step (delivery) routes the data object back to the requestor. Depending on how these steps are combined, ICN routing schemes can be categorized as: Route-By-Name Routing (RBNR), Lookup-By-Name Routing (LBNR), and Hybrid Routing (HR).

4.3.1. Route-By-Name Routing (RBNR)

RBNR omits the first name resolution step. The name of data object is directly used to route the request to the data object. Therefore, routing information for each data object has to be maintained in the routing table. Since the number of data objects is very large (estimated as 10^11 back in 2007 [DONA] but this may be significantly

larger than that, e.g., 10^15 to 10^22), the size of routing tables becomes a concern, as it can be proportional to the number of data object unless an aggregation mechanism is introduced. On the other hand, RBNR reduces overall latency and simplifies the routing process due to the omission of the resolution process. For the delivery step, RBNR needs another identifier (ID) of either host or location to forward the requested data object back to the requestor. Otherwise, an additional routing mechanism has to be introduced, such as bread-crumbs routing [BREADCRUMBS], in which each request leaves behind a trail of breadcrumbs along its forwarding path, and then the response is forwarded back to the requestor consuming the trail. Specific challenges include:

- How to aggregate the names of data objects to reduce the number of routing entries?

- How does a user learn the name which is designed for aggregation by provider? (For example, although we name our contribution as "ICN research challenge", IRTF (provider) may want to change the name to "/IETF/IRTF/ ICN/Research challenge" for aggregation. In this case, how does a user learn the name "/IETF/IRTF/ICN/Research challenge" to retrieve the contribution initially named "ICN research challenge" without any resolution process?)

- Without introducing the name aggregation scheme, can we still achieve scalable routing by taking advantage of topological structure and distributed copies? For example, employing compact routing [COMPACT], random walk [RANDOM] or Greedy routing [GREEDY].

- How to incorporate copies of a data object in in-network caches in this routing scheme?

4.3.2. Lookup-By-Name Routing (LBNR)

LBNR uses the first name resolution step to translate the name of requesting data object into its locator. Then, the second discovery step is carried out based on the locator. Since IP addresses could be used as locators, the discovery step can depend on the current IP infrastructure. The delivery step can be implemented similarly to IP routing. The locator of the requestor is included in the request message, and then the requested data object is delivered to the requestor based on the locator. A specific instantiation of such a system is [MDHT]. Specific challenges include:

- How to build a scalable resolution system which provides

  * Fast lookup: mapping the name of data object to its locators (copies as well).

  * Fast update: the location of data object is expected to change frequently. Also, multiple data objects may change their locations at the same time, e.g., data objects in laptop.

- How to incorporate copies of a data object in in-network caches in this routing scheme?

4.3.3. Hybrid Routing (HR)

HR combines RBNR and LBNR to benefit from their advantages. For instance, within a single administrative domain, e.g., an ISP, where scalability issues can be addressed with network planning, RBNR can be adopted to reduce overall latency by omitting the resolution process. On the other hand, LBNR can be used to route between domains which have their own prefix (locator). A specific challenge here is:

- How to design a scalable mapping system which, given the name of data object, it should return a destination domain locator so that a user request can be encapsulated and forwarded to the domain?

4.4. Mobility Management

Mobility management has been an active field in host-centric communications for more than two decades. In IETF in particular, starting with [RFC2002], a multitude of enhancements to IP have been standardized aiming to "allow transparent routing of IP datagrams to mobile nodes in the Internet" [RFC5944]. In a nutshell, mobility management for IP networks is locator-oriented and relies on the concept of a mobility anchor as a foundation for providing always-on connectivity to mobile nodes. Other standards organizations, such as 3GPP, have followed similar anchor-based approaches. Traffic to and from the mobile node must flow through the mobility anchor, typically using a set of tunnels, enabling the mobile node to remain reachable while changing its point of attachment to the network.

Needless to say, an IP network which supports node mobility is more complex than one that does not, as specialized network entities must be introduced in the network architecture. This is reflected in the control plane as well, which carries mobility-related signaling messages, establishes and tears down tunnels and so on. While mobile connectivity was an afterthought in IP, in ICN this is considered a primary deployment environment. Most, if not all, ICN proposals consider mobility from the very beginning, although at varying levels of architectural and protocol detail. That said, no solution has so

far come forward with a definite answer on how to handle mobility in ICN using native primitives. In fact, we observe that mobility appears to be addressed on ICN proposal specific basis. That is, there is no single paradigm solution, akin to tunneling through a mobility anchor in host-centric networking, that can be applied across different ICN proposals. For instance, although widely-deployed mobile network architectures typically come with their own network entities and associated protocols, they follow the same line of design with respect to managing mobility. This design thinking, which calls for incorporating mobility anchors, permeates in the ICN literature too.

However, employing mobility anchors and tunneling is probably not the best way forward in ICN research for mobile networking. Fundamentally this approach is anything but information-centric and location-indepedent. In addition, as argued in [SEEN], current mobility management schemes anchor information retrieval not only at a specific network gateway (e.g., home agent in Mobile IP) but due to the end-to-end nature of host-centric communication also at a specific correspondent node. However, once a change in the point of attachment occurs, information retrieval from the original "correspondent node" may be no longer optimal. This was shown in [MANI], for example, where a simple mechanism that triggers the discovery of new retrieval providers for the same data object, following a change in the point of attachment, clearly outperforms a tunnel-based approach like Mobile IP in terms of object download times. The challenge here is how to capitalize on location information while facilitating the use of ICN primitives which natively support multicast and anycast.

ICN naming and name resolution, as well as the security features that come along should natively support mobility. For example, CCN [CCN] does not have the restriction of spanning tree routing, so it is able to take advantage of multiple interfaces or adapt to the changes produced by rapid mobility (i.e., there is no need to bind a layer 3 address with a layer 2 address). In fact, client mobility can be simplified by allowing requests for new content to normally flow from different interfaces, or through newly connected points of attachment to the network. However, when the node moving is the (only) content source, it appears that more complex network support might be necessary, including forwarding updates and cache rebuilding. A case in point is a conversation network service, such as a voice or video call between two parties. The requirements in this case are more stringent when support for seamless mobility is required, esp. when compared to content dissemination that is amenable to buffering. Another parameter that needs to be paid attention to is the impact of using different wireless access interfaces based on different technologies, where the performance and link conditions can vary

widely depending of numerous factors.

In host-centric networking, mobility management mechanisms ensure optimal handovers and (ideally) seamless transition from one point of attachment to another. In ICN, however, the traditional meaning of "point of attachment" no longer applies as communication is not restrained by location-based access to data objects. Therefore, a "seamless transition" in ICN ensures that content reception continues without any perceptible change from the point of view of the ICN application receiving that content. Moreover, this transition needs to be executed in parallel with ICN content identification and reaching mechanisms enabling scenarios, such as, preparation of the content reaching process at the target connectivity point, prior to the handover (to reduce link switch disturbances). Finally, these mobility aspects can also be tightly coupled with network management aspects, in respect to policy enforcement, link control and other parameters necessary for establishing the node's link to the network.

In summary, the following research challenges on ICN mobility management can be derived:

- How can mobility management take full advantage of native ICN primitives?
- How do we avoid the need for mobility anchors in a network that by design supports multicast, anycast and location-indepedent information retrieval?
- How can content retrieval mechanisms interface with specific link operations, such as identifying which links are available for certain content?
- How can mobility be offered as a service, which is only activated when the specific user/content/conditions require it?
- How can mobility management be coordinated between the node and the network for optimization and policing procedures?
- How do we ensure that managing mobility does not introduce scalability issues in ICN?

4.5. Wireless Networking

Today, all layer 2 wireless network radio access technologies (L2) are developed with a clear assumption in mind: the waist of the protocol stack is IP and it will be so for the foreseeable future. By fixing the protocol stack waist, engineers can answer a large set of questions, including how to handle conversational traffic (e.g.,

voice calls) vs. web access to online resources, how to support multicast (the IP flavor), and so on, in a rather straightforward manner. Broadcast, on the other hand, which is inherent in wireless communication is not fully taken advantage off. On the contrary, researchers are often more concerned about introducing mechanisms that ensure that "broadcast storms" do not take down a network. The question of how broadcast can serve ICN needs better has yet to be thoroughly investigated.

Wireless networking is often intertwined with mobility but this is not always the case. In fact, empirical measurements often indicate that many users tend to connect (and remain connected) to a single Wi-Fi access point for considerable amounts of time. A case in point, which is frequently cited in different variations in the ICN literature, is access to a document repository during a meeting. For instance, in a typical IETF working group meeting, a scribe takes notes which are uploaded to a centralized repository (see Figure 1). Subsequently, each meeting participant obtains a copy of the document on their own devices for local use, annotation, and sharing with colleagues that are not present at the meeting. Note that in this example there is no node mobility and that it is not important whether the document with the notes is uploaded in one go at the end of the session or in a streaming-like fashion as is typical today with online (cloud-based) document processing.

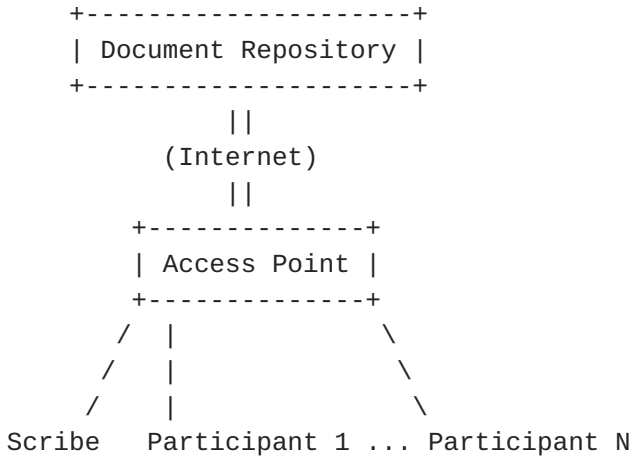

Figure 1: Document sharing during a meeting

In this scenario we observe that the same data object bits (corresponding to the meeting notes) need to traverse the wireless medium at least N+1 times, where N is the number of meeting participants obtaining a copy of the notes. In effect, a broadcast medium is shoehorned into N+1 virtual unicast channels. One could argue that wireless local connectivity is inexpensive, but this is

not the critical factor in this example. The actual information exchange wastes N times the available network capacity, no matter what is the spectral efficiency (or the economics) underlying the wireless technology. This waste is a direct result of extending the remote access paradigm from wired to wireless communication, irrespective of the special characteristics of the latter.

It goes without saying that an ICN approach that does not take into consideration the wireless nature of an interface will waste the same amount of resources as a host-centric paradigm. In-network caching at the wireless access point could reduce the amount of data carried over the backhaul link but, if there is no change in the use of the wireless medium, the NDO will still be carried over the wireless ether N+1 times. Intelligent caching strategies, replica placement cooperation and so on simply cannot alleviate this. On the other hand, promiscuous interface operation and opportunistic caching would maximize wireless network capacity utilization in this example.

Arguably, if one designs a future wireless access technology with an information-centric "layer 3" in mind, many of the design choices that are obvious in an all-IP architecture may no longer be valid. Although this is clearly outside the scope of this document, a few research challenges that the wider community may be interested in include:

- Can we use wireless resources more frugally with the information-centric paradigm than what is possible today in all-IP wireless networks?

- In the context of wireless access, how can we leverage the broadcast nature of the medium in an information-centric network?

- Would a wireless-oriented ICN protocol stack deliver significant performance gains? How different would it be from a wired-oriented ICN protocol stack?

- Is it possible that by changing the network paradigm to ICN we can in practice increase the spectral efficiency (bits/s/Hz) of a wireless network beyond what would be possible with today's host-centric approaches? What would be the impact of doing so with respect to energy consumption?

- Can wireless interface promiscuous operation coupled with opportunistic caching increase ICN performance, and if so, by how much?

- How can a conversational service be supported at least as efficiently as today's state-of-the-art wireless networks deliver?

- What are the benefits from combining ICN with network coding in wireless networks?

- How can MIMO and Coordinated Multipoint Transmission (CoMP) be natively combined with ICN primitives in future cellular networks?

4.6. Transport Services

ICN's receiver-driven communication model as described above creates new options for transport protocol design, as it does not rely solely on end-to-end communication from a sender to a receiver. A requested object can be accessible in multiple different network locations. A node can thus decide how to utilize multiple sources, e.g., by sending parallel requests for the same object or by switching sources (or next hops) in a suitable schedule for a series of requests.

In this model, the requestor would control the data rate by regulating its request sending rate and next by performing source/ next-hop selections. Specific challenges depend on the specific ICN approach, but general challenges for receiver-driven transport protocols (or mechanisms, since dedicated protocols might not be required) include flow and congestion control, fairness, network utilization, stability (of data rates under stable conditions) etc. [HRICP] describes a sample request rate control protocol and corresponding design challenges.

As mentioned above, the ICN communication paradigm does not depend strictly on end-to-end flows, as contents might be received from mid-network caches. The traditional concept of a flow is then somewhat cancelled as sub-flows, or flowlets might be formed on the fly, when fractions of an NDO are transmitted from in-network caches. For a transport layer protocol this is challenging, as any measurement related to this flow, as traditionally done by transport protocols such as TCP, will be hugely misleading. For example, false RTT measurements will lead to largely variable average and smooth RTT values, which in turn will trigger false timeout expirations.

Furthermore, out-of-order delivery is expected to be common in a scenario where parts of a content file are retrieved from in-network caches, rather than from the origin server. Several techniques for dealing with out-of-order delivery have been proposed in the past for TCP, some of which could potentially be modified and re-used in the context of ICN. Further research is needed on this direction though to i) choose the right technique and ii) adjust it according to the requirements of the ICN architecture and transport protocol in use.

ICN offers routers the possibility to aggregate requests and can use several paths, meaning that there is no such thing as a (dedicated)

end-to-end communication path, e.g., a router that receives two requests for the same content at the same time only sends one request to its neighbor. The aggregation of requests has a general impact on transport service design.

Achieving fairness for requestors can be one challenge as it is not possible to identify the number of requestors behind one particular request. A second problem related to request aggregation is the management of request retransmissions. Generally, it is assumed that a router will not transmit a request if it transmitted an identical request recently and because there is no information about the requestor, the router cannot distinguish the initial request from a client from a retransmission from the same client. In such a situation, how routers can adapt their timers to use the best of the communication paths. Finally, aggregation of requests has an impact on the server (producer) side. This last has no way to determine the number of clients actually consuming the content it is producing. This shift of model influences the business model of the server, e.g., how to implement pay-per-click

4.7. In-Network Caching

Explicitly named content objects allow for caching at virtually any network element, including routers, proxy caches and end-host machines. In-network caching can therefore improve network performance by fetching content from nodes geographically placed closer to the end-user. Several issues that need further investigation have been identified with respect to in-network caching. Here we list some of the most important challenges that relate to the properties of the new ubiquitous caching system.

4.7.1. Cache Placement

The declining cost of fast memory gives the opportunity to deploy caches in network routers and take advantage of explicitly named cached contents. There exist two approaches to in-network caching, namely, on-path and off-path caching. Both approaches have to consider the issue of cache location. Off-path caching is similar to traditional proxy-caching or CDN server placement. Retrieval of contents from off-path caches requires redirection of requests and, therefore, is closely related to the Request-to-Cache Routing problem discussed later. Off-path caches have to be placed in strategic points within a network in order to reduce the redirection delays and the number of detour hops to retrieve cached contents. Previous research on proxy-caching and CDN deployment is helpful in this case.

On the other hand, on-path caching requires less network intervention and fits more neatly in ICN. However, on-path caching requires line-

speed operation, which places more constraints on the design and operation of in-network caching elements. Furthermore, the gain of such a system of on-path in-network caches relies on opportunistic cache hits and has therefore been considered of limited benefit, given the huge amount of contents hosted in the Internet. For this reason, network operators might initially consider only a limited number of network elements to be upgraded to in-network caching elements. The decision on which nodes should be equipped with caches is an open issue and might be based, among others, on topological criteria, or traffic characteristics. These challenges relate to both the Content Placement Problem and the Request-to-Cache Routing Problem discussed below.

In call cases, however, the driver for the implementation, deployment and operation of in-network caches will be its cost. Operating caches at line speed inevitably requires faster memories, which increase the implementation cost. Based on the capital to be invested, ISPs will need to make strategic decisions on the cache placement, which can be driven by several factors, such as: avoid inter-domain/expensive links, centrality of nodes, size of domain and the corresponding spatial locality of users, traffic patterns in a specific part of the network (e.g., university vs business vs fashion district of a city).

4.7.2. Content Placement -- Content-to-Cache Distribution

Given a number of (on-path or off-path) in-network caching elements, content-to-cache distribution will affect both the dynamics of the system, in terms of request redirections (mainly in case of off-path caches) and the gain of the system in terms of cache hits. A straightforward approach to content placement is on-path placement of contents as they travel from source to destination. This approach reduces the computation and communication overhead of placing content within the network but, on the other hand, might reduce the chances of hitting cached contents. This relates to the Request-to-Cache Routing problem discussed next.

Furthermore, the number of replicas held in the system brings up resource management issues in terms of cache allocation. For example, continuously replicating content objects in all network elements results in redundant copies of the same objects. The issue of redundant replication has been investigated in the past for hierarchical web caches. However, in hierarchical web-caching, overlay systems coordination between the data and the control plane can guarantee increased performance in terms of cache hits. Line-speed, on-path in-network caching poses different requirements and therefore, new techniques need to be investigated. In this direction, there already exist some studies that attempt to reduce

redundancy of cached copies. However, the issue of coordinated content placement in on-path caches still remains open.

The Content-to-Cache Allocation problem relates also to the characteristics of the content to be cached. Popular content might need to be placed where it is going to be requested next. Furthermore, issues of "expected content popularity" or temporal locality need to be taken into account in designing in-network caching algorithms in order for some contents to be given priority (e.g., popular content vs. one-timers). The criteria as to which contents should be given priority in in-network content caches relate also to the business relationships between content providers and network operators. Business model issues will drive some of these decisions on content-to-cache distribution, but such issues are outside the scope of this note and are not discussed here further.

4.7.3. Request-to-Cache Routing

In order to take advantage of cached contents, requests have to be forwarded to the nodes that temporarily host (cache) the corresponding contents. This challenge relates to name-based routing, discussed before. Requests should ideally follow the path to the cached content. However, instructions as to which content is cached where cannot be broadcast throughout the network. Therefore, the knowledge of a content's location at the time of the request might either not exist, or it might not be accurate (i.e., contents might have been removed by the time a request is redirected to a specific node).

Coordination between the data and the control planes to update information of cached contents has been considered, but in this case scalability issues arise. We therefore, have two options. We either have to rely on opportunistic caching, where requests are forwarded to a server and in case the content is found on the path, then the content is fetched from this node (instead of the original server); or we employ cache-aware routing techniques. Cache-aware routing can either involve both the control and the data plane, or only one of them. Furthermore, cache-aware routing can be done in a domain-wide scale or can involve more than one individual Autonomous System (AS). In the latter case, business relationships between ASes might need to be exploited in order to build a scalable model.

4.7.4. Staleness Detection of Cached NDOs

Due to the largely distributed copies of NDOs in in-network caches, ICN should be able to provide a staleness verification algorithm which provides synchronization of NDOs located at their providers and in-network caching points. It is often argued that ignoring stale

NDOs in caches and simply providing new names for updated NDOs might be sufficient rather than using a staleness verification algorithm to manage them. However, the renaming scheme still causes following problems.

First, notifying the new names of updated NDOs to users is not a trivial task. Unless the update is informed to entire users at the same time, some users would use the old acquainted name by intending to retrieve the updated NDO. Second, the renaming scheme does not provide a mechanism to limit access to the stale NDOs which may be required for a security reason occationally.

One research challenge is how to design consistency and coherence models for caching NDOs along with their revision handling and updating protocols in a scalable manner.

4.8. Network Management

Managing networks has been a core craft in the IP-based host-centric paradigm ever since the technology was introduced in production networks. However, at the onset of IP, management was considered primarily as an add-on. Essential tools that are used daily by networkers, such as ping and traceroute, did not become widely available until more than a decade or so after IP was first introduced. Management protocols, such as SNMP, also became available much later than the original introduction of IP and many still consider them insufficient despite the years of experience we have running host-centric networks. Today, when new networks are deployed network management is considered a key aspect for any operator, a major challenge which is directly reflected in higher OPEX if not done well. If we want ICN to be deployed in infrastructure networks, development of management tools and mechanisms must go hand-in-hand with the rest of the architecture design.

Although defining an FCAPS model for ICN is clearly outside the scope of this document, there is a need for creating basic tools early on while ICN is still in the design and experimentation phases that can evolve over time and help network operations centers (NOC) to define policies, validate that they are indeed used in practice, be notified early on about failures, determine and resolve configuration problems. AAA as well as performance management, from a NOC perspective, will also need to be considered. Given the expectations for a large number of nodes and unprecedented traffic volumes, automating tasks, or even better employing self-management mechanisms is preferred. The main challenge here is that all tools we have at our disposal today are node-centric, end-to-end oriented, or assuming a packet-stream communication environment. Rethinking reachability

and operational availability, for example, can yield significant insights into how information-centric networks will be managed in the future.

With respect to network management we see three different aspects. First, any operator needs to manage all resources available in the network, which can range from node connectivity to network bandwidth availability to in-network storage to multi-access support. In ICN, users will also bring into the network significant resources in terms of network coverage extension, storage, and processing capabilities. DTN characteristics should also be considered to the degree that this is possible (e.g., content dissemination through data mules). Secondly, given that nodes and links are not at the center of an information-centric network, network management should capitalize on native ICN mechanisms. For example, in-network storage and name resolution can be used for monitoring, while native publish/subscribe functionality can be used for triggering notifications. Finally, management is also at the core of network controlling capabilities by allowing operating actions to be mediated and decided, triggering and activating networking procedures in an optimized way. For example, monitoring aspects can be conjugated with different management actions in a coordinated way, allowing network operations to flow in a concertated way.

However, the considerations on leveraging intrinsic ICN mechanisms and capabilities to support management operations go beyond a simple mapping exercise. In fact, not only it raises a series of challenges on its own, but also opens up new possibilities for both ICN and "network management" as a concept. For instance, naming mechanisms are central to ICN intrinsic operations, which are used to identify and reach content under different aspects (e.g., hierarchically structured vs. 'flattish' names). In this way, ICN is decoupled from host-centric aspects on which traditional networking management schemes rely upon. As such, questions are raised which can directly be translated into challenges for network management capability, such as, for example how to address a node or a network segment in a ICN naming paradigm, how to identify which node is connected "where", and if there is a host-centric protocol running from which the management process can also leverage upon.

But, on the other hand, these same inherent ICN characteristics also allow us to look into network management through a new perspective. By centering its operations around content, one can conceive new management operations addressing, for example, per-content management or access control, as well as analyzing performance per content name instead of per link or node. Moreover, such considerations can also be used to manage operational aspects of ICN mechanisms themselves. For example, [NDN-MGMT] re-utilizes inherent content-centric

capabilities of CCN to manage optimal link connectivity for nodes, in concert with a network optimization process. Conversely, how these content-centric aspects can otherwise influence and impact management in other areas (e.g., security, resilience) is also important, as exemplified by in [ccn-access], where access control mechanisms are integrated into a prototype of the [PURSUIT] architecture.

In this way, a set of core research challenges on ICN management can be derived as:

- Manage and control content reception at the destination
- Coordination of management information exchange and control between ICN nodes and ICN network control points Identification of management and controlling actions and items through information naming
- Relationship between NDOs and host entities identification (i.e., how to identify a particular link, interface or flow that need to be managed)

5. Link to and Impact on IETF Technologies

TBW later.

6. Security Considerations

Security related questions related to ICN are discussed in Section 4.2.

7. Informative References

[BACKSCATTER]
Waehlisch, M., Schmidt, TC., and M. Vahlenkamp, "Backscatter from the Data Plane - Threats to Stability and Security in Information-Centric Network Infrastructure", Computer Networks Vol 57, No. 16, pp. 3192-3206, November 2013.

[BREADCRUMBS]
Rosensweig, E. and J. Kurose, "Breadcrumbs: Efficient, Best-Effort Content Location in Cache Networks", In Proceedings of the IEEE INFOCOM 2009, April 2009.

[CCN] Jacobson, K, D, F, H, and L, "Networking Named Content",

CoNEXT 2009 , December 2009.

[COMPACT] Cowen, L., "Compact routing with minimum stretch", In Journal of Algorithms, vol. 38, pp. 170--183, 2001.

[Chaum] Chaum, D. and E. van Heijst, "Group signatures", In Proceedings of EUROCRYPT, 1991.

[DONA] Koponen, T., Ermolinskiy, A., Chawla, M., Kim, K., gon Chun, B., and S. Shenker, "A Data-Oriented (and Beyond) Network Architecture", In Proceedings of SIGCOMM 2007, August 2007.

[GREEDY] Papadopoulos, F., Krioukov, D., Boguna, M., and A. Vahdat, "Greedy forwarding in dynamic scale-free networks embedded in hyperbolic metric spaces", In Proceedings of the IEEE INFOCOM, San Diego, USA, 2010.

[HRICP] Carofiglio, G., Gallo, M., and L. Muscariello, "Joint hop-by-hop and receiver-driven interest control protocol for content-centric networks", In Proceedings of ACM SIGCOMM ICN 2012, DOI 10.1145/2342488.2342497, 2012.

[ICNNAMING] Ghodsi, A., Koponen, T., Rajahalme, J., Sarolahti, P., and S. Shenker, "Naming in Content-Oriented Architectures", In Proceedings ACM SIGCOMM Workshop on Information-Centric Networking (ICN), 2011.

[ICNSURVEY] Ahlgren, B., Dannewitz, C., Imbrenda, C., Kutscher, D., and B. Ohlman, "A Survey of Information-Centric Networking", In Communications Magazine, IEEE , vol.50, no.7, pp.26-36, DOI 10.1109/MCOM.2012.6231276, 2012.

[MANI] Pentikousis, K. and T. Rautio, "A multiaccess Network of Information", WoWMoM 2010, IEEE , June 2010.

[MDHT] D'Ambrosio, M., Dannewitz, C., Karl, H., and V. Vercellone, "MDHT: A hierarchical name resolution service for information-centric networks", ACM SIGCOMM workshop on Information-centric networking Toronto, Canada, 2011, August 2011.

[NDN-MGMT] Corujo, D., Aguiar, R., Vidal, I., and J. Garcia-Reinoso, "A named data networking flexible framework for management communications", Communications Magazine, IEEE , vol.50,

no.12, pp.36-43 , December 2012.

[PURSUIT] Fotiou et al., N., "Developing Information Networking Further: From PSIRP to PURSUIT", In Proceedings of Proc. BROADNETS. ICST, 2010.

[RANDOM] Gkantsidis, C., Mihail, M., and A. Saberi, "Random walks in peer-to-peer networks: algorithms and evaluation", In Perform. Eval., vol. 63, pp. 241--263, 2006.

[RFC2002] Perkins, C., "IP Mobility Support", RFC 2002, October 1996.

[RFC5246] Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, August 2008.

[RFC5280] Cooper, D., Santesson, S., Farrell, S., Boeyen, S., Housley, R., and W. Polk, "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile", RFC 5280, May 2008.

[RFC5944] Perkins, C., "IP Mobility Support for IPv4, Revised", RFC 5944, November 2010.

[RFC6920] Farrell, S., Kutscher, D., Dannewitz, C., Ohlman, B., Keranen, A., and P. Hallam-Baker, "Naming Things with Hashes", RFC 6920, April 2013.

[SEEN] Pentikousis, K., "In search of energy-efficient mobile networking", Communications Magazine, IEEE, vol. 48, no. 1, pp.95-103 , January 2010.

[ccn-access] Fotiou, N., Marias, G., and G. Polyzos, "Access control enforcement delegation for information-centric networking architectures", In Proceedings of the second edition of the ICN workshop on Information-centric networking (ICN '12). ACM, New York, NY, USA, 85-90., 2012.

Authors' Addresses

Dirk Kutscher (editor)
NEC
Kurfuersten-Anlage 36
Heidelberg,
Germany

Phone:
Email: kutscher@neclab.eu

Suyong Eum
National Institute of Information and Communications Technology
4-2-1, Nukui Kitamachi, Koganei
Tokyo 184-8795
Japan

Phone: +81-42-327-6582
Email: suyong@nict.go.jp

Kostas Pentikousis
EICT GmbH
Torgauer Strasse 12-15
Berlin 10829
Germany

Email: k.pentikousis@eict.de

Ioannis Psaras
University College London, Dept. of E.E. Eng.
Torrington Place
London WC1E 7JE
United Kingdom

Email: i.psaras@ucl.ac.uk

Daniel Corujo
Universidade de Aveiro
Instituto de Telecomunicacoes, Campus Universitario de Santiago
Aveiro P-3810-193
Portugal

Email: dcorujo@av.it.pt

Damien Saucez
INRIA
2004 route des Lucioles - BP 93
Sophia Antipolis 06902 Cedex
France

Email: damien.saucez@inria.fr

Thomas C. Schmidt
HAW HAmburg
Berliner Tor 7
Hamburg 20099
Germany

Email: t.schmidt@ieee.org

Matthias Waehlisch
FU Berlin
Takustr. 9
Berlin 14195
Germany

Email: waehlisch@ieee.org