

The Benefits & Challenges of Networking Named Data

Mark Baugher, Bruce Davie, Ashok Narayanan and Dave Oran
{mbaugher, bdavie, ashokn, oran}@cisco.com
Cisco Systems, Inc.



Overview: Networking Named Data^{*}

Benefits

- *Persistence*
- *Availability*
- *Authenticity*
- *Cache-ability*

Challenges

- *Public key cryptography*
- *Integrity*
- *Privacy*
- *Inter-domain routing*

^{*}Jacobson et. Al., Networking Named Content, CoNEXT'09, December 1–4, 2009, Rome, Italy.

^{*}Koponen et. Al., A Data-Oriented (and Beyond) Network Architecture, SIGCOMM'07, August 27–31, 2007, Kyoto, Japan

The Benefits of CCN and DONA

Content-Centric Networking (CCN) and Data Oriented Network Architecture (DONA) work differently but both envision a future Internet that is optimized for content delivery. Today, web content can be authenticated or cached but not both. These designs separate content delivery from a network endpoint so that content can be both cached and authenticated throughout any network.

Four Benefits of CCN and DONA

- 1. Persistence:** *No broken links*
 - Data packet is associated with a name and not an address
- 2. Availability:** Reliable, low-latency and global delivery
 - No need for ad-hoc CDN or application-layer P2P delivery
- 3. Authenticity:** Object is what the publisher published
 - Not limited to pair-wise, transient HTTPS integrity
- 4. Caching:** Data have integrity and known provenance

CCN and DONA Work Differently

	CCN	DONA
Naming	Structured, human-readable	Unstructured, self-certifying name carry a hash of the public key
Public Key Cryptography	PKI associates a key with name; names and content are signed	A directory maps public key to real-world identity, content is signed
Name governance	Centralized or distributed	Distributed
Name resolution	Integrated with routing	DHT resolution to Internet address
Routing	Routes hierarchical names	Uses Internet routing for data
Caching	Integrated with routing	Can route through caches
Layering	Independent of IP and other layers	Shim between IP and transport

Content-Centric Networking

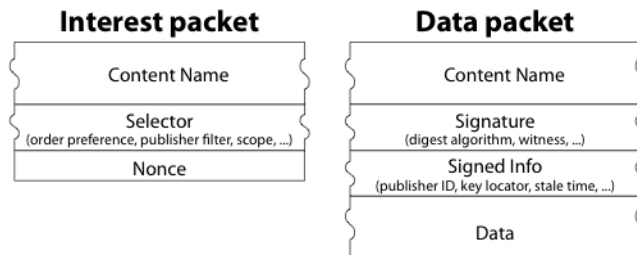


Figure 2: CCN packet types

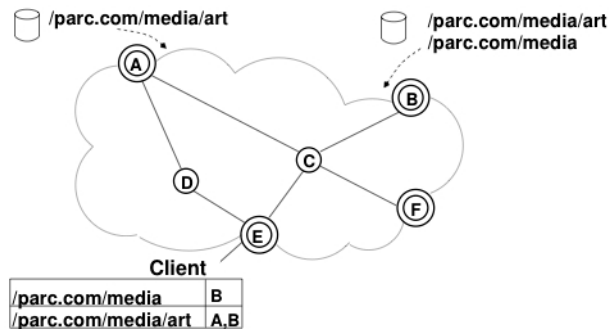


Figure 6: Routing Interests to a domain's media content

- 1 Interest per Data packet
 - Runs on anything
 - Anything can run on it
- CCN names
 - Hierarchical human-readable
 - Designed to use OSPF, BGP, etc.
- Ubiquitous Caching
 - Any router can be a cache
- “Muscular” forwarding plane
 - E.g. Adaptive Interest Forwarding¹

From: Jacobson et. Al., Networking Named Content, CoNEXT'09, December 1–4, 2009, Rome, Italy.

¹Li et. Al., Robust Packet Delivery via Named Data, 2011

Data-Oriented Network Architecture

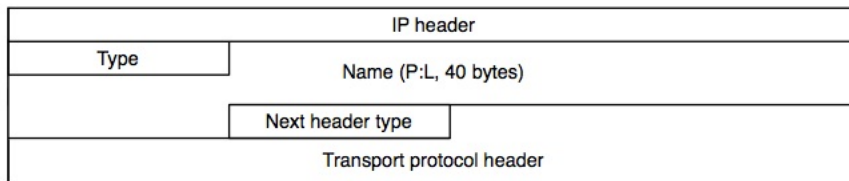


Figure 3: Protocol headers of a FIND packet. Type is to separate FINDs from their responses.

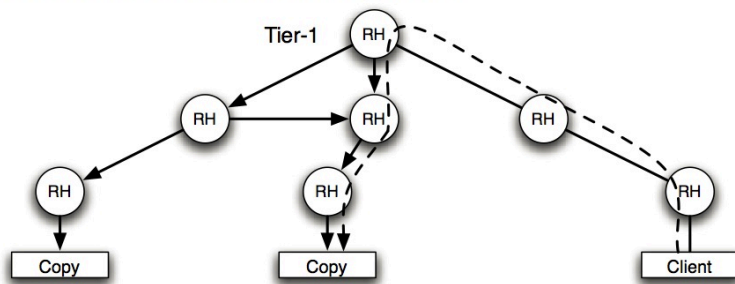


Figure 1: Registration state (solid arrows) in RHs after copies have registered themselves. RHs route client-issued FIND (dashed arrow) to a nearby copy.

- Find and Register
 - like SIP Invite/Register
 - Runs on Internet Protocol
- Flat, self-certifying names
 - SFS and P2P inspired, DHT model
 - Built-in authenticity
 - Name is Principal:Label (P:L)
- Name names a three-tuple
<data, public key, signature>

CCN and DONA Challenges

There are technical challenges in the areas of privacy, public key management, integrity, denial of service, and the inter-domain scalability of hierarchical or flat names.

Four Challenges

1. **Authenticity using public-key cryptography (PKC)**

- CCN needs a PKI to bind a name to a public key
- DONA needs no PKI but must bind a public key to authorized identity

2. **Integrity**

- Neither say what happens when a data packet does not verify

3. **Privacy**

- Both trade IP address exposure for data-object exposure

4. **Inter-domain Routing**

- Neither have a detailed design or proof of concept

CCN & DONA Authenticity using PKC

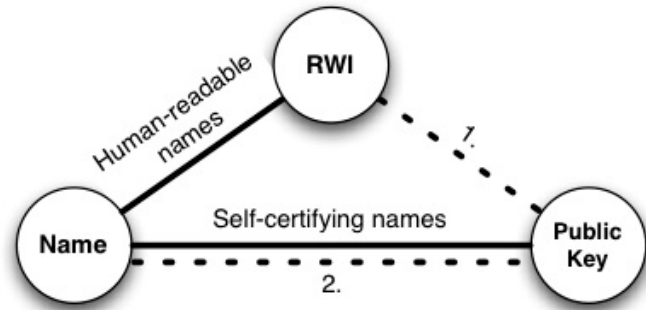


Figure 1: A depiction of the three entities and the different bindings between them. Two naming schemes provide different intrinsic bindings (solid lines) but require both an external authority to provide one additional binding (dashed line): with self-certifying names it's the binding (1), whereas with human-readable names it's the binding (2).

From: Ghodsi, et. Al. Naming in Content-Oriented Architectures, Workshop on Information-Centric Networking, Sigcomm 2011

- CCN and DONA differ in key authorization method
- CCN names are human readable and signed with the data packets
- DONA has self-certifying names, P:L, with two variants
 1. P:L where P is a hash of P's public key and L is any unique label
 2. P:L where P is a hash of P's public key and L is a hash or digest of the packet
- Smetters & Jacobson noted that the first form is insecure
 - #2 is a “self-verifying name” that has light-weight content-integrity verification

CCN Authenticity: Merkle Trees

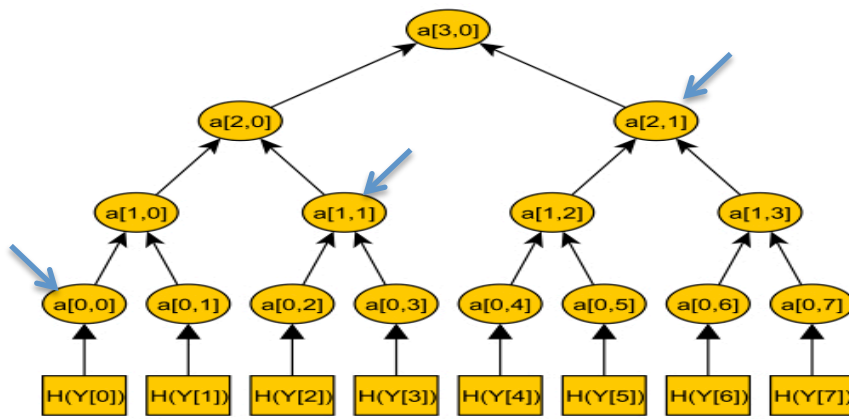


Figure 3.1: Merkle tree with 8 leaves

From: G. Becker, Merkle Signature Schemes, Merkle Trees and Their Cryptanalysis, Ruhr-Universität Bochum, 2008

- Y_i leafs are packets from a file
- Each node is SHA-1 of children
- Example: Y_1 packet carries log N hashes (shown with arrows)
- Hash $a[3,0]$ is RSA signed
- Only 1 PKC operation per file
- Must manage a public key

Authenticity: PKI and PKC Challenges

Google URLs (2008)	$O(10^{12})$	googleblog.blogspot.com/2008/07/we-knew-web-was-big.html
Web pages	$O(10^{10})$	www.worldwidewebsite.com
Internet users	$O(10^9)$	www.internetworldstats.com/stats.htm
Web servers	$O(10^8)$	news.netcraft.com/archives/category/web-server-survey/
PGP global directories	$O(10^6)$	sks-keyservers.net/status/info/pgp.mit.edu
SSL PKI	$O(10^5)$	blog.ivanristic.com/2011/09/ssl-survey-protocol-support.html

If data objects are to be signed:

- DONA needs key directories
 - $O(10^9)$ users who might publish
 - $O(10^6)$ users have public keys today
 - 3 orders of magnitude more than today
- CCN needs a PKI
 - $O(10^8)$ web servers who might sign data
 - $O(10^5)$ web servers use SSL today
 - 3 orders of magnitude more than today

But, should all data object be signed?

Authenticity Without Object Signatures

“The main lesson learned was that the trust structure embodied in X.509 and related standards is not suited for such applications [publishing digital books]. Indeed, it may turn out that digital signatures are not the appropriate tool for the job, but rather secure cataloguing and notarisation services based on trees of hash values.”

From: Anderson, R. J., et. Al., Secure Books: Protecting the Distribution of Knowledge, IN PROCEEDINGS OF SECURITY PROTOCOLS WORKSHOP '97

X.509 public key cryptography has several problems when used in digital publishing:

1. X.509 key lifetimes are often too short for a published work.
2. What does key expiration/revocation mean for a published work?
3. How will people manage a key-pair for a lifetime across various computer systems and with attendant risks of exposure?
4. Some users need non-repudiation where a one-time signature is ideal but the typical 2-3 year lifetime of a key-pair is too long.

Integrity Without Packet Signatures

- Content router may need to integrity-check packets
 - No way to signal a router that it's forwarding bad packets
 - No obvious way to clean up paths to bogus data; edge content router may need a reverse "no cache" option
 - Can routers fetch keys, verify signatures in forwarding path?
- No signature required: split integrity from authenticity
 - A simpler, SHA-1 hash is all that's needed to verify integrity
 - This is one form of a DONA P:L name, but without the P

Solution: Sign Collections, Not Objects

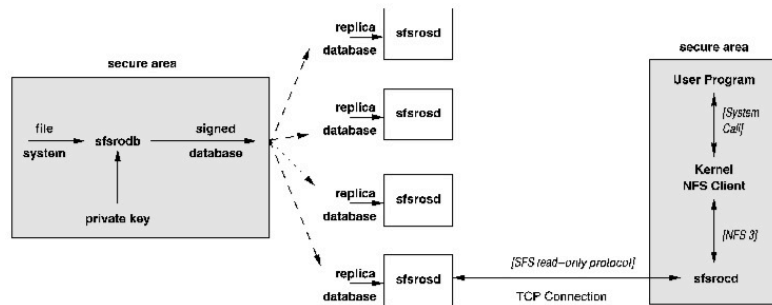


Fig. 1. The SFS read-only file system. Shaded boxes show the trusted computing base.

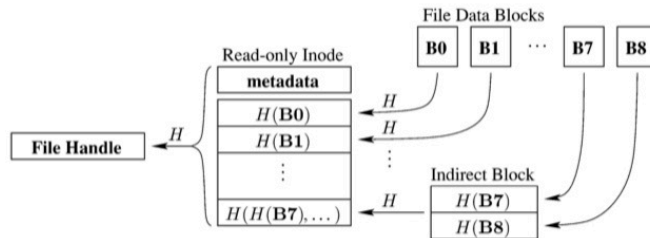


Fig. 3. Format of a read-only file system inode.

From: Fu, et. Al. Fast and Secure Distributed Read-Only File System, ACM Transactions on Computer Systems, 2002

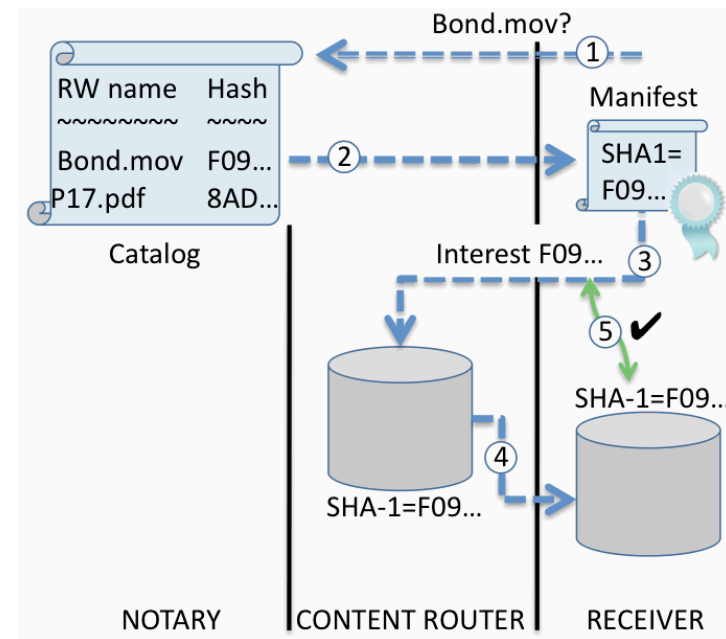
- Fu, et. Al. sign collections, not objects using a signed database
 - Of one or more NFS file systems
 - Each Inode or block is named by its hash
- When client fetches an inode
 - 1 database check for authenticity
 - Every name is checked for integrity
- When client fetches a block
 - Computes block hash
 - Compares hash with name

Proposal: Self-Verifying Names

- Self-verifying name might be a SHA-1 hash of the named object
 - Verifiable: compute the object hash and compare with the object name
 - Receiver gets authenticity/provenance by matching hash to name **it requested**
 - Content routers need only ensure integrity, i.e. hash matches name
 - Robust: check integrity when public keys are not used – or used improperly
 - A brittle solution would require that signatures always be verified using PKC
 - Robust solutions tolerate poor PKC implementations - or user refusal to use PKC
 - Robust solutions ensures integrity, and also authenticity if PKC is used properly
- Receiver needs each packet's hash/name before sending Interest
 - Append the hash name to CCN name, or use L = hash in DONA name

Self-Verifying Names & Certified Catalogs

- Notary signs catalog of name pairs: <real-world name, hash>
 - Sink verifies authenticity using PKC
 - Or assumes catalog is authentic
- Notary serves a hash name
 - To a sink, given a real-world name
 - May send the sink a list so it can request authentic packets by name
- If catalog is private, hash name & real-world name are private



Self-Verifying Names and Privacy

- A self-verifying name (not human readable) needs a catalog
 - A catalog associates hash name with real-world name of an object
 - A secure catalog is signed by the cataloger and verified by receiver
 - A receiver using an unverified catalog assumes catalog integrity
 - Security requires verification and PKC, but there is value even without it
- A self-verifying name is private if the catalog is private
 - A private catalog might serve keys to encrypted data objects
 - Group key encryption allows efficient, nearest-node caching

Privacy in Data Networking Designs

- Seen as an important problem for CCN, DONA and others
 - Without privacy, at least one router can identify “who” gets “what”
- The web today exposes an endpoint addresses
 - And usually exposes URL as well as the content
- CCN and DONA expose the content name
 - And exposes receiver endpoint information as well as content
- If privacy method must encrypt, it requires key management
- Recent works use Tor and content-oriented mixing

Receiver, Name, and Content Privacy

Onion Routing^x

- PARC/UCI presented a two-hop solution at CCNxCon
- Improves on the 3-hop solution for IP
- Hides receiver

^xUzun, et. Al., Anonymity in Named Data Networking, CCNxCon 2011,
http://www.ccnx.org/wp-content/uploads/2011/08/ccnxcom_slides.pdf

Content Mixes^{*}

- Applies censorship-resistant techniques
- Requires additional mix objects to be retrieved
- Hides name and content

^{*} Arianfar, Koponen, Raghavan, Shenker, On Preserving Property in Content Oriented Networks, Sigcomm 2011 ICN

Comparison of Content-Oriented Privacy Methods

Method	Routing	Caching	Transaction costs ^a
Onion-based routing	Expands path	Not cacheable at nearest neighbor	4 PKC operations ^b 4 encryptions ^c 4 decryptions ^d
Content mix	Does not change	Not cacheable	NM xor operations ^e NM block downloads
Self-verifying names	Does not change	Cacheable anywhere	≤ 1 PKC operation 1 encryption (group) 1 decryption (group)

^a All methods have a base-line transaction costs of 1 Interest packet and N blocks downloaded in a data packet

^b 2 PKC encryptions of symmetric keys for each layer plus 2 PKC decryptions of the key on 2 paths (forward path and reverse path)=4X2

^c 2 symmetric key encryptions for each layer on 2 paths

^d 2 decryptions with a symmetric key on 2 paths.

^e A file has each of its N blocks xored with M blocks from different files

Quadrillions of Names, Private and Public

Number of names	$O(10^{15})$	conferences.sigcomm.org/sigcomm/2011/papers/icn/p7.pdf
IP hosts today	$O(10^8)$	ftp.isc.org/www/survey/reports/current/
IP prefixes in DFZ today	$O(10^5)$	en.wikipedia.org/wiki/File:BGP_Table_growth.svg
AS numbers in DFZ	$O(10^4)$	Multihoming and mobility expand the DFZ IP prefix table size

Self-verifying names like DONA's self-certifying names are flat, but CCN's are hierarchical, human readable

- May have $O(10^{15})$ content names^x
 - There are only $O(10^8)$ host IP names
 - 7 orders of magnitude larger than today
- Challenges of route table expansion
 - CCN aggregates names, uses OSPF, BGP,...
 - DONA looks to ROFL and DHT research

^xD'Ambrosio, et. Al., MDHT: A Hierarchical Name Resolution Service for Information-centric Networks, ICN Workshop, Sigcomm 2011

Recent Work in Scalable Content Naming

- DONA-based Naming -



Figure 2: In deepest match, an exact match for each part of the name is looked for. The matching begins from the end of the name (above D) and proceeds a part by part to the beginning (A), until a match or there's nothing to look for.

Ghodsji, et. Al., Naming in Content-Oriented Architectures, Sigcomm 2011, ICN Workshop <http://conferences.sigcomm.org/sigcomm/2011/papers/icn/p1.pdf>



- CCN Naming -

Link State Routing

- Extend OSPF for name-based routing
 - Advertise name prefixes via name opaque LSAs
 - Advertise NDN links via adjacency opaque LSAs
 - Keep most protocol operations and algorithms
 - Update CCNx FIB
 - Based on Quagga, currently under testing, plan to deploy on the NDN testbed next month.
- Scale by aggregation
 - Map app names to ISP-assigned names
 - Global routing on aggregated ISP-assigned prefixes.

Zhang, B., NDN Routing Overview, CCNxCon 2011, http://www.ccnx.org/wp-content/uploads/2011/08/ccnxcom_slides.pdf

Summary: Answers to the Challenges

Challenges	Answers	Comment
Privacy	Tor	CCN: No nearest caching, lengthens paths
	Content mixes	CCN, DONA: Inflates object size, non-cacheable
	Self-verifying names	Requires a private catalog and group-key encryption
Global Scalability	Hierarchical naming	CCN: Mutable names for OSPF and BGP
	Routing on Flat Labels	DONA: ROFL and “deepest match” method
Integrity	Bind name to a key	CCN: PKI maps RW name to key
	Bind name to content hash	DONA: A P:L when L is a (SHA-1) hash of content object
	Self-verifying name	Name is a hash (e.g. SHA-1 hash) of content object
Authenticity and PKC	Root certification trees	CCN: Browser forest of root certs
	Public key directories	DONA: Keys and real-world names
	Catalog signed by a notary	A self-verifying name and real-world name pair

Conclusion: Benefits and Challenges

- Data networking research may transform the Internet for content delivery by making replicable content the end point
- CCN and DONA are two archetypes of data networking
 - They take different approaches to naming, routing and caching
 - ccnx development effort, drives leading-edge research with NDN project
 - They both have privacy, integrity, global scalability challenges
- There are three different approaches to using public key cryptography
 - DONA (self-certifying names) and CCN (PKI names) use PKC in data path
 - Self-verifying names use only SHA-1 in data path, PKC in session control
- A challenge not well-addressed in this presentation is global scalability



CISCO