Information Centric Networking(ICN) for Delivering Big Data with Persistent Identifiers(PID)

Andreas Karakannas

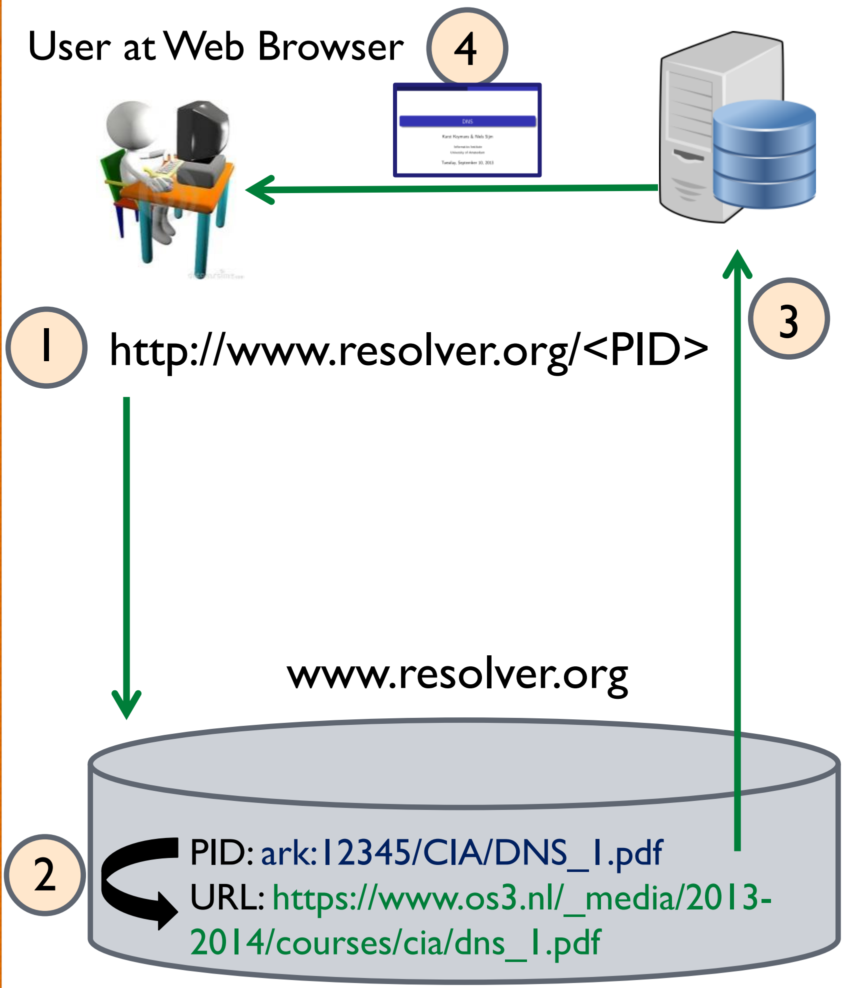
Research Project 2

Supervised by:

Zhiming Zhao

Background

PIDs in IP Network



Information Centric Networking

- A new network concept
- Based on the idea that users are interested in accessing Digital Objects regardless of their locations.
- No end-to-end communication
- Digital Objects are uniquely identified
- Request for Objects are routed based on the Digital Object unique name (NO IP ROUTING!!!)
- Objects are cached in the path from source to destination(In-Network Caching).
- In-Network Caching aims to achieve efficient & reliable distribution of the contents among the network infrastructure.

Research Questions

- How can PID types be mapped/resolved to ICNs' Object Identifiers?
- What is the efficiency of ICNs' caching algorithms for delivering Big Data?

Approach

- Theoretical Studies on latest ICN Projects and PID Standards.
- Propose Mapping Architecture Design based on the Theoretical study
- Evaluate In-Network Caching Performance for Big Data Objects

Theoretical Studies
ICN Approaches

ICN approaches

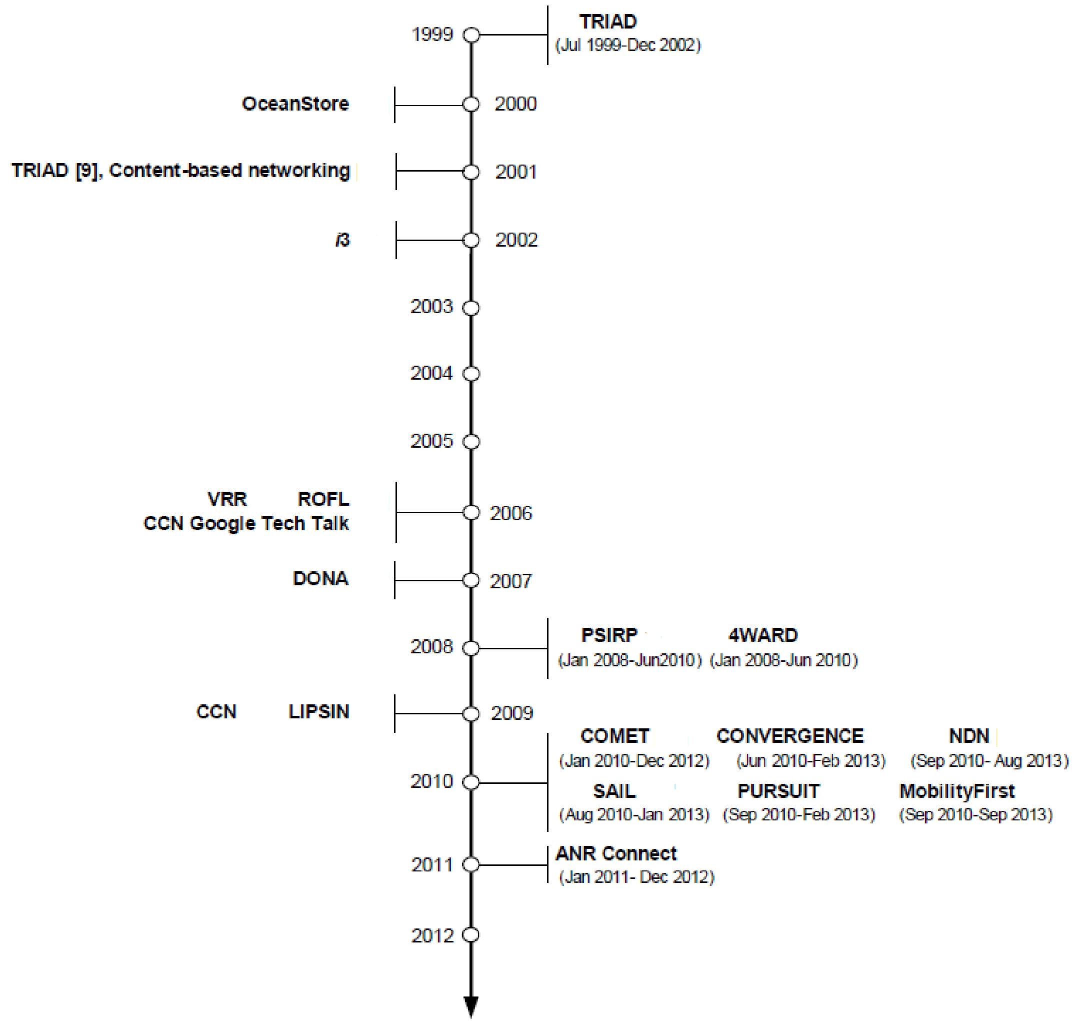


Fig. 1. Timeline of key ICN milestones. Seminal ICN papers are shown on the left hand side, while ICN-related projects are shown on the right hand side.

- A Survey of Information-Centric Networking Research

Theoretical Studies
ICN Approaches

ICN approaches

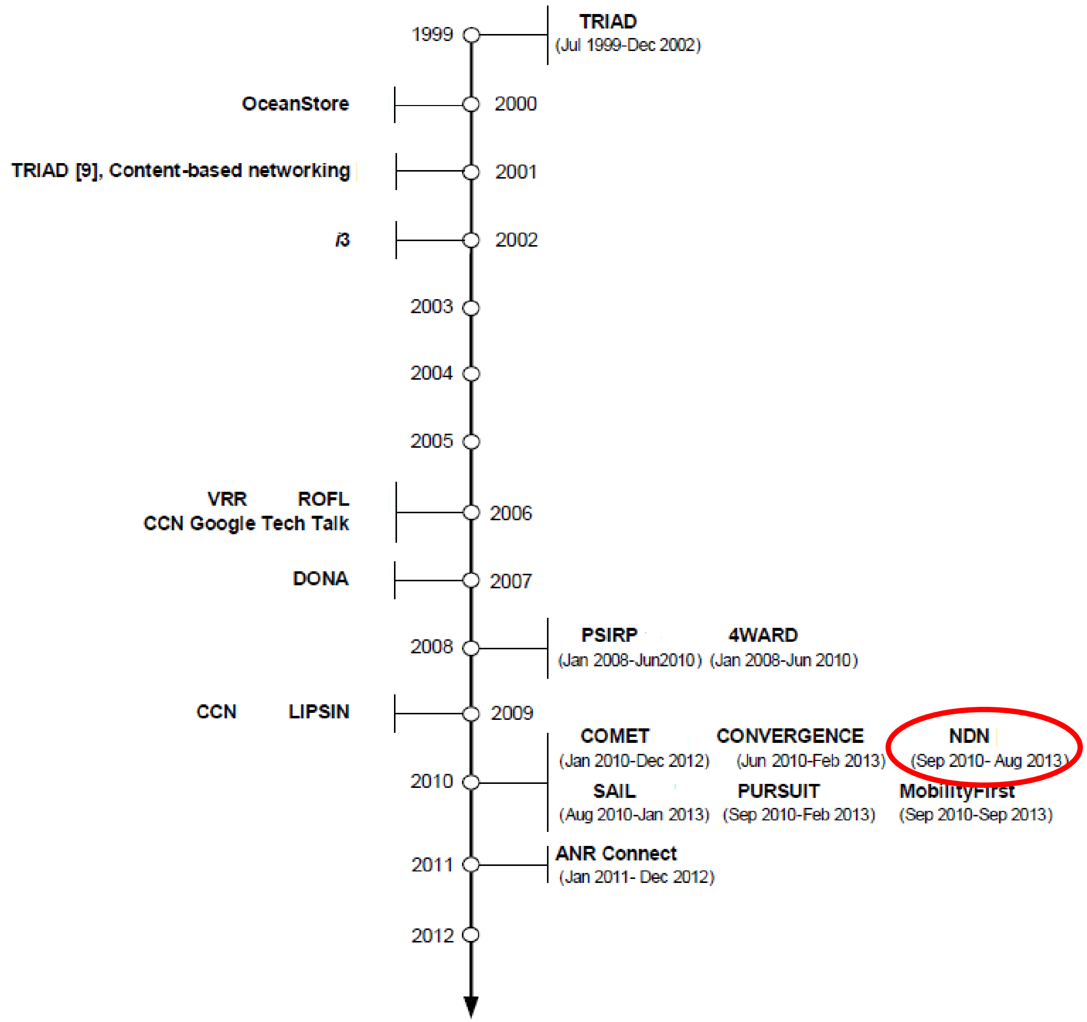


Fig. 1. Timeline of key ICN milestones. Seminal ICN papers are shown on the left hand side, while ICN-related projects are shown on the right hand side.

- A Survey of Information-Centric Networking Research

Named Data Networking(NDN)

- The most mature ICN approach.
- The only approach with published specification.(Packet Format 0.1a2 published on March 27,2014).
- Most research in caching algorithms in ICN is based on NDN.
- Only one with available open source simulators(ndnSIM,ccnSIM) for evaluating caching performance under different scenarios.

Named Data Networking(NDN)

- Names in NDN
  - Based on URI syntax
  - Have hierarchical structure (e.g. /NL/Amsterdam/UVA/ComputerScience/OS3/CIA/DNS.pdf)
  - Names can be anything: a pdf file, a video, an endpoint, a command to turn on some lights.
  - Names are used in the Routing procedure.

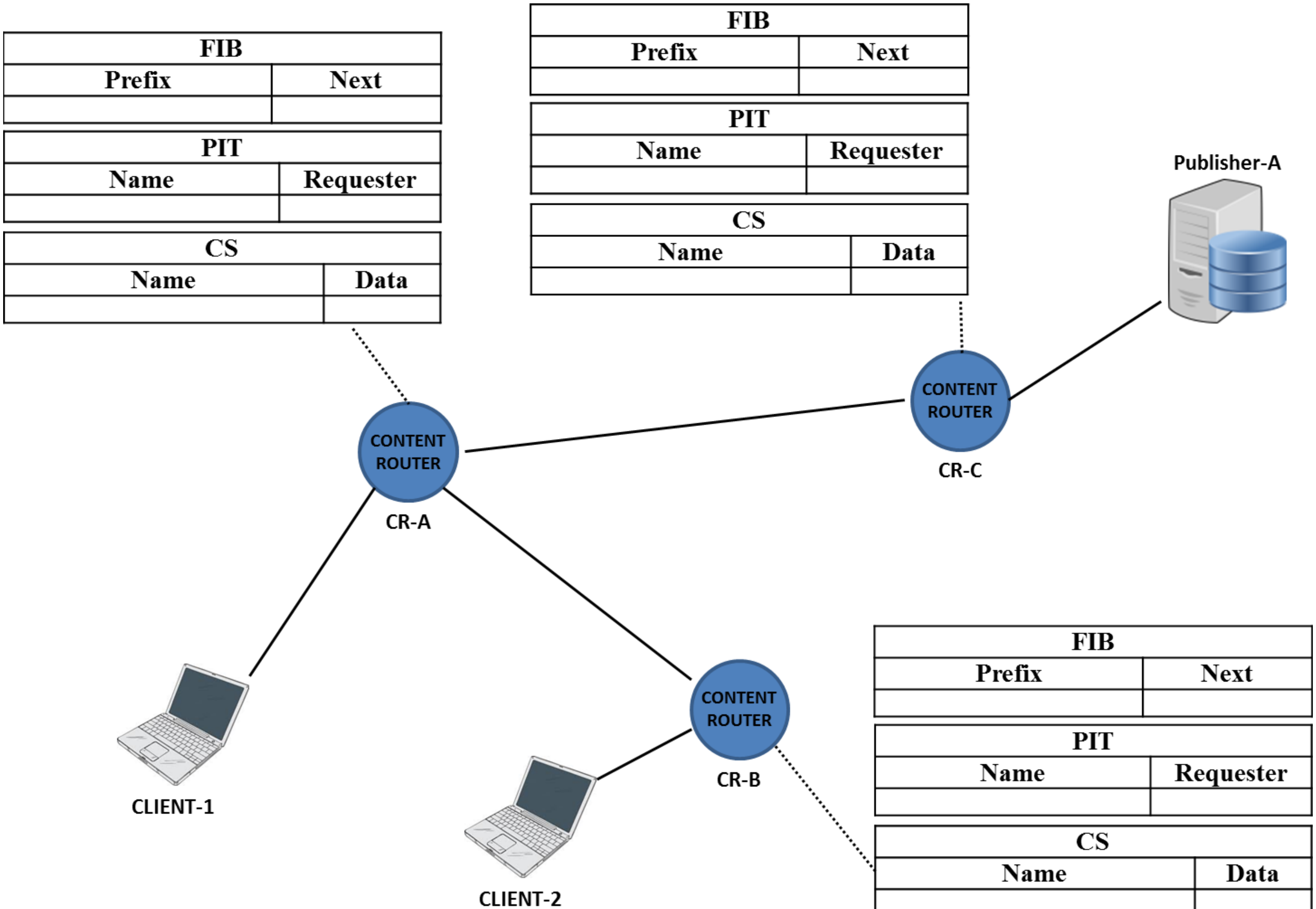
- 2 Types of packets
  - INTEREST(request) packets
    - Contains the Name of the Request

      e.g. INTEREST(/NL/Amsterdam/UVA/ComputerScience/OS3/CIA/DNS.pdf)
  - DATA(answer) packets
    - Contains the Name of the Request & the Data

      e.g. DATA(NL/Amsterdam/UVA/ComputerScience/OS3/CIA/DNS.pdf, <DATA>)

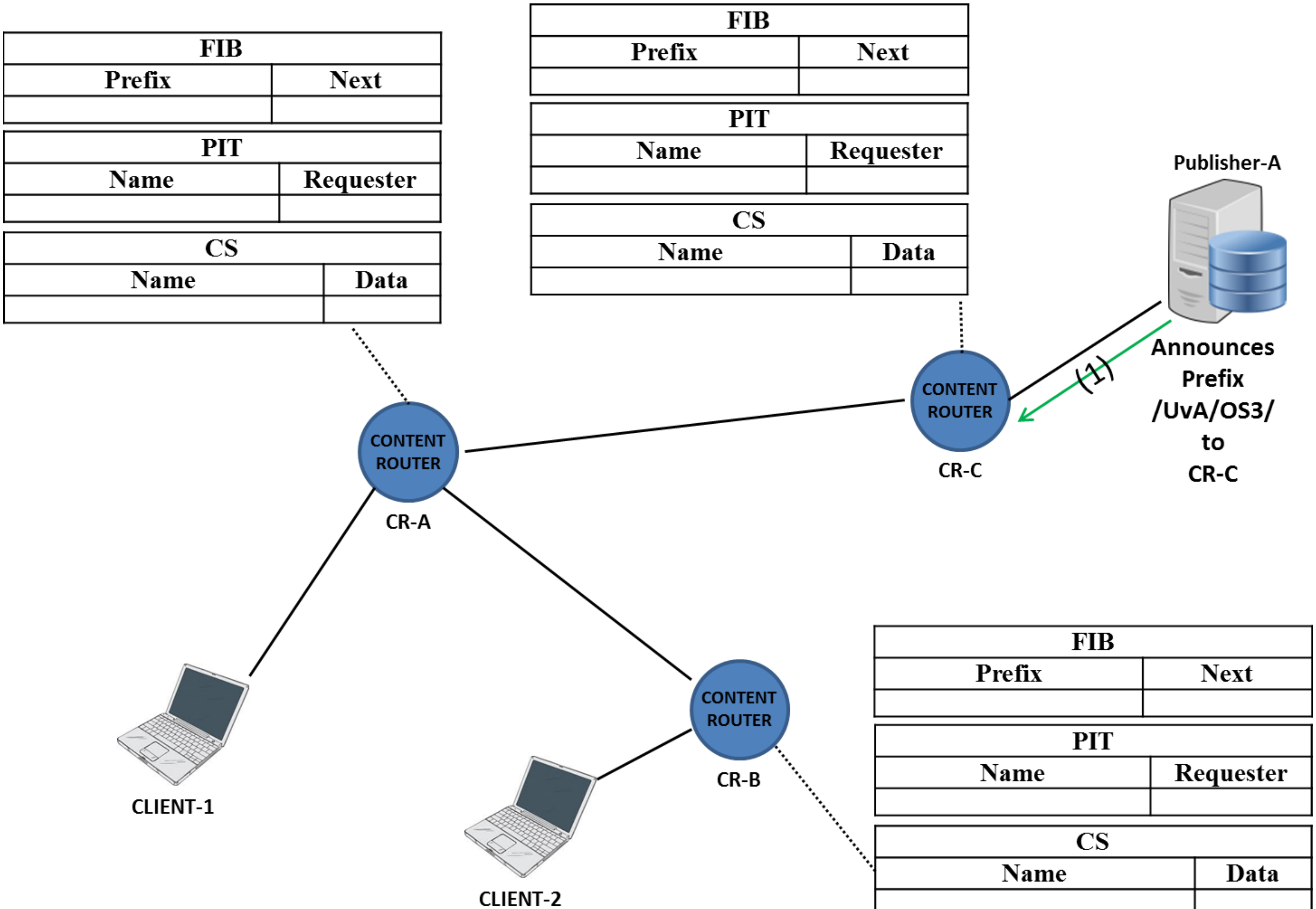
Named Data Networking(NDN)

Theoretical Studies
NDN
Populating the Name Prefix



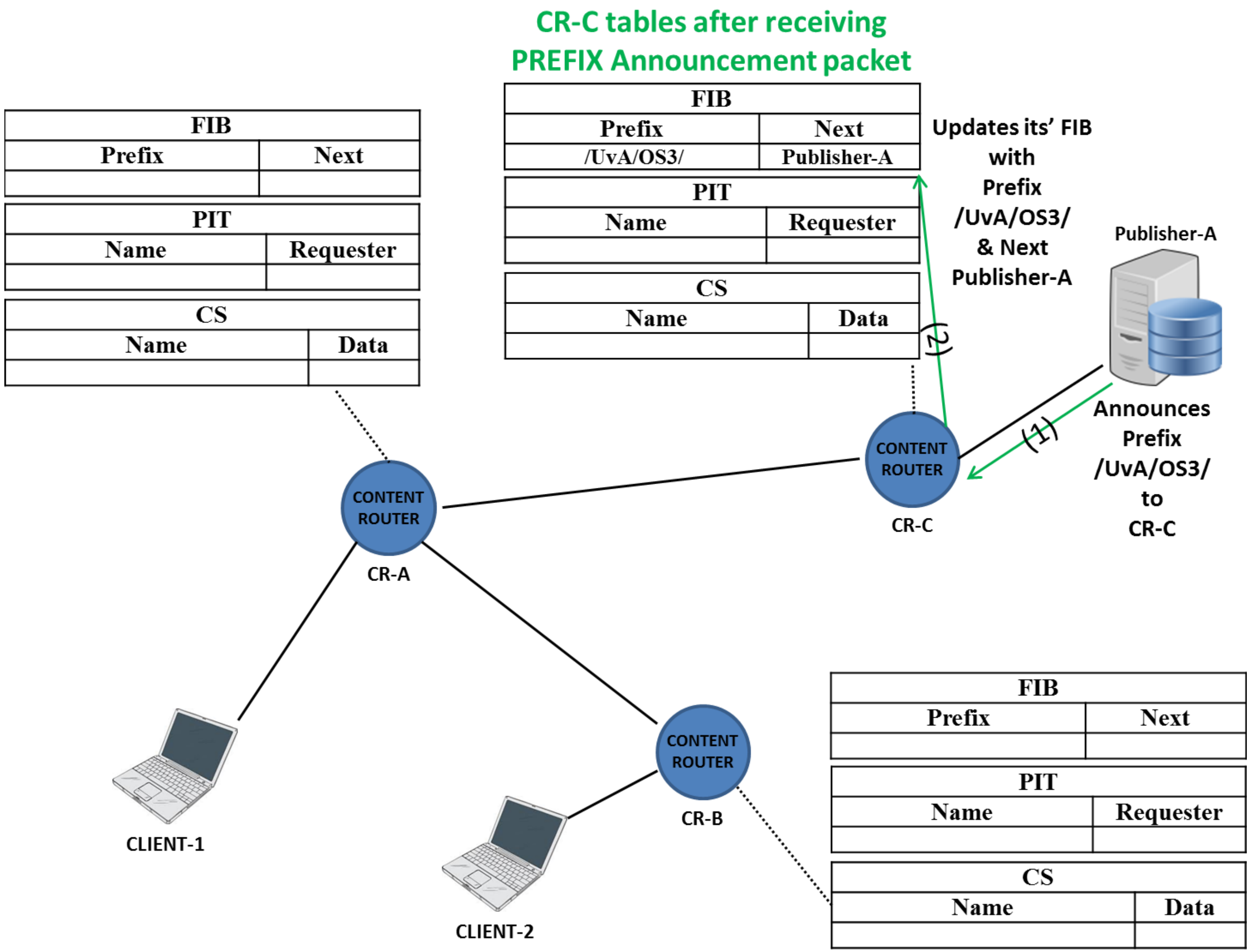
Named Data Networking(NDN)

Theoretical Studies
NDN
Populating the Name Prefix



Named Data Networking(NDN)

Theoretical Studies
NDN
Populating the Name Prefix



Named Data Networking(NDN)

CR-A tables after receiving PREFIX Announcement packets

FIB	
Prefix	**Next**
/UvA/OS3/	CR-C

PIT	
Name	**Requester**

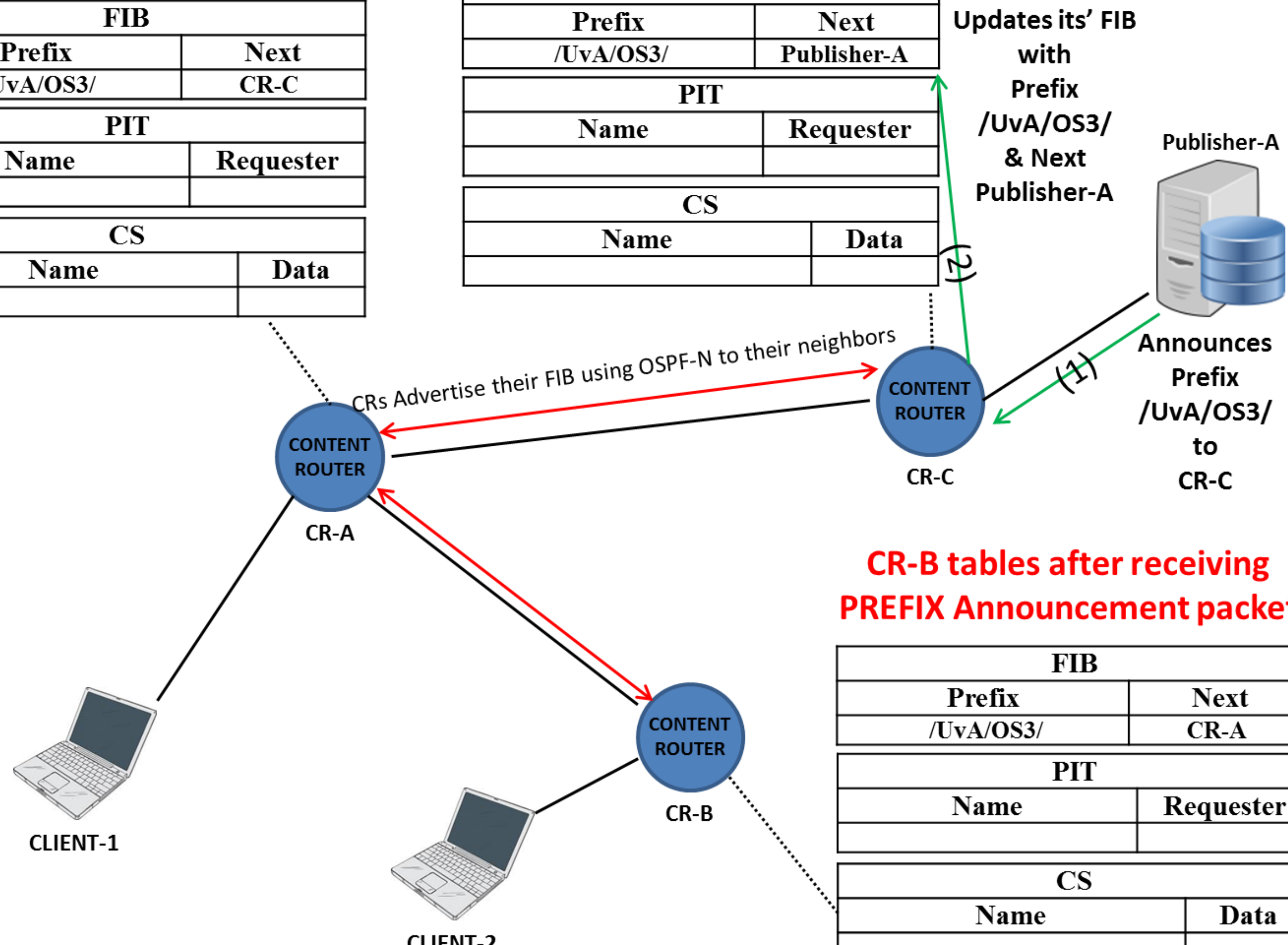
CS	
Name	**Data**

CR-C tables after receiving PREFIX Announcement packet

FIB	
Prefix	**Next**
/UvA/OS3/	Publisher-A

PIT	
Name	**Requester**

CS	
Name	**Data**



CR-B tables after receiving PREFIX Announcement packet

FIB	
Prefix	**Next**
/UvA/OS3/	CR-A

PIT	
Name	**Requester**

CS	
Name	**Data**

Named Data Networking(NDN)

Theoretical Studies
NDN
Routing the
INTEREST
packet

CR-A tables after receiving INTEREST packet

FIB	
Prefix	Next
/UvA/OS3/	CR-C

PIT	
Name	Requester
/UvA/OS3/CIA/DNS.pdf	CLIENT-1

CS	
Name	Data

CR-C tables:

FIB	
Prefix	Next
/UvA/OS3/	Publisher-A

PIT	
Name	Requester

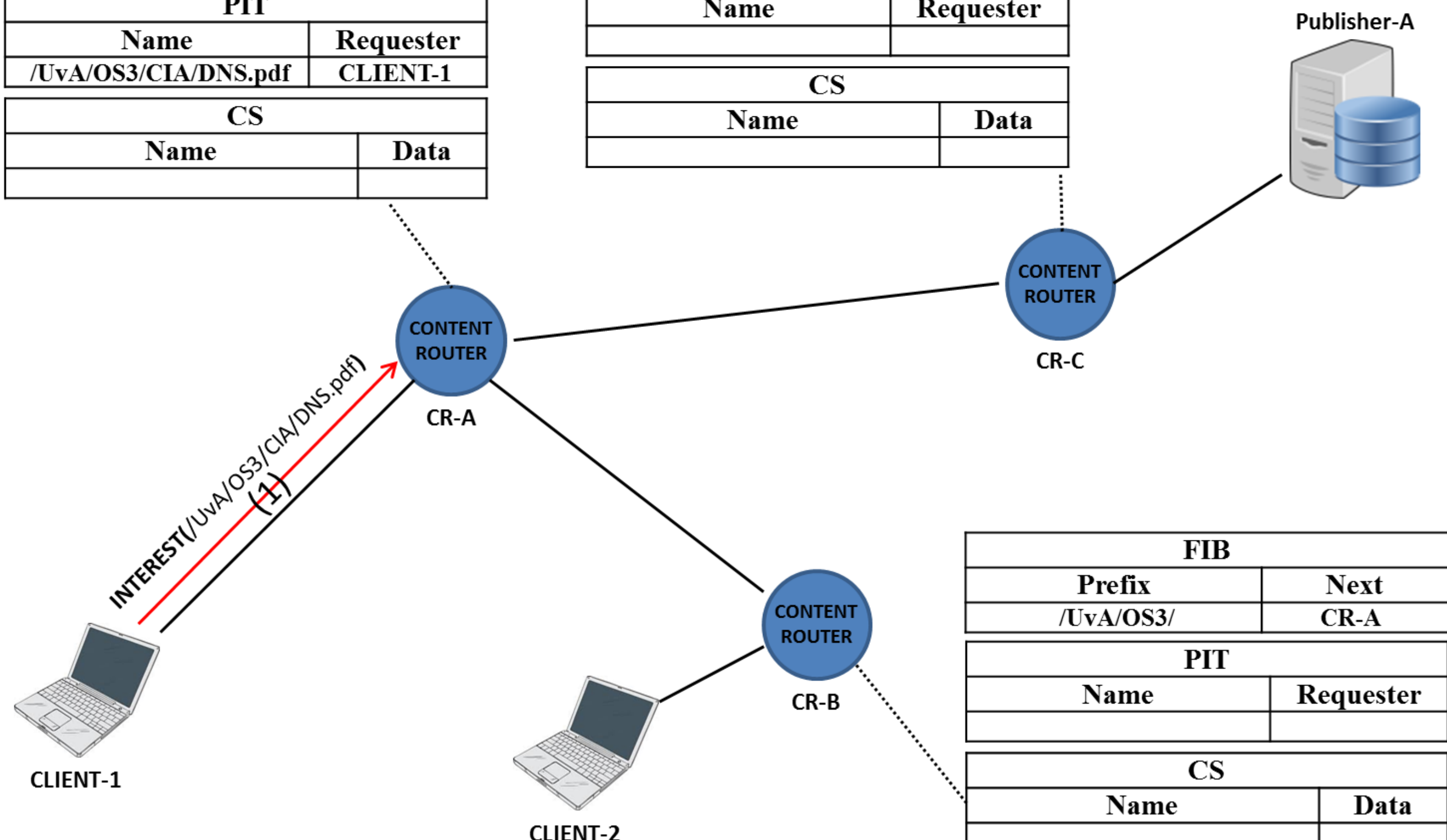
CS	
Name	Data

CR-B tables:

FIB	
Prefix	Next
/UvA/OS3/	CR-A

PIT	
Name	Requester

CS	
Name	Data



Theoretical Studies
NDN
Routing the
INTEREST
packet

Named Data Networking(NDN)

CR-A tables after receiving INTEREST packet

FIB	
Prefix	**Next**
/UvA/OS3/	CR-C

PIT	
Name	**Requester**
/UvA/OS3/CIA/DNS.pdf	CLIENT-1

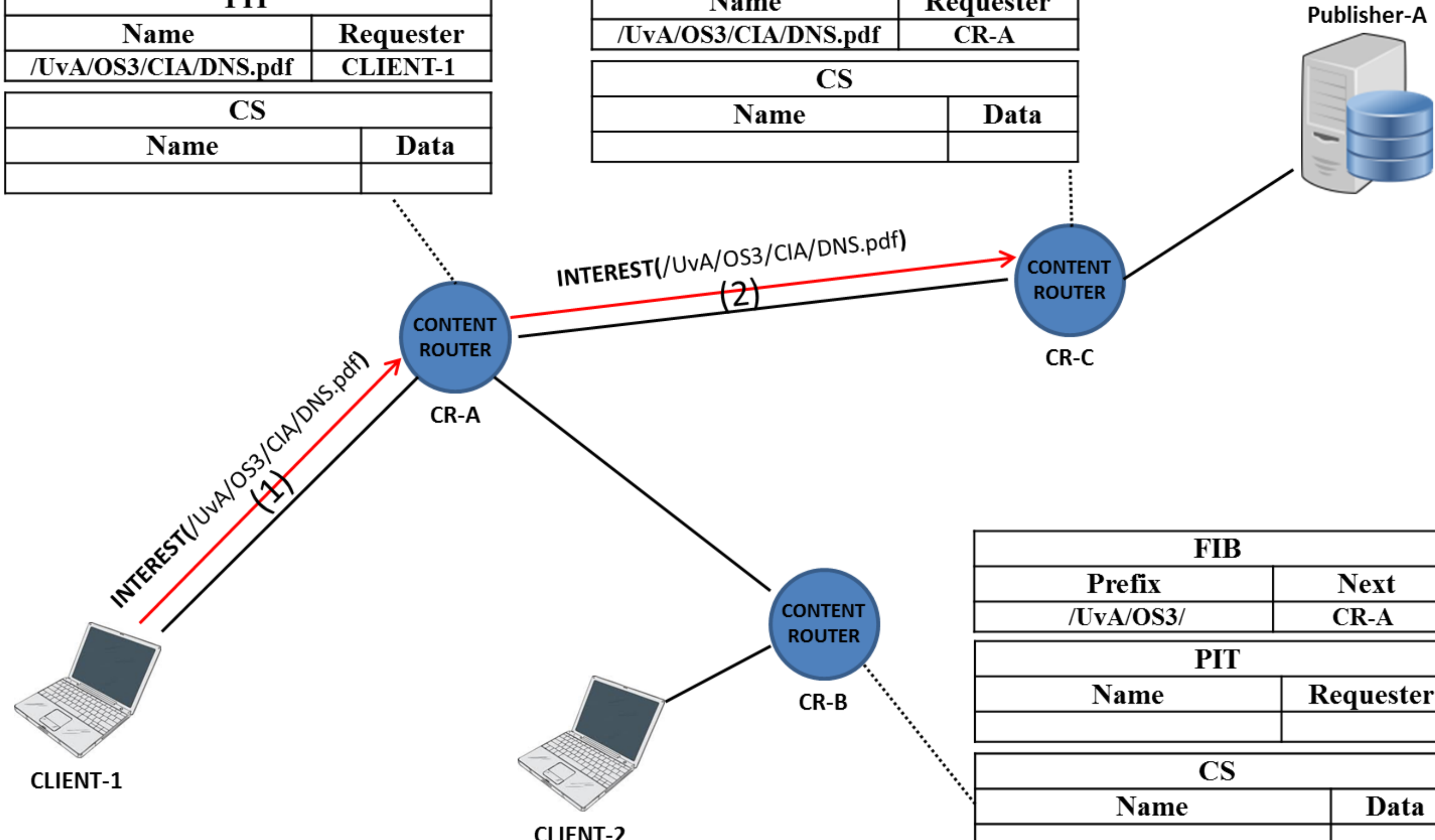
CS	
Name	**Data**

CR-C tables after receiving INTEREST packet

FIB	
Prefix	**Next**
/UvA/OS3/	Publisher-A

PIT	
Name	**Requester**
/UvA/OS3/CIA/DNS.pdf	CR-A

CS	
Name	**Data**



FIB	
Prefix	**Next**
/UvA/OS3/	CR-A

PIT	
Name	**Requester**

CS	
Name	**Data**

Named Data Networking(NDN)

Theoretical Studies
NDN
Routing the
INTEREST
packet

CR-A tables after receiving INTEREST packet

FIB	
Prefix	**Next**
/UvA/OS3/	CR-C

PIT	
Name	**Requester**
/UvA/OS3/CIA/DNS.pdf	CLIENT-1

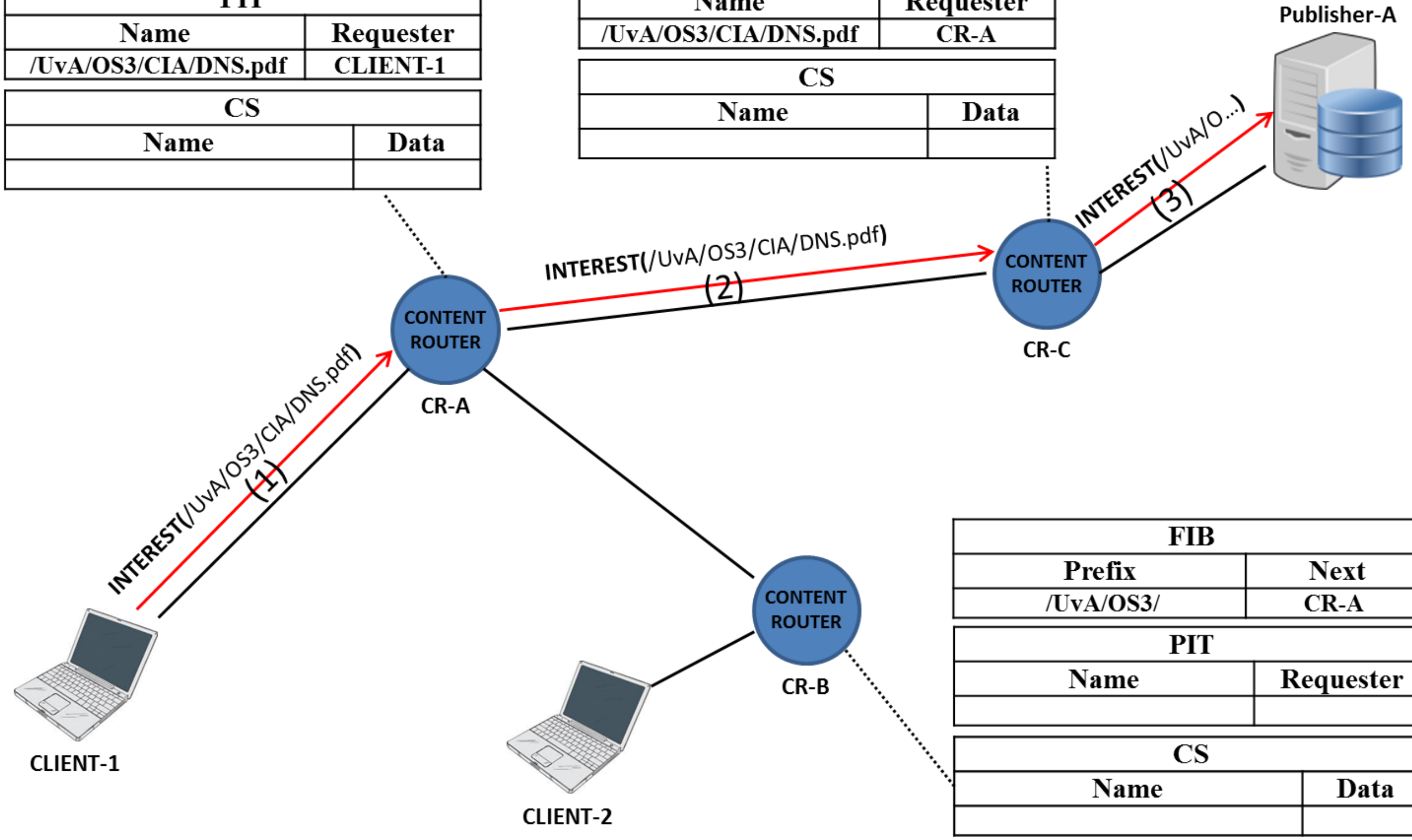
CS	
Name	**Data**

CR-C tables after receiving INTEREST packet

FIB	
Prefix	**Next**
/UvA/OS3/	Publisher-A

PIT	
Name	**Requester**
/UvA/OS3/CIA/DNS.pdf	CR-A

CS	
Name	**Data**



FIB	
Prefix	**Next**
/UvA/OS3/	CR-A

PIT	
Name	**Requester**

CS	
Name	**Data**

Named Data Networking(NDN)

CR-A tables after receiving INTEREST packet

FIB	
Prefix	**Next**
/UvA/OS3/	CR-C

PIT	
Name	**Requester**
/UvA/OS3/CIA/DNS.pdf	CLIENT-1

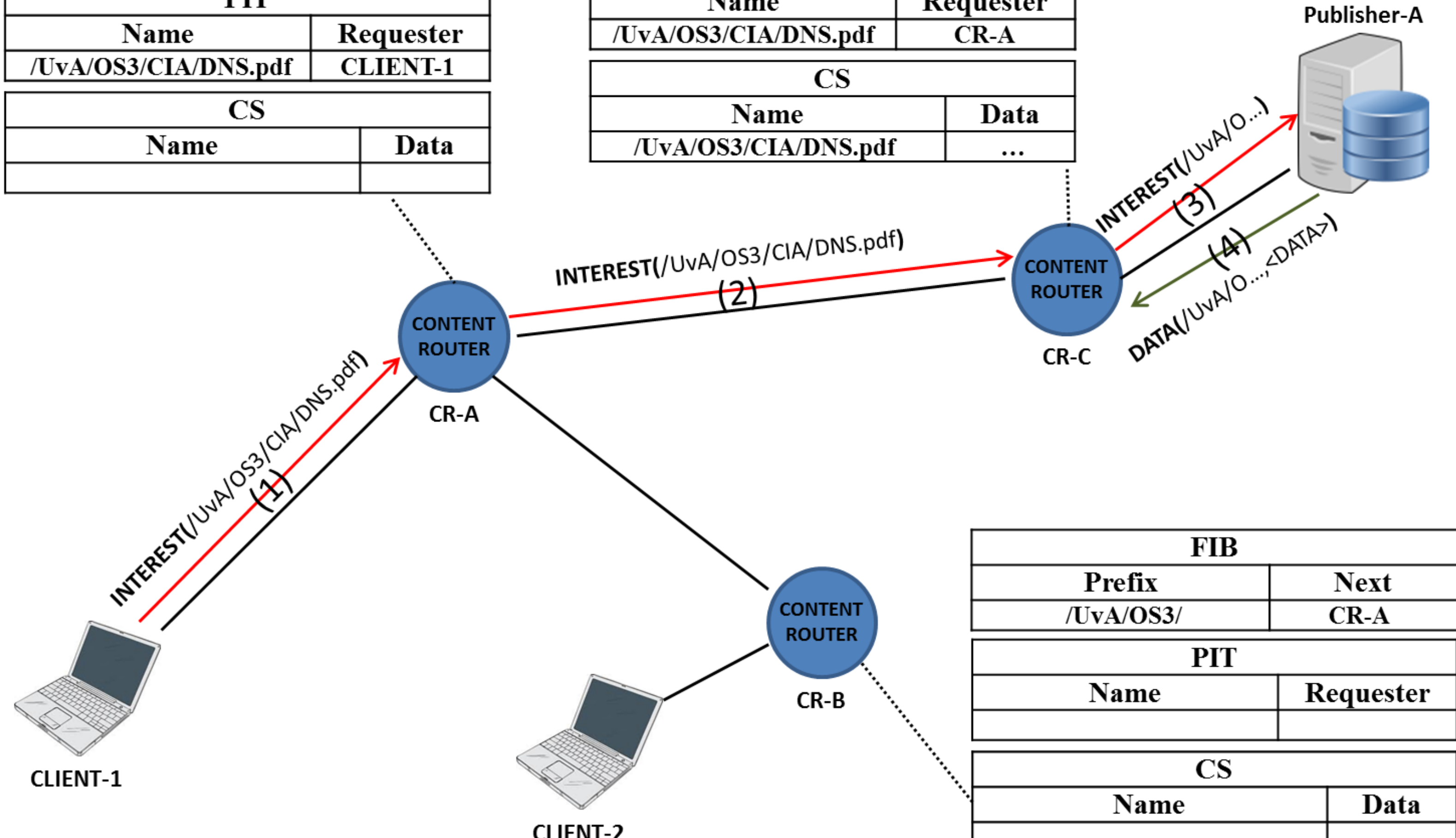
CS	
Name	**Data**

CR-C tables after receiving DATA packet

FIB	
Prefix	**Next**
/UvA/OS3/	Publisher-A

PIT	
Name	**Requester**
/UvA/OS3/CIA/DNS.pdf	CR-A

CS	
Name	**Data**
/UvA/OS3/CIA/DNS.pdf	…



FIB	
Prefix	**Next**
/UvA/OS3/	CR-A

PIT	
Name	**Requester**

CS	
Name	**Data**

Theoretical Studies
NDN
Routing the
DATA
packet

Named Data Networking(NDN)

CR-A tables after receiving DATA packet

FIB	
Prefix	**Next**
/UvA/OS3/	CR-C

PIT	
Name	**Requester**
/UvA/OS3/CIA/DNS.pdf	CLIENT-1

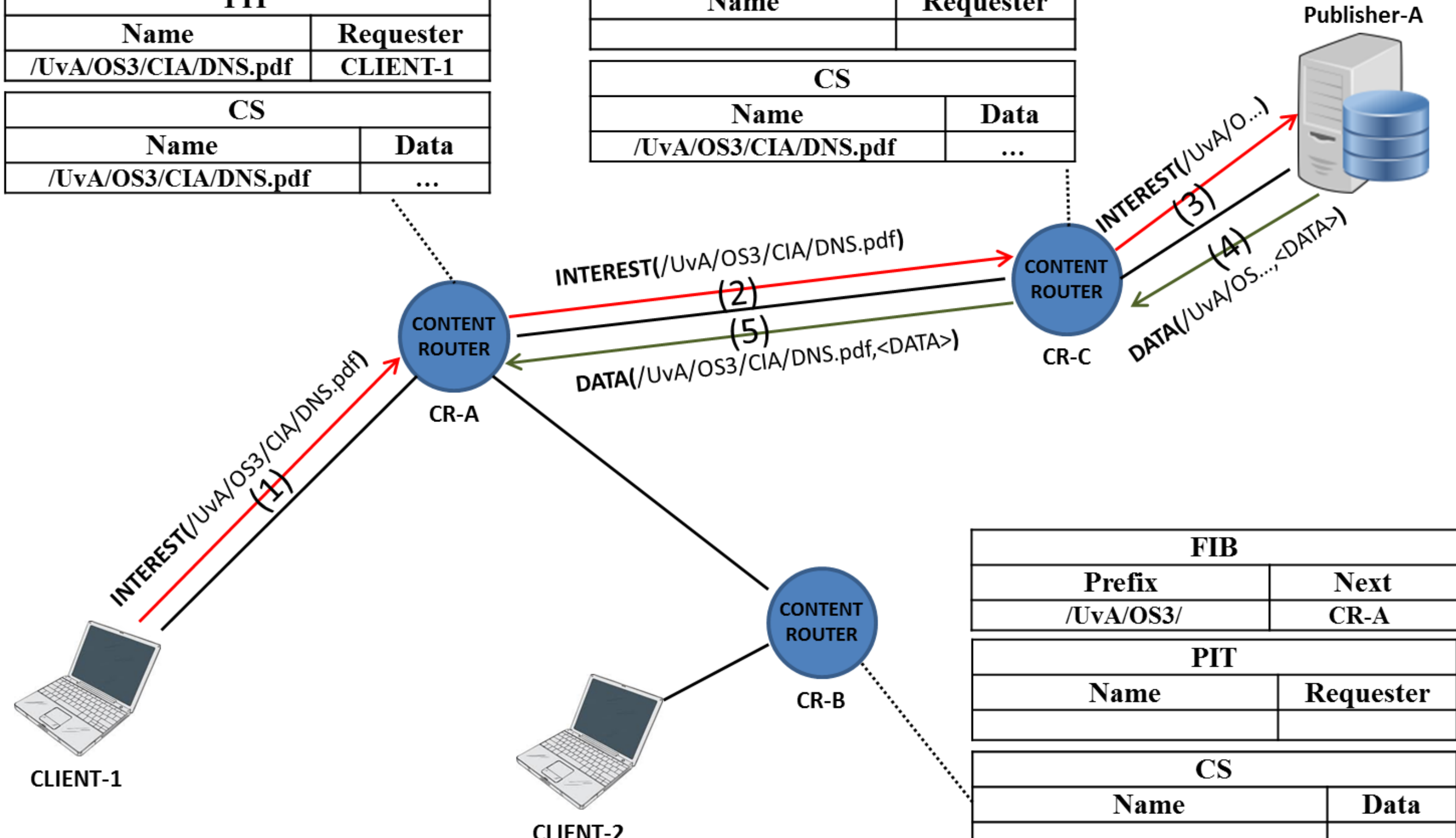
CS	
Name	**Data**
/UvA/OS3/CIA/DNS.pdf	…

CR-C tables after receiving DATA packet

FIB	
Prefix	**Next**
/UvA/OS3/	Publisher-A

PIT	
Name	**Requester**

CS	
Name	**Data**
/UvA/OS3/CIA/DNS.pdf	…



FIB	
Prefix	**Next**
/UvA/OS3/	CR-A

PIT	
Name	**Requester**

CS	
Name	**Data**

Named Data Networking(NDN)

Theoretical Studies
NDN
Routing the
DATA
packet

CR-A tables after receiving DATA packet

FIB	
Prefix	Next
/UvA/OS3/	CR-C

PIT	
Name	Requester

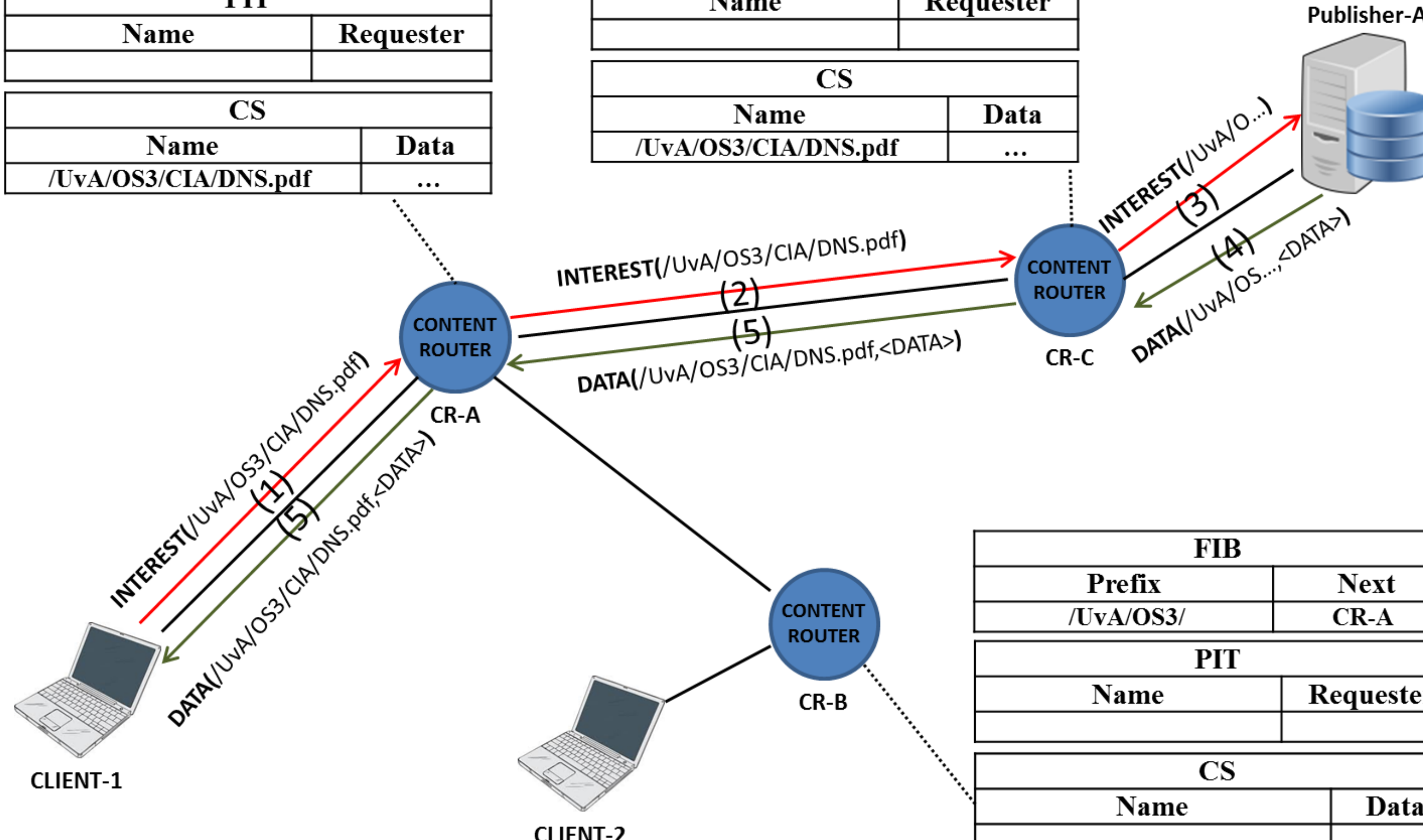
CS	
Name	Data
/UvA/OS3/CIA/DNS.pdf	…

CR-C tables after receiving DATA packet

FIB	
Prefix	Next
/UvA/OS3/	Publisher-A

PIT	
Name	Requester

CS	
Name	Data
/UvA/OS3/CIA/DNS.pdf	…



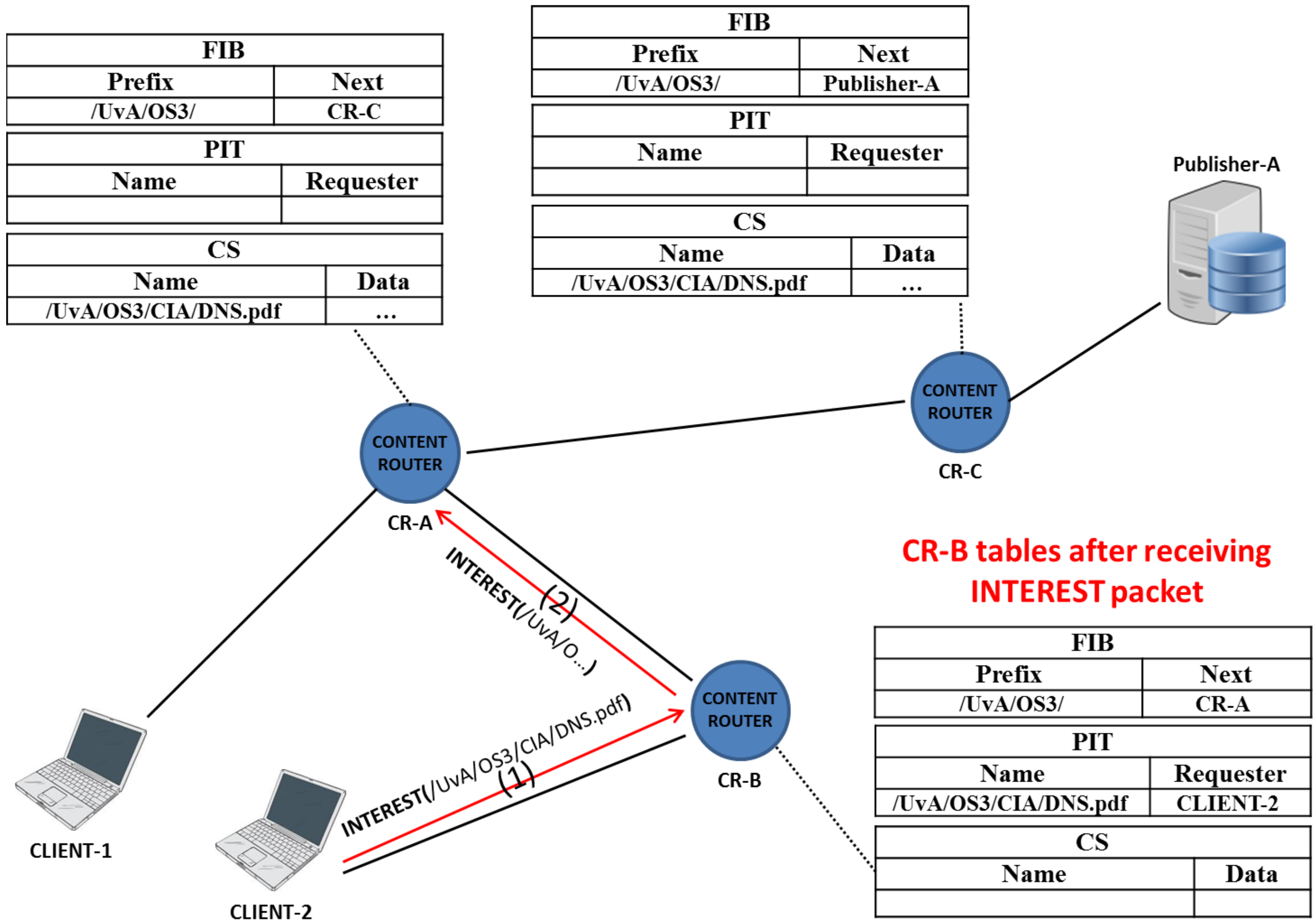
FIB	
Prefix	Next
/UvA/OS3/	CR-A

PIT	
Name	Requester

CS	
Name	Data

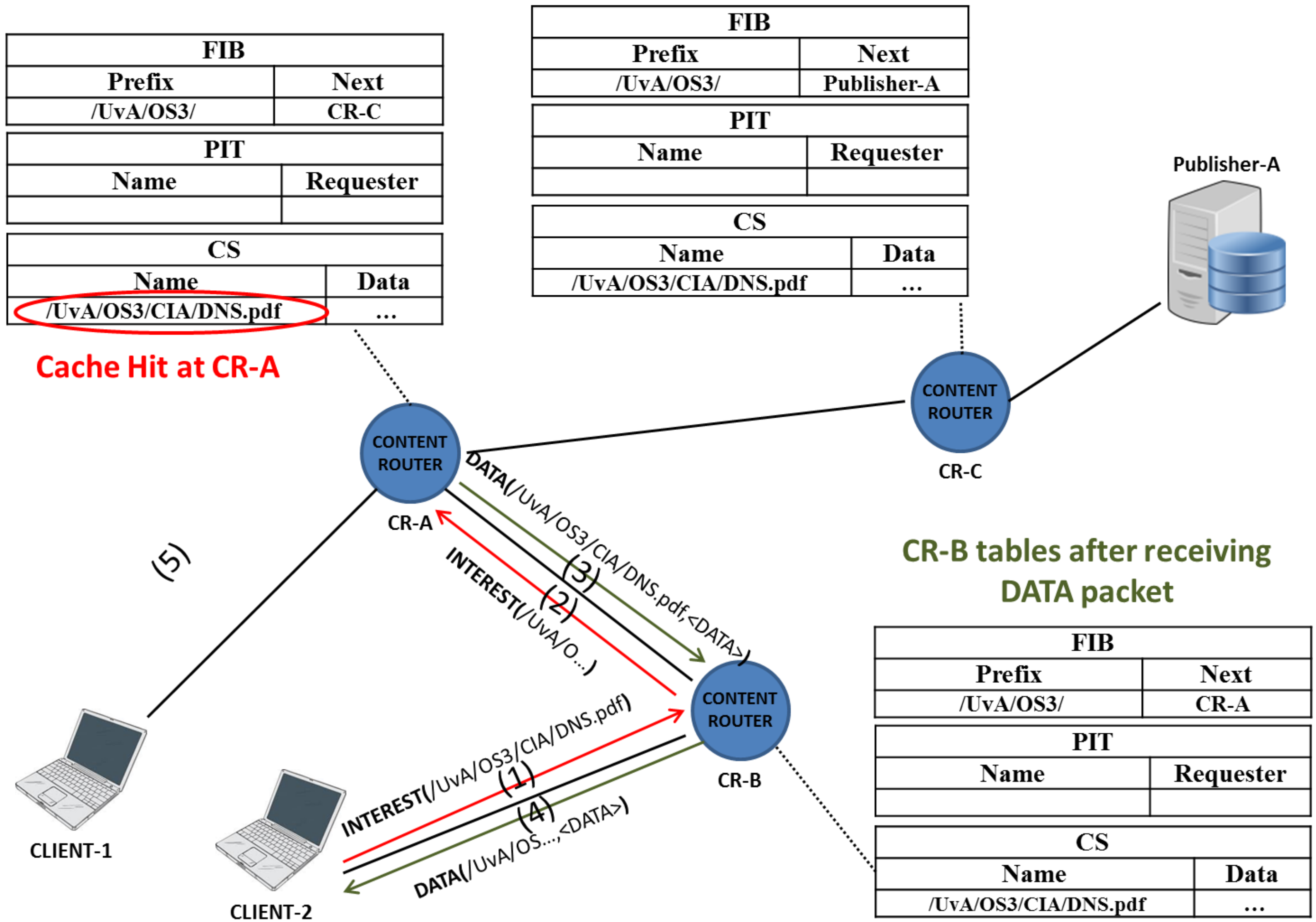
Named Data Networking(NDN)

Theoretical Studies
NDN
Cache HIT



Named Data Networking(NDN)

Theoretical Studies
NDN
Cache HIT



CR-A:

FIB	
Prefix	Next
/UvA/OS3/	CR-C

PIT	
Name	Requester

CS	
Name	Data
/UvA/OS3/CIA/DNS.pdf	…

CR-C:

FIB	
Prefix	Next
/UvA/OS3/	Publisher-A

PIT	
Name	Requester

CS	
Name	Data
/UvA/OS3/CIA/DNS.pdf	…

CR-B:

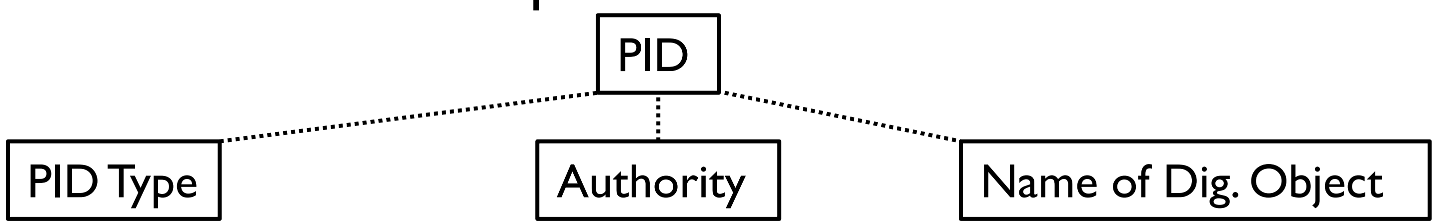
FIB	
Prefix	Next
/UvA/OS3/	CR-A

PIT	
Name	Requester

CS	
Name	Data
/UvA/OS3/CIA/DNS.pdf	…

Persistent Identifiers(PIDs)

- A name with specific syntax that uniquely identifies an object for a long-lasting period regardless of its' location and lifespan.
- Different PID types are available for naming digital objects.
- Each PID has three parts:



Unique Identifier of the PID Type(e.g.urn:,ark:)

A Unique Identifier of the Authority(e.g. isbn,ietf)

A Unique Identifier of the Digital Object (e.g. 0-7645-2641-3)
Further Delegation to sub-Authorities is possible

Example : urn:isbn:0-7645-2641-3

Theoretical Studies
PID Standards

Persistent Identifiers(PIDs)

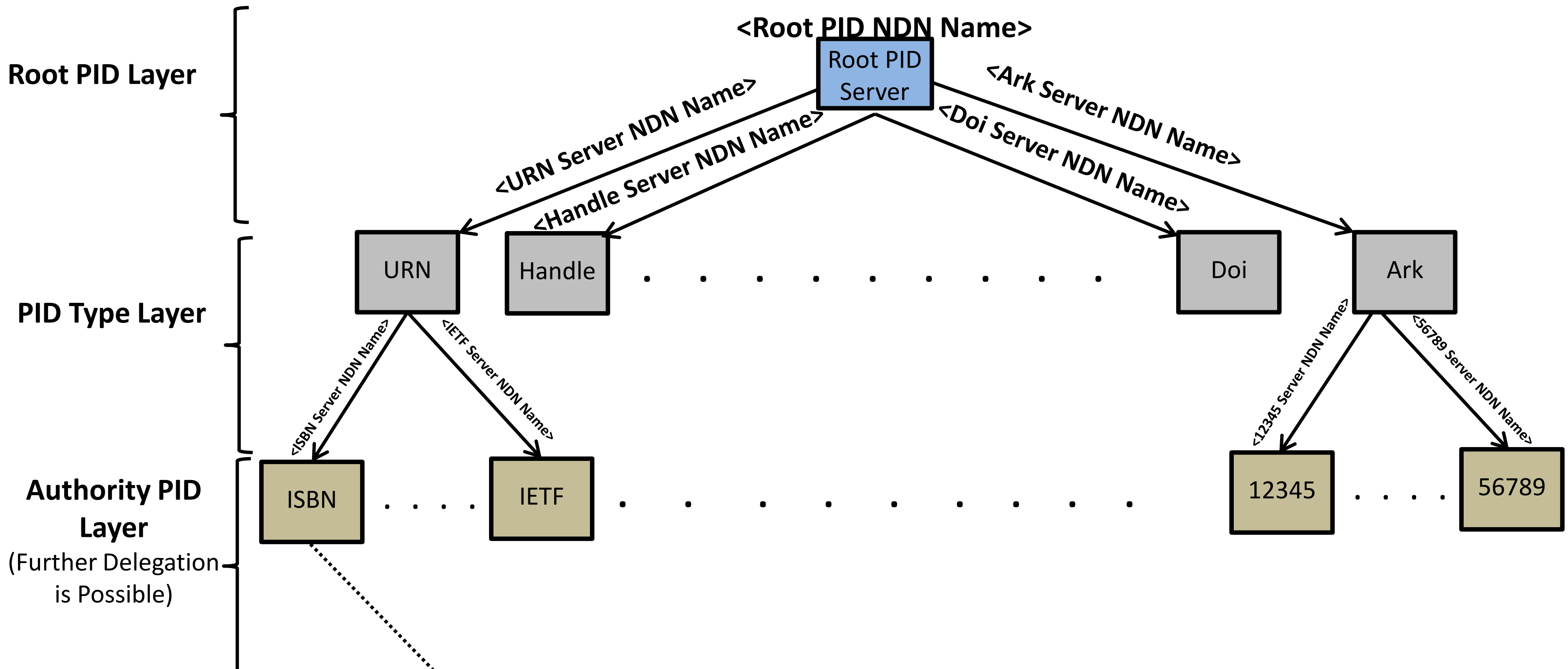
Most-well known PID Types

PID Types	PID Type Identifier	Authority	Name
URL	url:	<protocol><host>:<port>	[/<path>[?<searchpart>]]
URN	urn:	<NID>:	<NSS>
ARK	ark:	<NAAN>	/"<Name>[<Qualifier>]
HANDLE	handle:	<Handle Naming Authority>	/<Handle Local Name>
PURL	purl:	<protocol><resolver address>	/<name>
DOI	doi:	10.<Naming Authority>	/<doi name syntax>

Mapping Architecture Design Goals

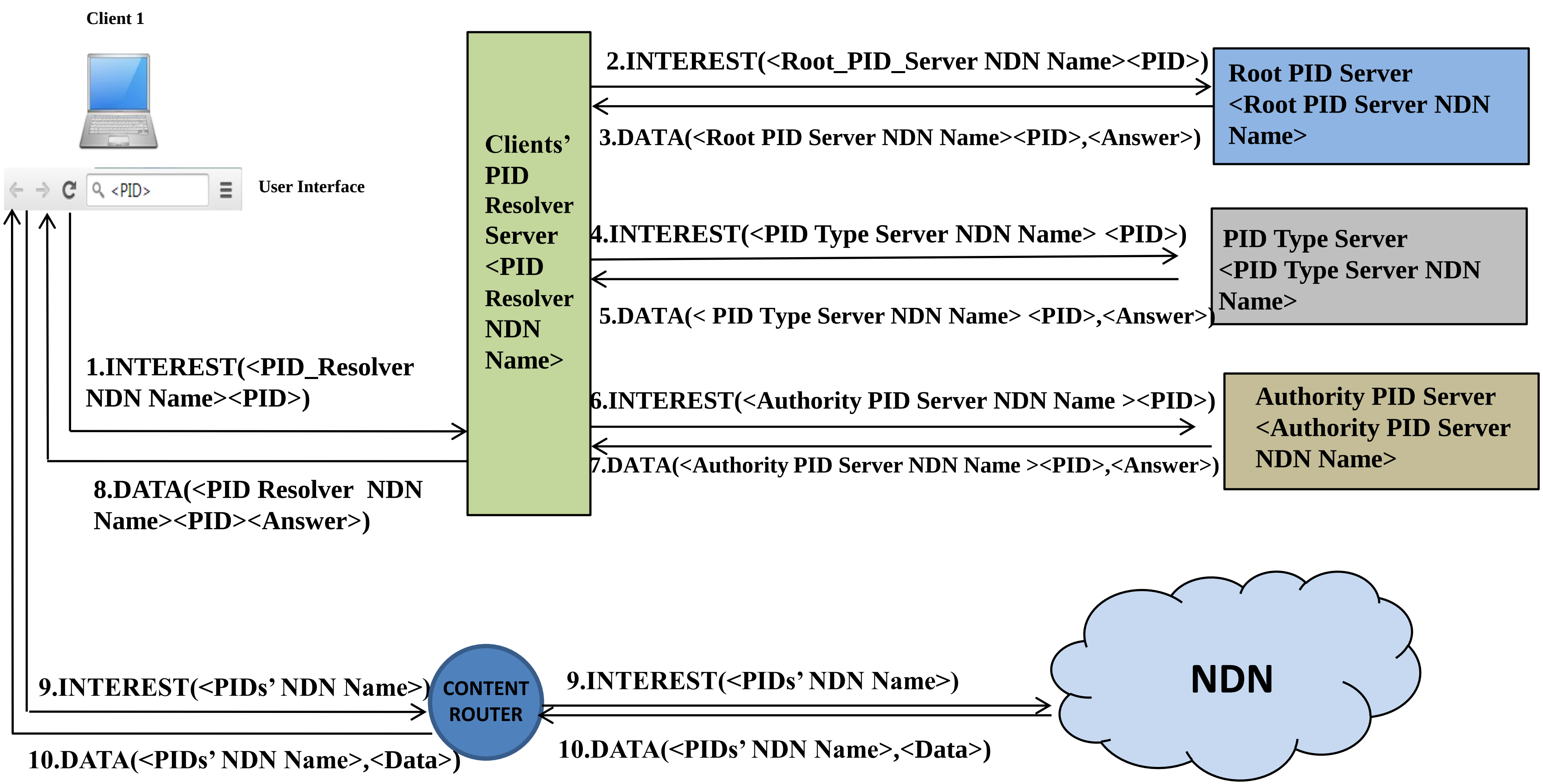
- Generic
- Extensible
- Scalable
- Easy to Implement, Manage & Administrate

Mapping Architecture Name-Space Implementation



PID	NDN-Name
urn:isbn:0-7645-2641-3	/UvA/NaturalScience/CS/CIA/DNS.pdf
.	.
.	.

Iterative Resolution of PIDs to NDN names



Caching Strategies

- Decision Algorithms(DA)

  Which Content Router caches what?

  LCE,LCD,FIX(P),ProbCache

- Replacement Algorithm(RA)

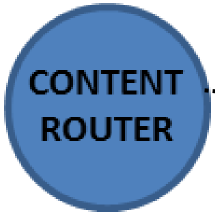
  How are Content Routers replaced Objects in the Content Store?

  FIFO,RANDOM,LRU,LFU

Simulation Parameters

Evaluate In-Network Caching Performance

Parameter	Description	Values
R	Big Data Repository Size	51.2TBytes
|R|	Num. of Big Data Objects in R	150
B	Size of Big Data Object	350GBytes
c	Num. of sub-Objects a Big Data Object is consisted of	[1,2,4,6..20]
a	Popularity of Big Data sets is based on Zipf Distribution: P(x=i)=(1/i^a)/C C= $\sum_{i=1}^{\|R\|} 1/i^\wedge a$	1



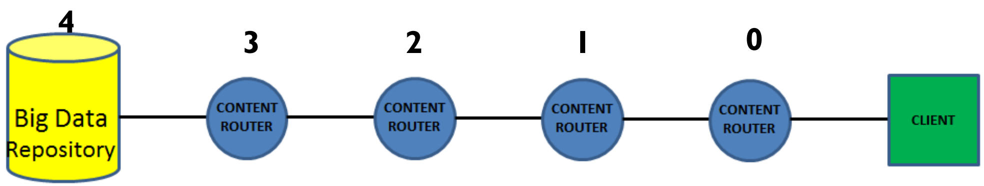
Parameter	Description	Values
C	The Content Store Size in each Content Router expressed as Size of a Big Data Object	[0.5B,1B,2B,4B,8B,16B]
CA	Caching Algorithm	[LCE,LCD,FIX(0,5),FIX (0.25),ProbCache]
RA	Replacement Algorithm	LRU

Parameter	Description	Values
T	Indicated the number of Requests for a Big Data Object the Client has send so far	-

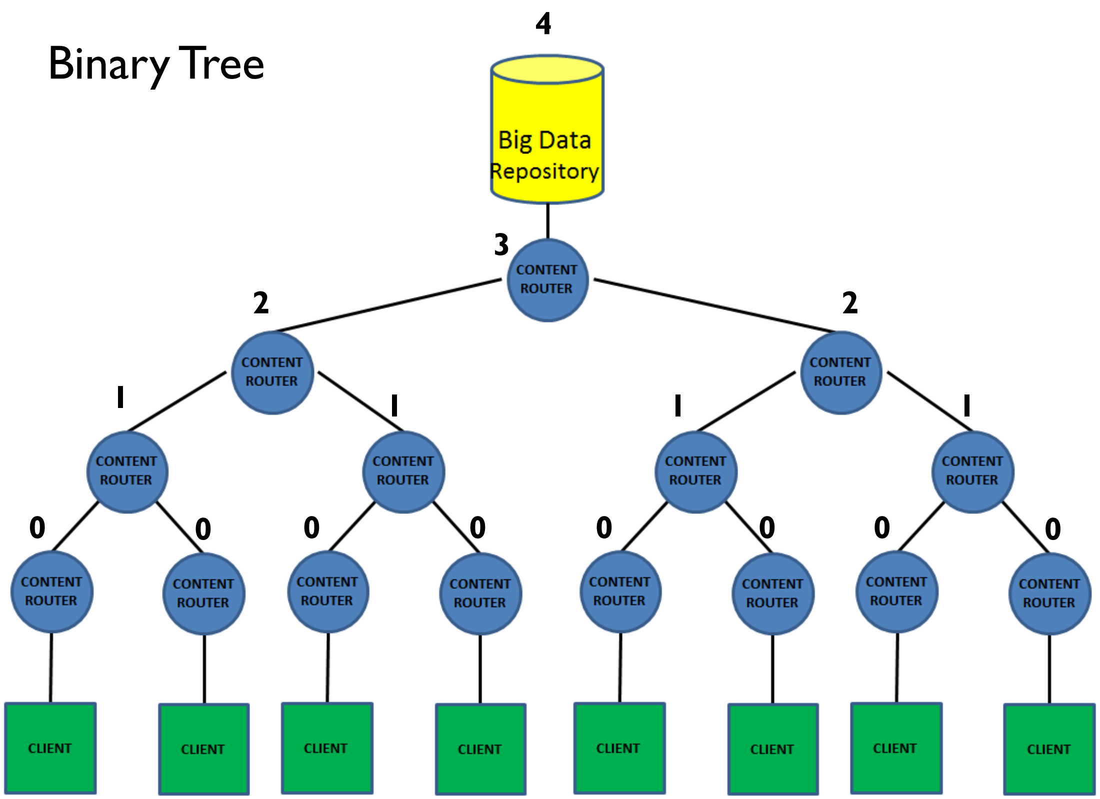
Network Topologies

Evaluate In-Network Caching Performance

String



Binary Tree



In both Network Topologies the distance between the client and the Big Data Repository is 4 Hops(Content Routers)

Performance Metrics

In ICN the in-network caching aims to:

- From the Customer point of view:

  Reduce the average time required to download the requested content.

- From the Publisher point of view:

  Reduce the number of requests the publisher needs to serve.

- From the Network point of view

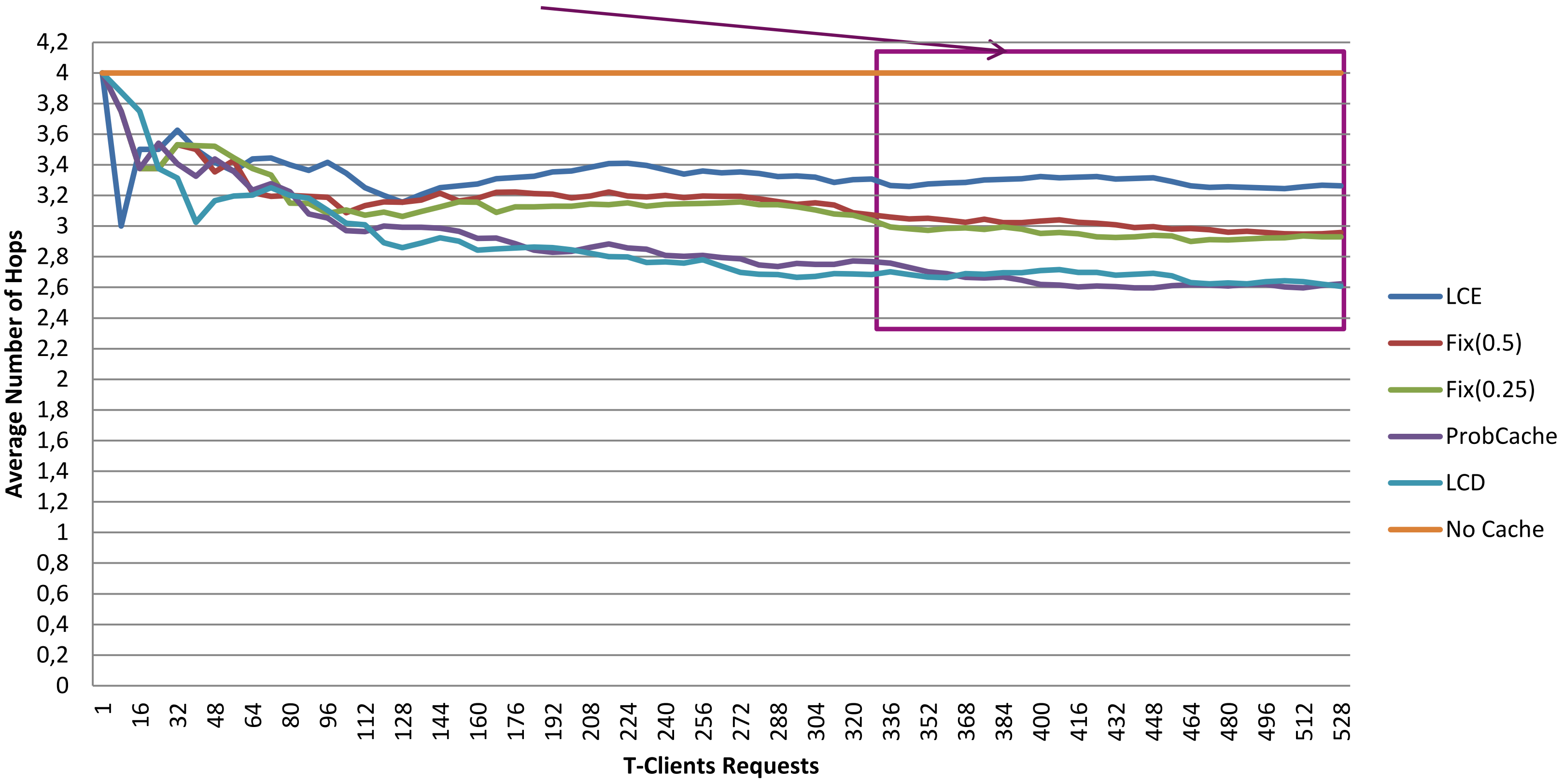
  Reduce the network traffic.

Average Number of Hops per simulation describes all the above benefits.

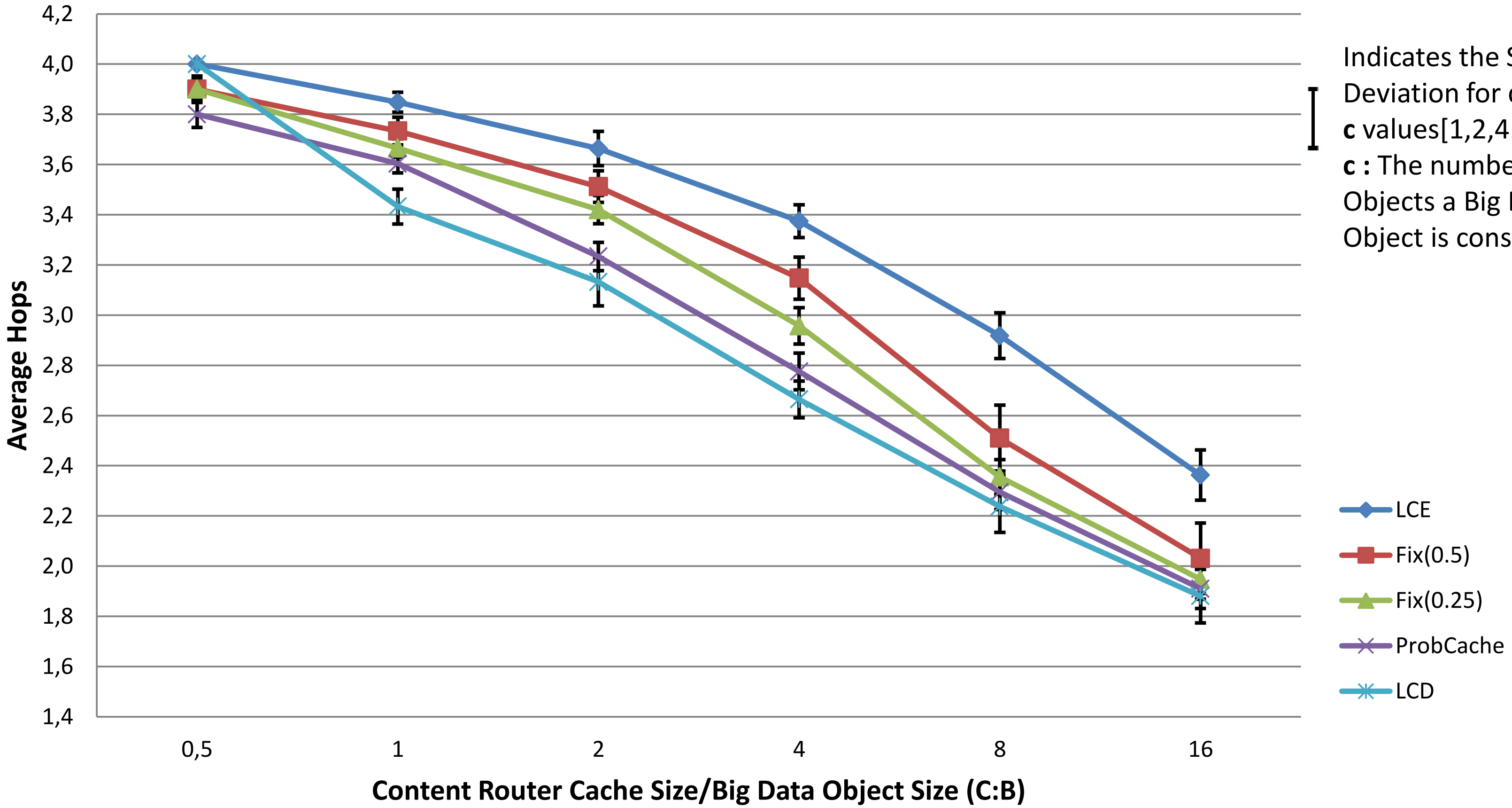
Evaluate In-Network Caching Performance

Collection of Measurements

Collection of the Average number of Hops for each simulation starts when the Average Number of Hops converges for at least 50T.



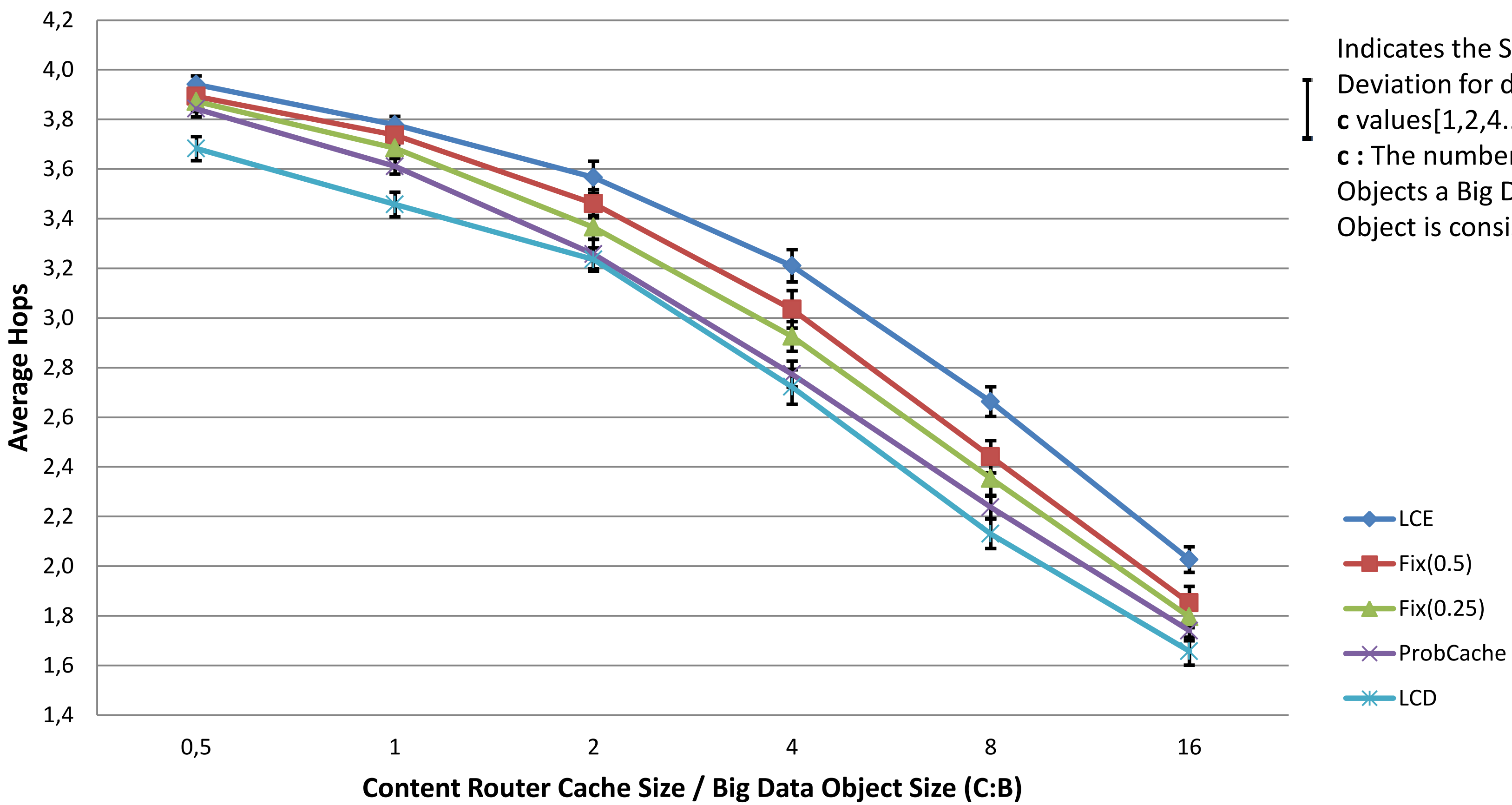
Results : String Network (1 Client)



Indicates the Standard Deviation for different **c** values[1,2,4..20].
c : The number of sub-Objects a Big Data Object is consisted of.

- Number of sub-Objects**(c)** a Big Data Object is consisted of has neglectable impact on the performance of caching algorithms.
- **C:B ≤ 1** Low Caching Algorithms Performance
- **C:B ≥ 2** Significant Benefits can be gained from this point and onwards.

Results : Binary Tree Network (8 Clients)



Indicates the Standard Deviation for different **c** values[1,2,4..20].
c : The number of sub-Objects a Big Data Object is consisted of.

- Number of sub-Objects**(c)** a Big Data Object is consisted of has neglectable impact on the performance of caching algorithms.
- **C:B ≤ 1** Low Caching Algorithms Performance
- **C:B ≥ 2** Significant Benefits can be gained from this point and onwards.

Conclusion

- **Based on our research in ICN approaches & PID Standards, mapping PIDs to ICN Names is possible**
  - Decentralized Solution Proposed for NDN approach.
    - Generic
    - Extensible
    - Scalable
    - Administration & Management is needed on each Layer
- **Evaluation of Caching Algorithms gave us**
  - Cache Size/Big Data Set Size(**C:B**), plays critical role on the efficiency of current caching algorithms.
    - C:B ≤ 1 Insignificant gain from Caching.
    - C:B ≥ 2 Significant Benefits can be gained from this point and onwards.
  - Number of sub-Object the Big Data Object is segmented does not significantly affect the efficiency of caching algorithms.

Questions?