Information Centric Networking (ICN)

Giulia Mauri
PhD Student at Politecnico di Milano
giulia.mauri@polimi.it
http://home.deib.polimi.it/gmauri

World Wide Web

Past Present Future

Network of Nodes

Network of Contents

Problems with information distribution today

- CDNs and P2P applications provide a service model of accessing named data objects instead of host-to-host service model.
- However, the network is not aware of data requests and data transmissions because this functionality resides in overlays only. Thus:
  - Data traffic follows sub-optimal paths.
  - Network capabilities (multicast and broadcast) are largely underutilized or not employed at all.
  - Overlays require significant infrastructure support (authentication portals, content storage, and application servers).

ICN Goal

- Define and create a simple, universal, flexible architecture that:
  - Matches today's communication problems;
  - Is at least as scalable and efficient as TCP/IP;
  - Is much more secure;
  - Is easier to manage.

- Is it possible to create a network architecture that fulfills the previous requirements and that is based on named data instead of named host?

What is ICN?

- The term Information Centric Networking (ICN) appeared around 2006, inspired by Van Jacobson's Google Tech Talk "A new way to look at Networking" [1]
- The ICN principles are:
  - The *content* itself is the key player of the future Internet;
  - The *content* is wherever there is interest in it, it goes where requested;
  - The users ask for the *content* in which they are interested and do not care form where it comes.
- ICN concepts can be applied to different layers of 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 at different layers.

[1] https://www.youtube.com/watch?v=oCZMoY3q2uM
[2] https://irtf.org/icnrg

ICNRG

- The Internet Research Task Force (IRTF) is sponsoring a research group on Information Centric Networking [1] that couples ongoing ICN research with solutions that are relevant for evolving the Internet at large.
- The ICNRG will produce a document that provides guidelines for experimental activities in the area of ICN so that different, alternative solutions can be compared consistently, and information sharing accomplished for experimental deployments.
- The ICNRG is focusing on the following short-term goals:
  - To produce a document that provides a survey of different approaches and techniques.
  - To produce a document that describes the ICN problem statement, the main concepts and research challenges in depth.
  - To define reference baseline scenarios to enable performance comparisons between different approaches.
- Such documentation could become input to IETF working groups.

[1] https://irtf.org/icnrg

Relevant ICN initiatives

- There are numerous approaches aimed at defining the reference ICN framework. Here, we list some of them and we will use as reference the Named Data Networking (NDN) project, or the Content Centric Networking (CCN) project that is similar.
  - Named Data Networking (NDN) [2], a US funded project;
  - Content Centric Networking (CCN) [3], a US funded project;
  - Data-Oriented Architecture (DONA), a project at Berkeley;
  - Publish-Subscribe Internet Routing Paradigm (PSIRP), a EU funded project, now in Publish-Subscribe Internet Technology, PURSUIT [4];
  - Network of Information (NetInf), currently in the Scalable & Adaptive Internet soLutions (SAIL) [5], a EU funded project;
  - COntent Mediator architecture for content-aware nETworks (COMET) [6], a EU funded project.

[2] http://named-data.net/
[3] https://www.ccnx.org/
[4] http://www.fp7-pursuit.eu/PursuitWeb/
[5] http://www.sail-project.eu/
[6] http://www.comet-project.org/

Main features of ICN projects

- **Naming**: Each piece of content in the network has a name. Naming can be flat, the content identifier is a cryptographic hash of a public key, or hierarchical, the content identifier is like a web URL. Usually, hierarchical names are human-readable, while flat names are not.
- **Name resolution and data routing**: These two functions can be coupled or decoupled. In the first approach, the content request is routed to the provider and the data response follows the same path. While in the decoupled approach, the path followed by the data is not restricted to be the same of the request. Moreover, the name-based routing can be unstructured, mainly performed based on flooding, or structured exploiting a tree and a distributed hash table structure.

Main features of ICN projects

- **Caching**: There are two options: on-path and off-path caching. The on-path caching stores content along the path of the request, while off-path caching exploits content stored outside the path.
- **Mobility**: User mobility is easy to support, since new requests can be sent after a handoff. While, provider mobility is harder to manage, since name resolution and data routing should be updated.
- **Security**: It is highly related to the naming structure. The human readable names need a trusted agent to authenticate the relation between a content and its name. While, self-certified names require a trusted agent to map the name to a human-readable one.

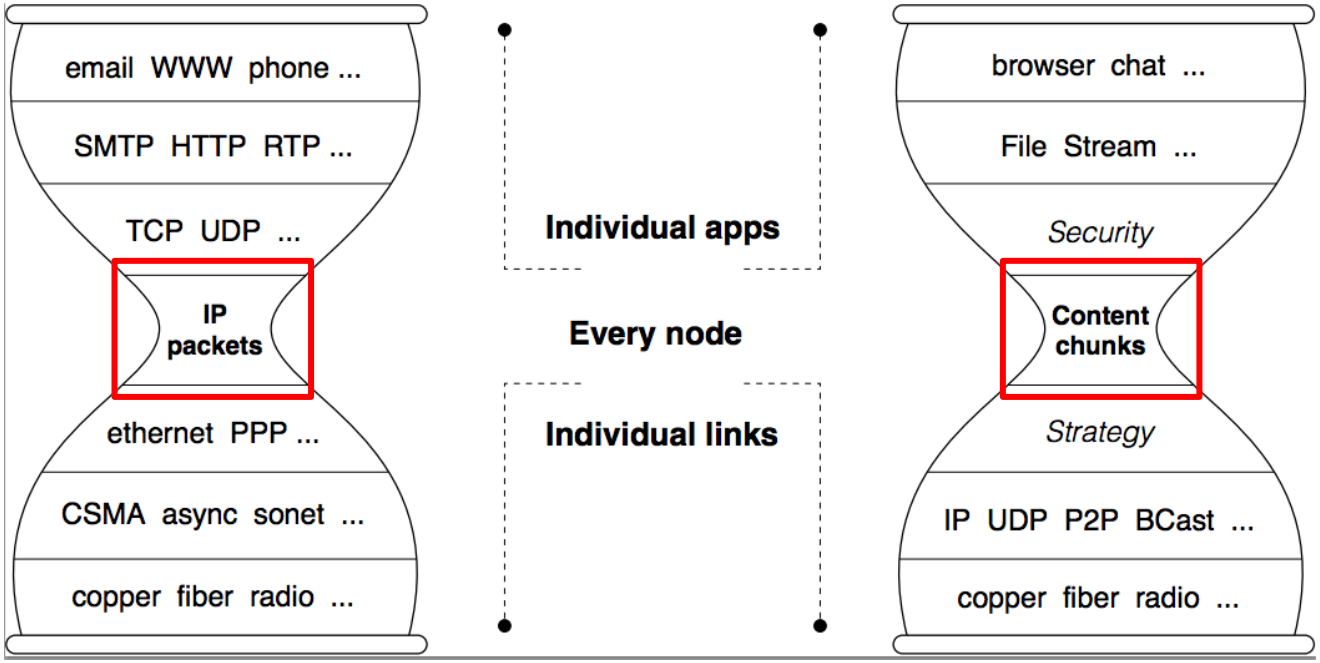
SUMMARY OF CHARACTERISTIC

	CCN/NDN	DONA	PURSUIT	SAIL	COMET
Naming	Hierarchical	Flat	Flat	Flat	Unspecified
Name Resolution and Data Routing	Coupled	Coupled and Decoupled	Decoupled	Coupled, Decoupled, and Hybrid	Coupled
Caching	On-path caching at content routers. Off-path caching with additional routing information.	On-path caching at resolution handlers. Off-path caching with additional registrations.	On-path caching difficult. Off-path caching with additional registrations.	On-path caching at content routers. Off-path caching with additional routing information or registrations.	Probabilistic on-path caching at content routers. Off-path caching with additional registrations.
Mobility	Subscriber mobility via new requests. Interest flooding protocol for publisher mobility.	Subscriber mobility via new requests. Publisher mobility requires additional registrations.	Subscriber mobility via new requests. Publisher mobility requires updating the topology manager.	Subscriber mobility via new requests. Support for publisher mobility via routing hints in hybrid operation.	Specialized mobility-aware content routers at access network that exchange mobile context state.

Named Data Networking

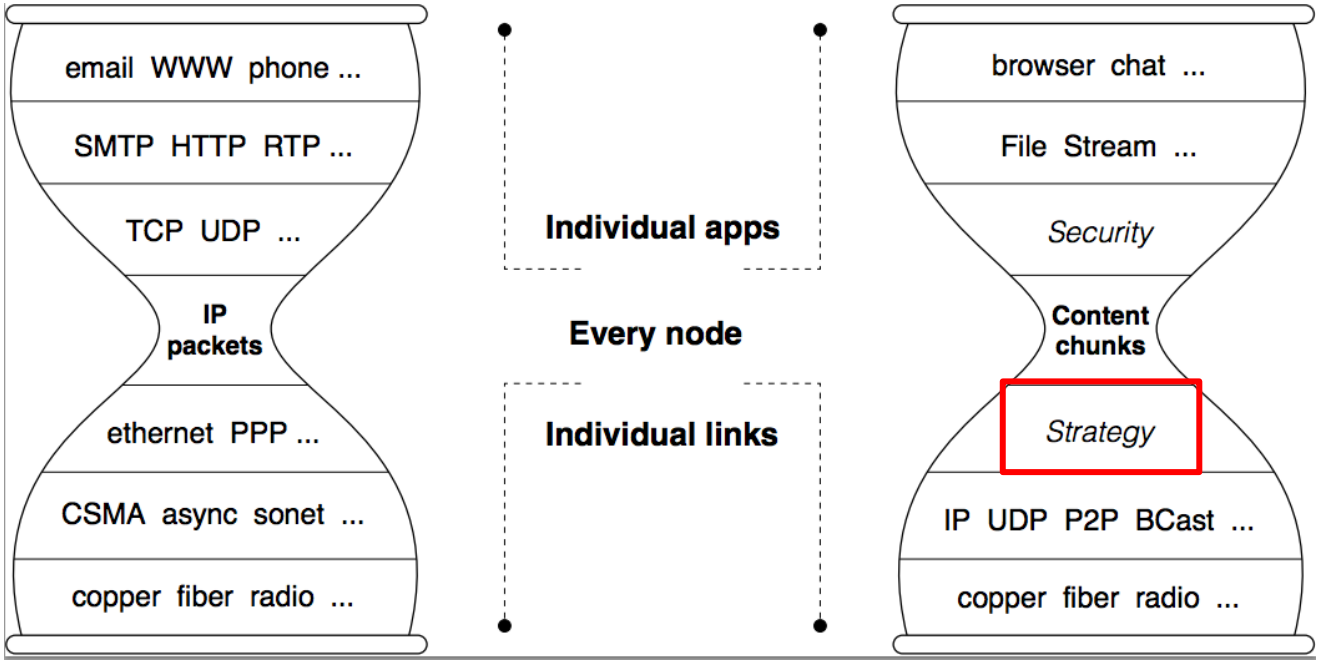
THE EVOLUTION FROM IP

- The main building blocks of the NDN architecture are **named content chunks**, in contrast to IP architecture's fundamental unit of communication, which is an end-to-end channel between two endpoints identified by IP addresses.



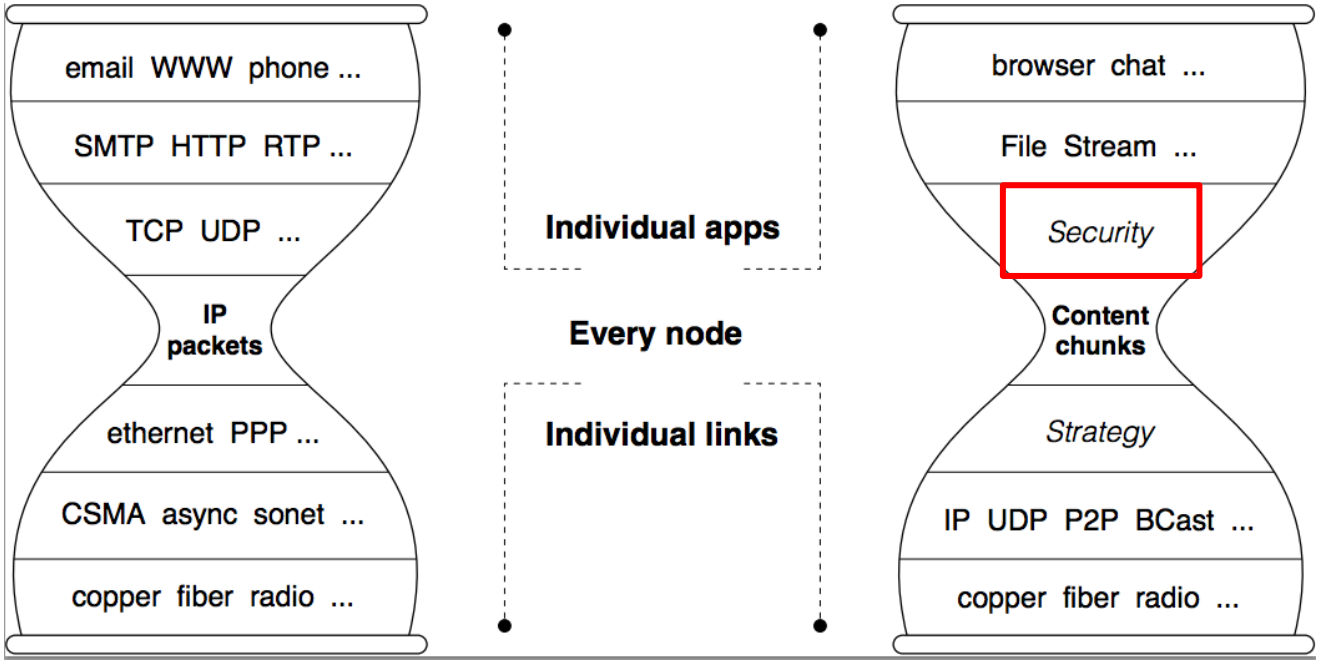
THE EVOLUTION FROM IP

- The **strategy layer** helps node to make best forwarding decision among multiple options. Which next hop to use? What to do when receiving a data?
- The **strategy layer** makes the fine-grained, dynamic optimization choices needed to best exploit multiple connectivity under changing conditions.

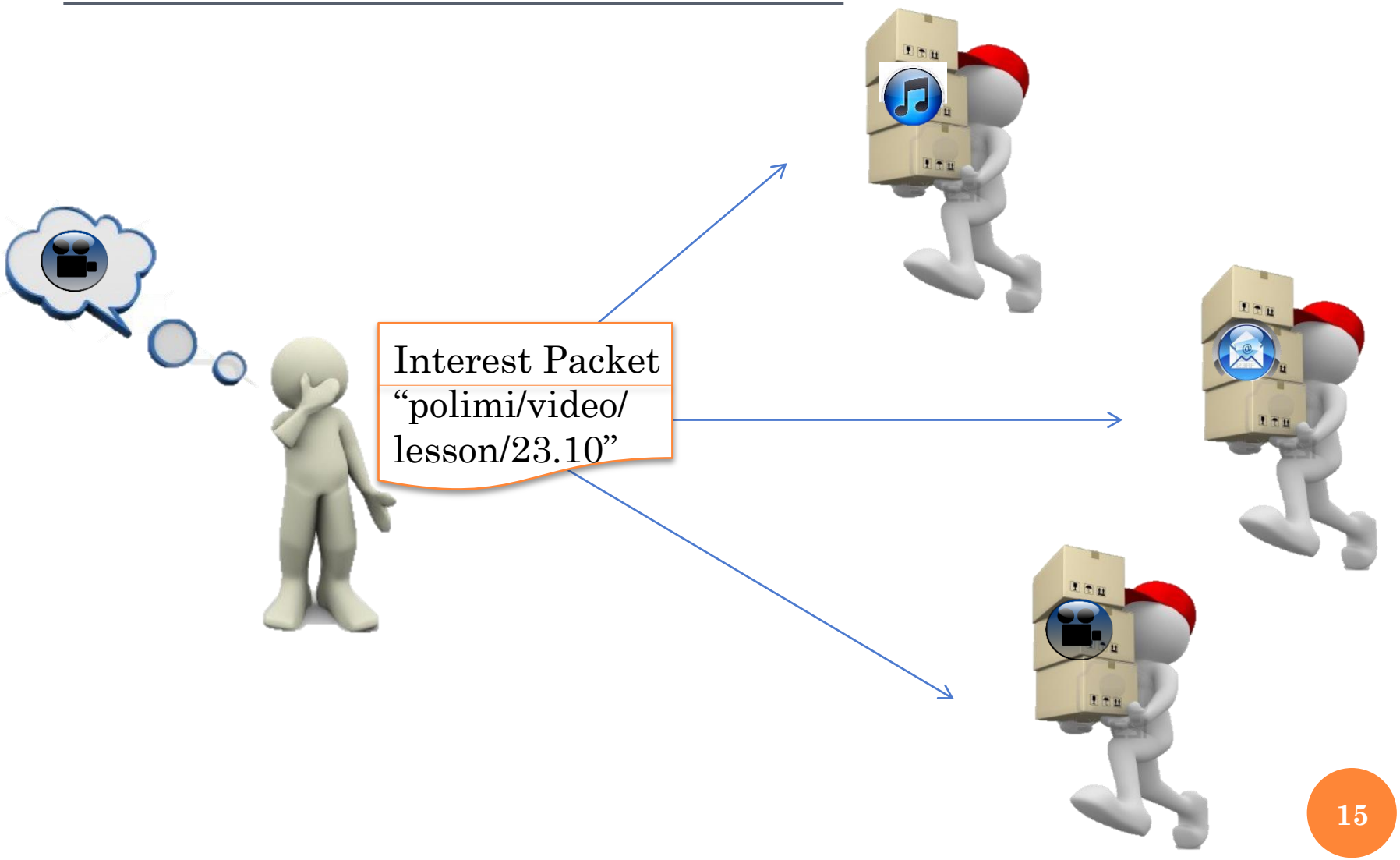


The evolution from IP

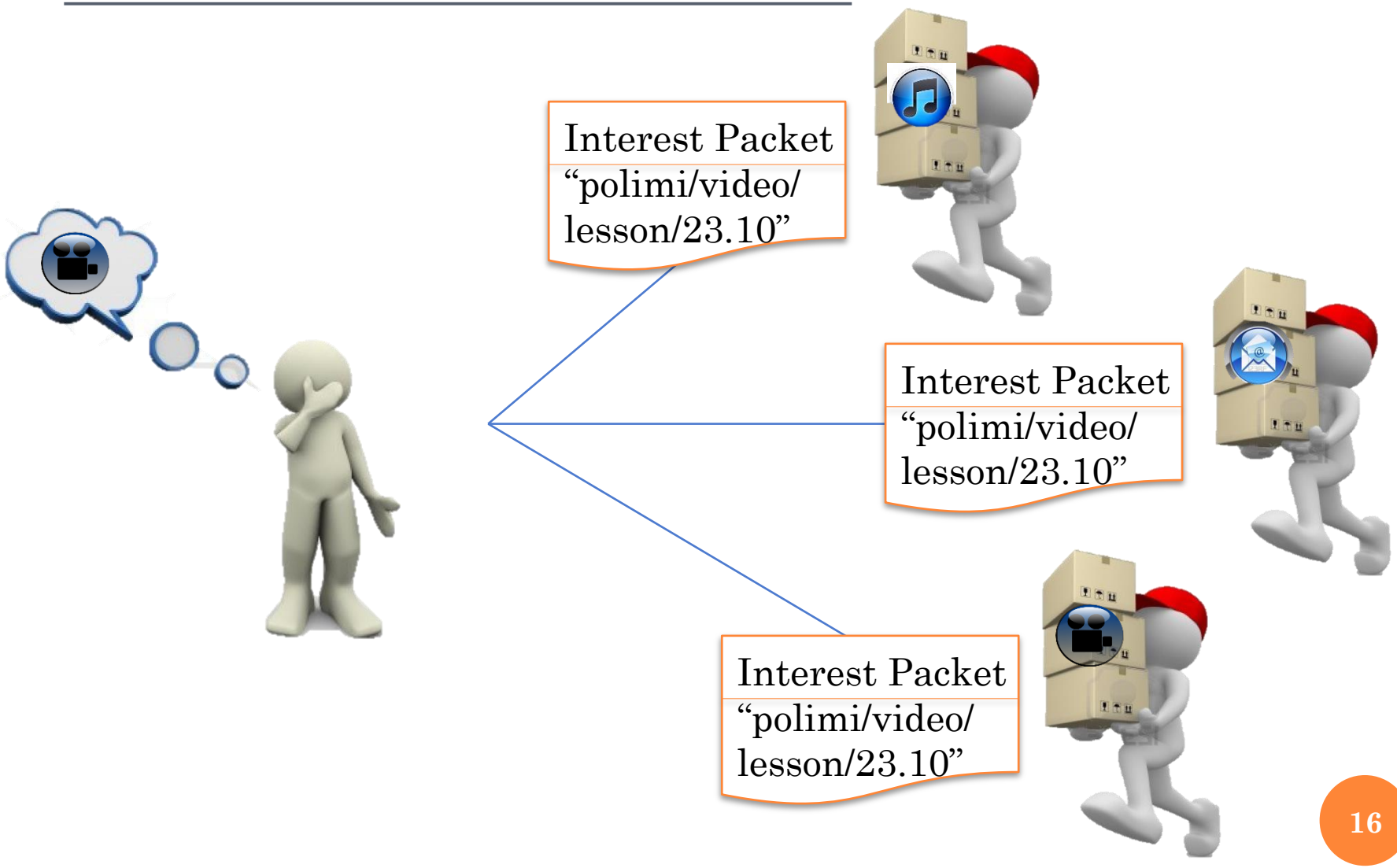
- The **security layer** enables scalable and cooperative consistency checking.
  - Each packet is authenticated and publicly verifiable;
  - The security is embedded in the content itself.



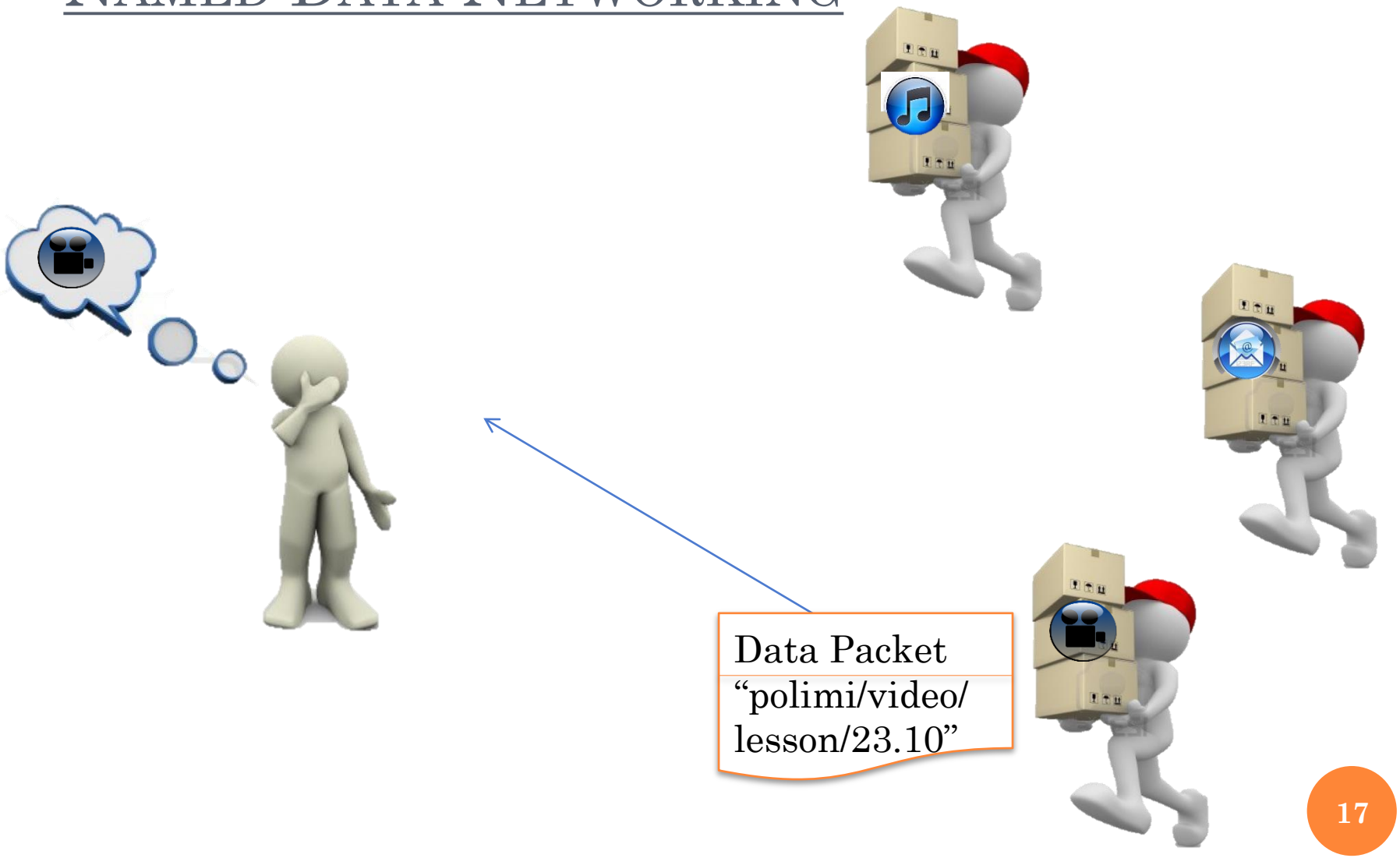
Named Data Networking



Named Data Networking



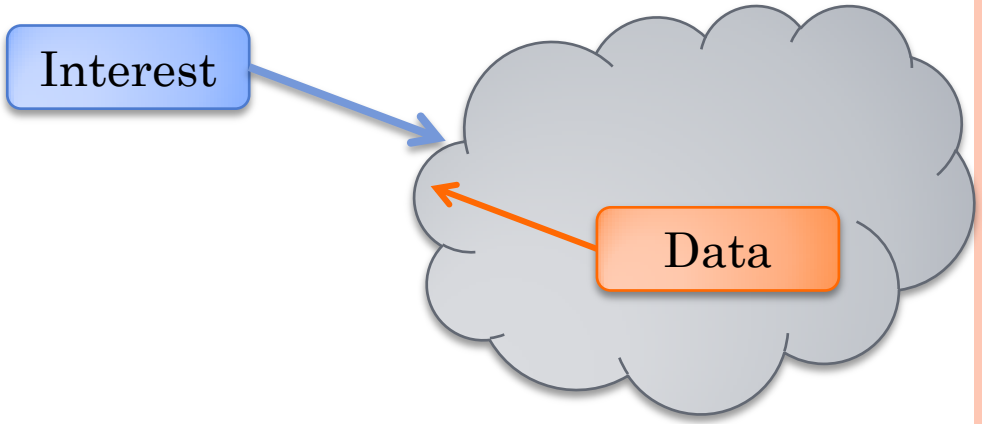
Named Data Networking



Packets in NDN

- A user asks for content by broadcasting its INTEREST over all available connectivity.
- Any node hearing the interest and having the content that *satisfies* it can respond with a DATA packet.
- A data packet *satisfies* an interest if the content NAME in the interest packet is a prefix of the content NAME in the data packet.
- Interest may be received for content that does not yet exist, allowing a publisher to generate that content on the fly in response to that query.

Packets in NDN



Interest Packet

Name
Selectors (order preference, publisher filter, exclude filter, …)
Nonce
Guiders (scope, Interest lifetime)

Data Packet

Name
MetaInfo (content type, freshness period, …)
Content
Signature (signature type, key locator, signature bits, …)

Packets in NDN- Interest

- **Name**: is a hierarchical name for NDN content, which contains a sequence of name components.
- **Selectors**:
  - *MinSuffixComponents / MaxSuffixComponents* refer to the name of name components.
  - P*ublisherPublicKeyLocator* specifies the name of the key used to sign the corresponding Data packet.
  - *Exclude* allows a consumer to choose whether to exclude list and/or ranges of name components from the responding Data packet.
  - *ChildSelector* expresses a preference for which of the matching Data within a given content store should be returned.
  - *MustBeFresh* means that the router should not answer with a Data packet from its content store whose FreshnessPeriod has expired.
- **Nonce**: is a random number that uniquely identifies the Interest packet.
- **Guiders**:
  - *Scope* limits how far the Interest may propagate.
  - *InterestLifeTime* is the time remaining before the Interest expires.

PACKETS IN NDN- DATA

- **Name**: is a hierarchical name for NDN content, which contains a sequence of name components. It must be the same of the corresponding Interest packet.
- **MetaInfo**:
  - *ContentType* could be default that is the actual data bits identified by the data name, LINK is a name that identifies the actual data content and, KEY is a public key.
  - *FreshnessPeriod* indicates how long a node should wait after the arrival of this data before marking it as stale.
  - *FinalBlockId* is equal to the last name component of the final block and indicates the final block in a sequence of fragments.
- **Content:** is the data itself.
- **Signature**:
  - *SignatureInfo* is included in the signature computation and describes the signature, signature algorithm, and other information such as the Key-Locator.
  - *SignatureValue* is excluded from signature computation and is the actual bits of the signature and other supporting information.

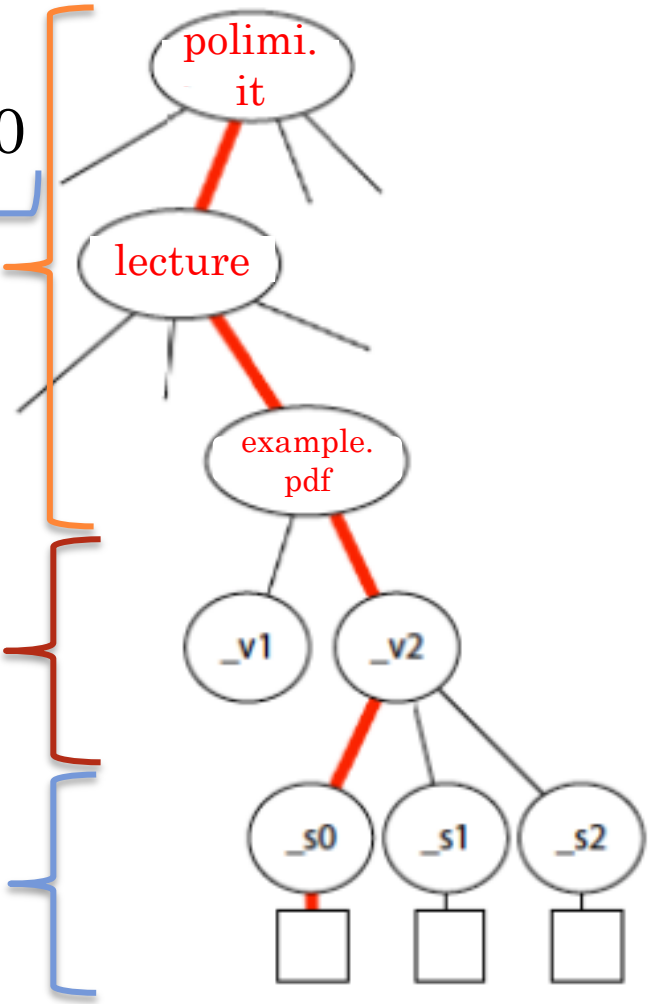
NAME

/polimi.it/lecture/example.pdf/_v2/_s0

User supplied name

Versioning w/ TimeStamp

Segmentation



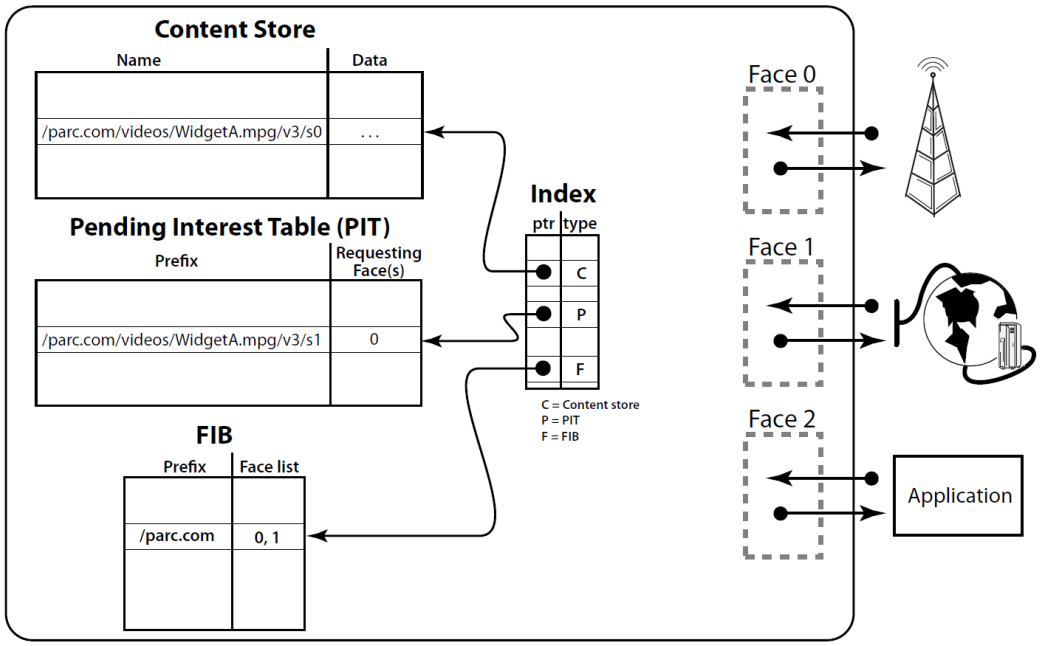
Name

- Names are opaque to the network: routers do not know the meaning of a name.
- Names are hierarchically structured. This is useful for applications to represent relationship between pieces of data. The hierarchy also allows routing to scale.
- Names do not need to be globally unique, although retrieving data globally requires a degree of global uniqueness.

- The naming system is the most important piece in the NDN architecture and still under active research.

NDN FORWARDING

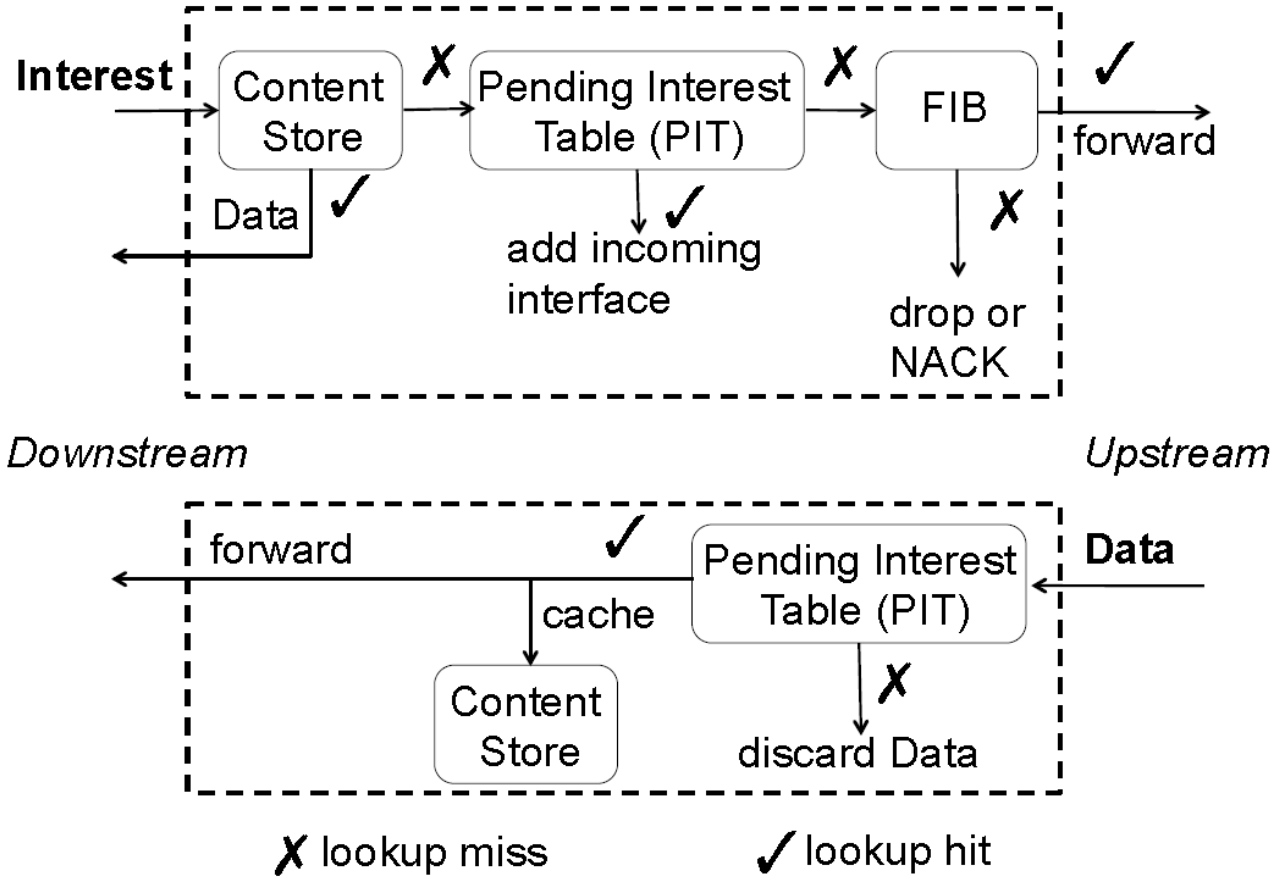
- A packet arrives on a face, a longest match lookup is done on its name, and then an action is performed based on the result of that lookup.
- The core NDN packet forwarding engine has three main data structures: CS, PIT, FIB.



NDN FORWARDING

- The **Forwarding Information Base (FIB)** is used to forward interest packets toward potential sources of matching data by registering the prefixes and the corresponding list of neighbors.
- The **Pending Interest Table (PIT)** keeps track of interests forwarded upstream toward content sources so that the returned data can be sent downstream to the requesters. PIT entries are erased as soon as they have been used to forward the data packet.
- The **Content Store (CS)** is an associative container of data. Which data is stored in a node at a given time is decided by means of a CS management policy (e.g. LRU, LFU).

NDN FORWARDING



NDN FORWARDING - INTEREST

- When a node receives an interest, it checks if there is a correspondence in its tables.

1. If the CS caches the requested data packet, the node sends out the content and drops the satisfied interest.
2. If the match is in the PIT, the corresponding entry is updated adding the requesting node and the interest is discarded.
3. If the match is in the FIB, the interest is sent out to the next hop(s) and it is created a new entry in the PIT.
4. If there is no match, the interest is discarded.

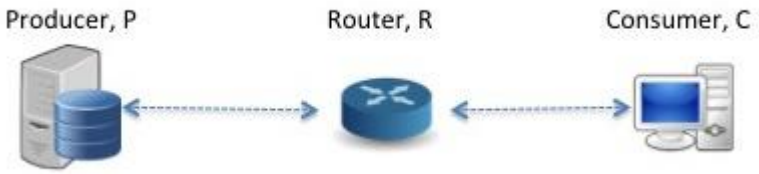
NDN FORWARDING - DATA

- The data packet processing is quite similar to the interest processing.

1. The node looks in the PIT and if there is a match, it sends the data to the requesting nodes. If not, it discards the packet.
2. Then, it stores the data packet in the Content Store.
3. A FIB match means an unrequested data, so the node drops the packet.

NETWORK ENTITIES

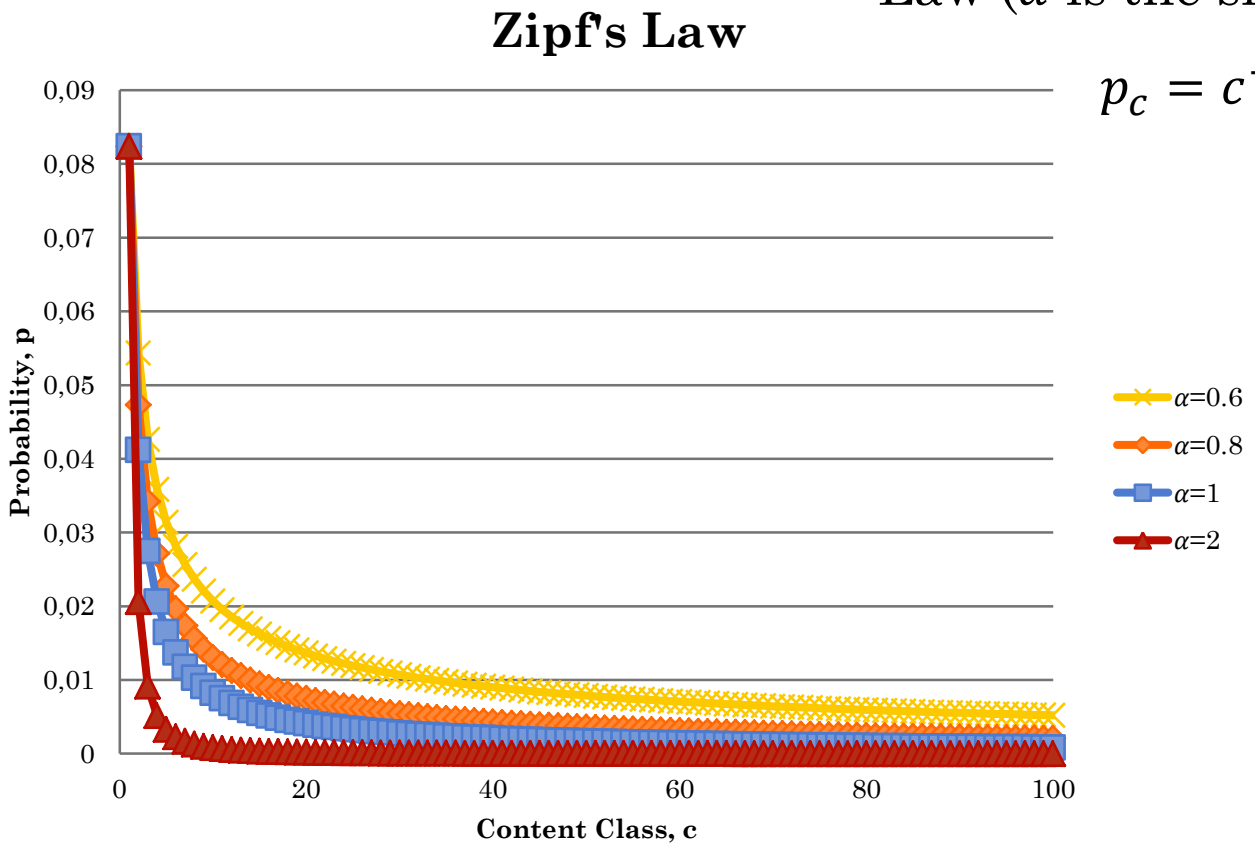
- **Data Producer**: announce name prefixes, upon reception of an Interest packet, it answers with the corresponding Data packet. It signs a content by using its key.
- **Data Router**: upon reception of an Interest packet, it answers with the corresponding Data packet, if it is present in its content store. Otherwise it forwards the request towards the correct Data Producer. Upon reception of a Data packet, it forwards it to the downstream Consumer. Moreover, it caches packet in its Content Store.
- **Data Consumer**: obtains data sending Interests with the desired data name.



POPULARITY MODEL

- Given a set of contents C, the contents are classified into K classes with respect to their popularity.
- Each content c has a probability of being requested that follows the Zipf's Law (α is the skew parameter):

$$p_c = c^{-\alpha} / \sum_{c=1}^{C} c^{-\alpha}$$



In Network Storage

- In network storage can improve network performance by fetching content from node geographically placed closer to the end-user.
- Each node has a Content Store where it caches data packets for satisfying future requests. Named data objects allow for caching at any network element: routers, proxy caches and end-host machines.
- The Content Store is like a buffer memory in IP routers; however, NDN nodes can reuse contents as long as they remain in the CS.
- There exist two approaches to in network caching: on path and off path caching. Off path caches are placed in strategic points within the network in order to improve the performance and reduce the redirection delays. While, on path caching relies on opportunistic cache hits and fits more neatly in ICN. In both cases, the cost for the implementation and deployment will be their driver.

CACHING MODEL

- There are various methods for choosing whether a content should be cached or not. When the cache is full, an algorithm must choose which items to discard to make room for the new ones. These algorithms can be distinguished in two classes: reactive and proactive.
- **Reactive protocols**:
  - *LRU*: the least recently used item is discarded first from the cache.
  - *LFU*: the content that are used least often are discarded first.
- **Proactive protocols**: the content are cached according to some pre-computation on the probability of the content to be chosen and according to the network scenario.
- The caching performance are affected by:
  - Hit ratio: describes how often a searched object is found in the cache.
  - Latency: is the delay in returning a object after receiving the request.

Data centric security

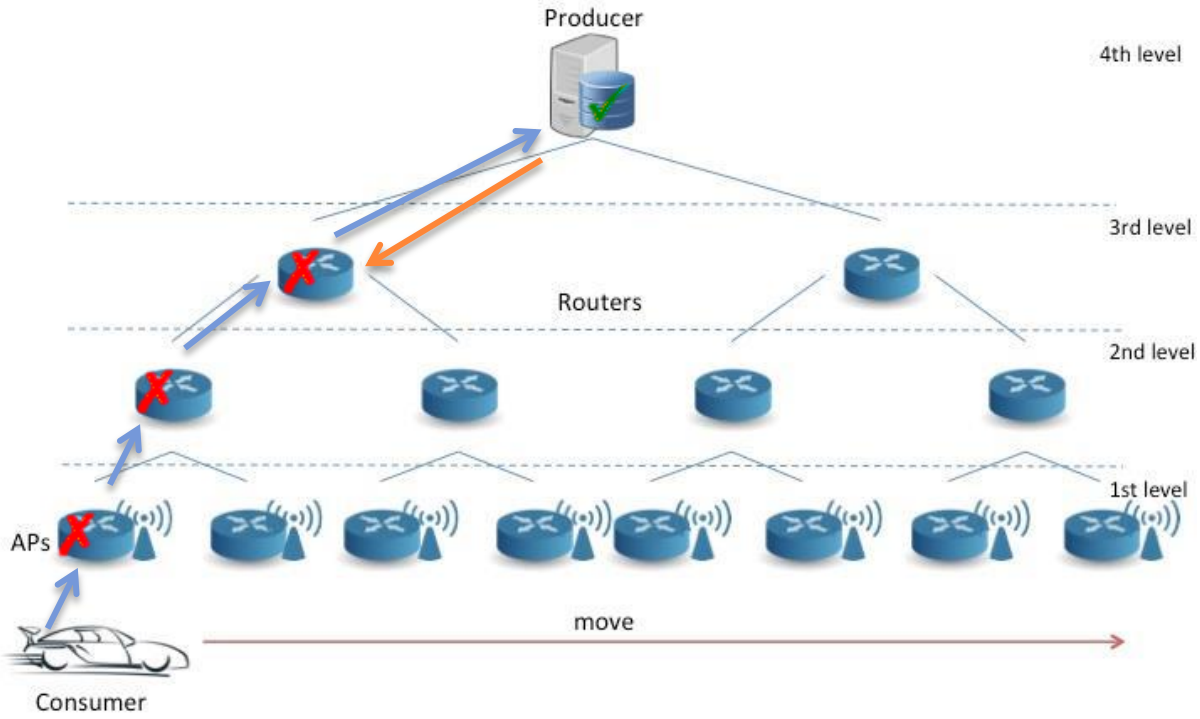
- The data itself is secured thanks to a digital signature over the content and its name, securely binding them.
- The data is publicly authenticable, anyone can verify that a name-content binding was signed by a particular key.
- Each signed data packet contains information to enable retrieval of the public key necessary to verify it.
- A digital signature guarantees integrity, provenance, and authenticity of the content, allowing the decoupling of the consumer's trust in data from where data is obtained.
- The data centric security can be used for content access control and infrastructure security.

Mobility

- 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.
- ICN naming and name resolution, as well as security features should natively support mobility.
- ICN is able to take advantage of multiple interfaces or adapt to the changes produced by rapid mobility.
- A request for a new content can flow from different interfaces, or through newly connected points of attachments in the network.
- 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.

MOBILITY

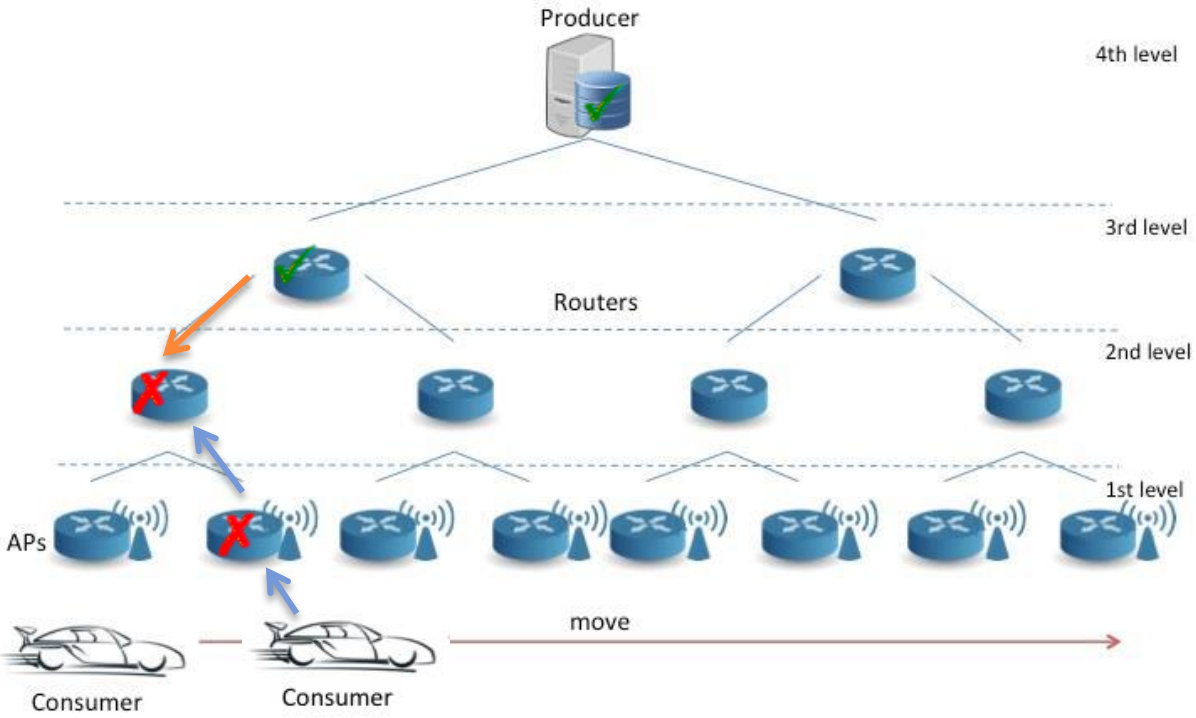
INTEREST DATA



✓Cache Hit ✗Cache Miss

MOBILITY

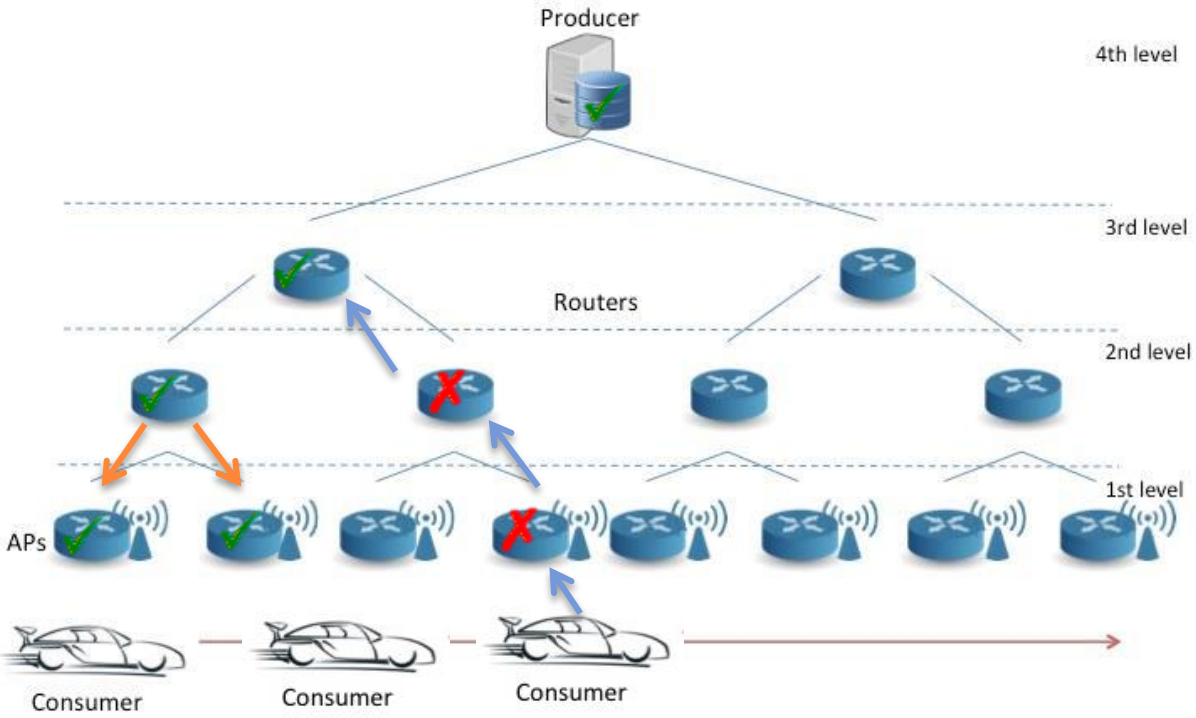
INTEREST DATA



✓Cache Hit ✗Cache Miss

MOBILITY

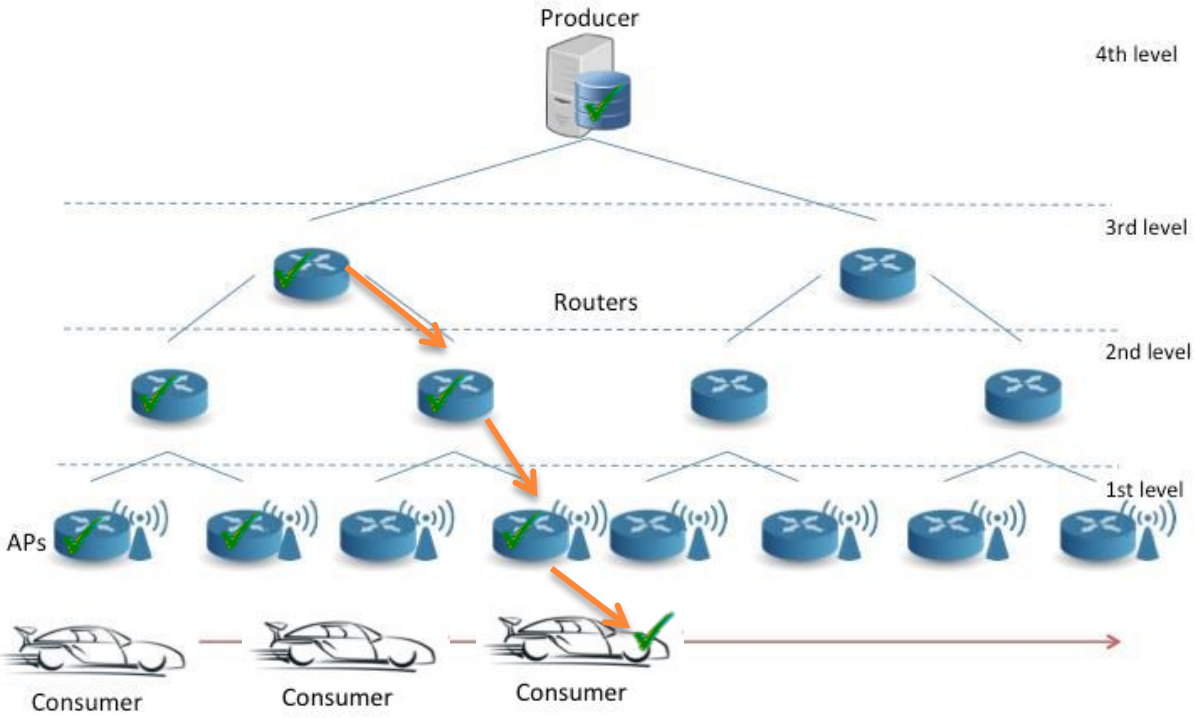
INTEREST DATA



✓Cache Hit ✗Cache Miss

MOBILITY

INTEREST → DATA ←



✓Cache Hit ✗Cache Miss

Open Challenges

Open challenges

TRUST MANAGEMENT

- Data object authentication is a fundamental ICN feature.
- Since data objects are replicated in network caches, they can be modified by malicious entities. The ICN architecture should provide a security mechanism to verify origin and integrity of contents.
- It is also necessary to define a **trust management** infrastructure to distribute the publisher's public key to the customers.

Privacy

- The ICN domain introduces new **privacy** issues related to the protection of what data could reveal, e.g. information about an individual along with his or her physical, cultural, economic, social characteristics, or personal behavior.
- Both the user requests and the cached content have a unique name that can reveal a lot of information about the users.
- Meanwhile, these information are important to improve the network performance.
- Thus, it is necessary to find a tradeoff between network performance and users' privacy.

Caching

- The in-network caching brings along improved efficiency, better scalability, and increased network performance, but also attracts new kinds of attacks, e.g. cache pollution.
- Moreover, the decision on which nodes should be equipped with caches is an open issue and could depend on topological criteria or traffic characteristics.
- The driver for the implementation, deployment and operation of in-network caches will be its costs.
- It should be decided also which content should be cached and where, considering both the replicas and the content popularity.
- Finally, since a lot of copies of named object are distributed among the in-network caches, a staleness verification algorithm should be defined.

Mobility

- The communication model and data replication in the network caches should facilitate a seamless handover in a **mobile** scenario.
- A seamless transition in ICN ensures that the content retrieval does not suffer from intermittent connectivity. The content reception continues without any perceptible change from the point of view of the ICN application receiving that content.
- Some open problems on ICN mobility are the following:
  - How to take full advantage of native ICN primitive?
  - How can mobility management be coordinated between the network nodes and the users for optimizing caching policies and sizing.
  - How is it possible to ensure that scalability issues are not introduced by the mobility management?
  - How the name resolution is affected by the rapid topological changes?

ROUTING

- ICN **routing** comprises name resolution, content discovery, and data delivery. ICN routing is a process that finds a data object based on its name.
- There is not a common consensus on how to manage these steps and different solutions are provided in literature.
- How to aggregate names of data objects to reduce the number of routing entries is a big challenge.
- Another problem is how to learn the object name which is designated by the provider.
- Also, how to manage the copies of a data object in in-network caches by the routing schemes is an open issue.
- The routing issues are strictly related to the naming convention.

Naming

- **Naming** data object is as important for ICN as naming host is for today's Internet.
- ICN requires unique names for individual data objects, because objects are identified independently of their location or container.
- Two possible naming schemes have been proposed: hierarchical and flat namespaces. Each solution has its own advantages and drawbacks but also in this case there is not a definitive accepted proposal.
- Updating and versioning named objects is challenging because it can contradict the fundamental ICN assumption: names have to be long-lived for retrieval. Thus, updating an object is not possible. Versioning is a possible solution.
- Names reveals what individuals request. Thus, except the problem of user privacy, the names can be used by malicious user to request the same object in the future and inferring an attack.

Conclusion

- The Internet has been a huge success, but the world has changed since it was created.
- The Internet architecture is no longer a good match to its primary use, so it is necessary to design a new architecture that addresses the today's problems.
- The answer is ICN, that generalizes the Internet by replacing the focus on where with what.
- The ICN paradigm is under development and standardization, however a lot of open challenges should be inspected and solved.
- The research community around ICN must grow and experiment with this new architecture. Thus, any volunteer to proceed in exploring these topics is welcome. Is this you?


Thank you

Bibliography

- L Zhang, A Afanasyev, J Burke, V Jacobson, KC Claffy, P Crowley, C Papadopoulos, L Wang, and B Zhang. **Named data networking**. Technical report, University of California, Los Angeles, 2014.
- Van Jacobson, Diana K. Smetters, James D. Thornton, Michael F. Plass, Nicholas H. Briggs, and Rebecca L. Braynard. **Networking named content**. In Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, CoNEXT '09, pages 1–12, New York, NY, USA, 2009. ACM.
- D. Kutsher, S. Eum, K. Pentikousis, I. Psaras, D. Corujo, D. Saucez, T. Schmidt, and M. Waehlisch, **ICN Research Challenges**. Draft IRTF, Feb. 2015.