Digital Fountains in Information-Centric Networking

George Parisis
Computer Laboratory
University of Cambridge
Cambridge, UK
george.parisis@cl.cam.ac.uk

Dirk Trossen
Computer Laboratory
University of Cambridge
Cambridge, UK
dirk.trossen@cl.cam.ac.uk

ABSTRACT

In this paper, we revisit digital fountains as an information theoretic approach for disseminating information. We embed this approach, however, into the architectural context of information centric networking (ICN). We discuss how our information-centric network architecture enables efficient information dissemination through fountain coding and we present the basic network operations for disseminating information from publishers to subscribers.

Categories and Subject Descriptors

D.3.3 [**Computer-Communication Networks**]: Network Architecture and Design – *distributed networks, network communications.*

Keywords

Digital fountains; information-centric networking; multi-path and multi-source information dissemination

1. INTRODUCTION

Information centric networking has been attracting significant attention in the networking community. Its focus on information dissemination rather than endpoint communication promises a better utilization of distributed computing and storage resources. Core to efforts such as [1][2][3][4] is the focus on information as the main principal of communication, recognizing that WHAT is communicated is often more important than WHO is communicating. This way, the notion of well specified endpoints becomes less important, moving the mental model from a well-defined endpoint-to-endpoint to a loosely coupled multipoint one, where information can be contributed by many sources for many receivers. Despite the shift from endpoints to information being delivered across networks, recent efforts on resource management within information centric networks, such as [6][7], seem to be stuck with the notion of *flows* between well-identified network entities, re-applying principles of TCP-like mechanisms as applied in the current IP world in an information-centric context.

In this paper we take a different approach by focusing on the information to be disseminated rather than the flows between the entities disseminating this information. For this, we utilize fountain coding [8] as an information theoretic approach for encoding content. Unlike previous work in reliable (IP) multicast transport [9], we directly embed the fountain codes into the (information centric) identification scheme that is used at the networking layer. For this, encoded symbols are separately identified in a way that makes them self-contained, and, therefore, cacheable at network nodes closer to subscribers.

ICN'13, August 12, 2013, Hong Kong, China.

2. DIGITAL FOUNTAINS IN OUR ICN

2.1 Enabling Efficient Information Dissemination

The digital fountains' principles are embedded into our ICN architectural context, which is described in detail in [2][4]. An implementation and evaluation of a network stack supporting the network architecture is presented in [5]. For simplicity, in this paper, we assume an intra-domain dissemination strategy, in which information is disseminated across a single network domain. Intra-domain rendezvous can be realized in dedicated network nodes that may share or replicate the information structure and the interested nodes. Topology management is realized in dedicated nodes that know the network topology and are notified when nodes attach to or detach from a forwarding node. Depending on the domain size, one or more cooperating topology managers may be required. Finally, forwarding is realized using LIPSIN [10] identifiers, which natively support source-based multicasting.

We see the following aspects within our architectural context as crucial for addressing the challenges in multipoint, fountain coding-based delivery. Firstly, the publish/subscribe service model naturally supports the notion of digital fountains for which a subscriber may be interested. This is similar to IP multicast solutions. However, ICN does not single out multicast as an exception but as a norm, unifying the service model across multicast and unicast delivery relations alike.

Secondly, the core functions separate the concerns of information management, communication resource management and delivery of the (coded) information. Hence, while we can utilize efficient stateless multicast solutions, such as [10], the separate topology formation and management function can realize opportunities for multi-source and multi-path dissemination as well as in-network caching. With the (network) resource owner receiving the most benefit for such optimized resource management, realizing these opportunities within the topology management function seems almost natural and is indeed the core of our proposal.

Thirdly, the identification principles of labeling and scoping information (see [4] for more information on these principles) allow for embedding algorithmic relationships in the delivery relations even after the initial demand/supply match occurred (through the rendezvous function) and the initial delivery relationship has been established (through the topology management and formation function). With that, LIPSIN identifiers can be reused for publishing algorithmically related information items.

Finally, identifying information items using statistically unique labels allows for caching individual encoded symbols in network nodes, which in turn can play them out again. Embedding the information for decoding the symbol into the identifier (through an algorithmic relation) makes these in-network stored symbols self-contained.

2.2 Basic Network Operation

The description of the basic network operation for disseminating information using digital fountains is based on LT codes [8], although other digital fountain schemes could be used as well. We utilize the principles of information labeling and scoping to create labels for all encoded symbols that can be produced when encoding some content. Information labels are used to communicate encoding-related information that is necessary for receivers (subscribers) to decode each symbol by calculating its *degree* and *neighbors* (see [8] for a detailed description of how fragments from the initial content are selected and XORed to create an encoded symbol). We call identifiers of an encoded symbol *algorithmic*, because they are all produced according to an algorithm that must be known to the entities participating in the communication while all other entities can remain agnostic to this algorithmic relationship. In other words, subscribers utilize the same algorithm to extract meaningful information out of the identifier, which will be later used to decode the initial content, while intermediary caches can utilize the statistical uniqueness of symbols' identifiers to possibly replay cached symbols.

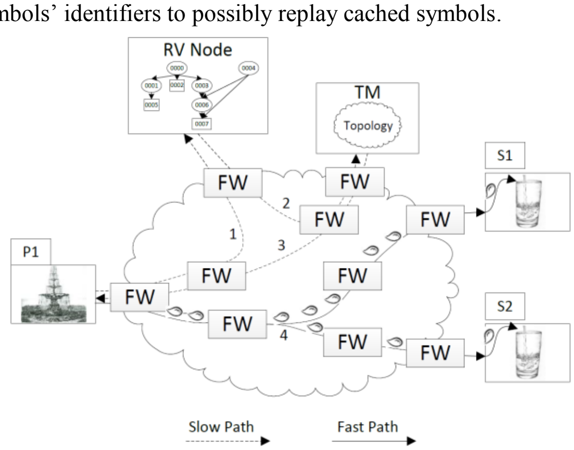

Figure 1. Intra-Domain Fountain

For our basic transport model, we assume, for reasons of simplicity, a single network domain, running a centralized topology manager and a dedicated rendezvous node that manages the information space (Figure 1). To keep the figure readable, we assume that subscribers *S1* and *S2* (shown with the glass waiting for drops from the fountain) are already subscribed to an item with ID */A/B* (or its parent scope */A*). Publisher *P1* (illustrated as a fountain), then advertises the information item */A/B* (step 1). The rendezvous node receives the request, matches the publisher with the subscribers and publishes a topology formation request to the domain's topology manager (step 2). The request is received by the topology manager, which is subscribed to a scope under which this information item resides, and knows the domain's topology. It creates a LIPSIN identifier from the publisher to the subscribers and publishes it to the publisher (step 3). Finally, the publisher application is notified that subscribers exist for the advertised item. At that point, the publisher's network stack stores a mapping of the advertised information item to the LIPSIN identifier and the digital fountain is subsequently enabled. Steps 1 to 3 comprise the slow-path operation of our ICN architecture. Step 4 takes place in the fast-path of our network. The publisher utilizes the previously created LIPSIN identifier to publish encoded symbols to the subscribers. Hence, there is no need for rendezvous for each encoded symbol, removing any extra delay. This identifier may be updated internally (in the network stack) as subscribers join or leave the digital fountain (steps 2 and 3 are repeated). This is transparent to the publishing application.

3. TOWARDS MULTI-SOURCE AND MULTI-PATH DISSEMINATION

We have implemented the basic network operation described above as a shim layer on top of the publish/subscribe network API exported by our network stack [5]. We are working on extending it towards supporting multi-source and multi-path dissemination. For this, we utilize the clean separation of our core network functions, allowing for transparently creating multiple paths from a set of publishers to a set of subscribers by just altering the topology management and formation functionality. This way multiple publishers could publish encoded symbols independently and in parallel. Subscribers could receive symbols from multiple sources and potentially from multiple network paths and/or network interfaces. As long as symbols from multiple sources are different, subscribers can combine all of them to decode the original content.

4. REFERENCES

[1] V. Jacobson and et al., "*Networking named content", Commun. ACM, 55, 1*, 117–124, 2012.

[2] D. Trossen, M. Sarela, and K. Sollins, "Arguments for an information-centric internetworking architecture," *SIGCOMM Comput. Commun. Rev., 40, 2*, 2010.

[3] T. Koponen and et al., "A data-oriented (and beyond) network architecture", *in Proc. of SIGCOMM '07*, 2007.

[4] D. Trossen and G. Parisis, "Designing and realizing an Information-centric Internet," *Communications Magazine, IEEE, 50, 7,* 60 –67, 2012.

[5] G. Parisis, D. Trossen and D. Syrivelis, "'Implementation and Evaluation of an Information-Centric Network", *In Proc. of IFIP Networking '13*, 2013.

[6] S. Arianfar, P. Nikander, L. Eggert, and J. Ott, "A transport protocol for content-centric networks," *in Proc. of IEEE ICNP '10*, 2010.

[7] G. Carofiglio, M. Gallo, and L. Muscariello, "ICP: Design and evaluation of an interest control protocol for content-centric networking," *in INFOCOM Workshops*, 2012.

[8] M. Luby, "LT codes", *in Proc. of the 43rd Symposium on Foundations of Computer Science (FOCS)*, 2002.

[9] J. Byers, M. Luby, M. Mitzenmacher and A. Rege, "A digital fountain approach to reliable distribution of bulk data*", In Proc. of SIGCOMM '98*, 1998.

[10] P. Jokela and et al., "LIPSIN: line speed publish/subscribe inter-networking", *in Proc. of SIGCOMM '09*, 2009.