

NDN-RTC: Real-time videoconferencing over Named Data Networking

Peter Gusev
UCLA REMAP
peter@remap.ucla.edu

Jeff Burke
UCLA REMAP
jburke@remap.ucla.edu

ABSTRACT

NDN-RTC is a real-time videoconferencing library that employs Named Data Networking (NDN), a proposed future Internet architecture. It was designed to provide a platform for experimental research in low-latency, real-time multimedia communication over NDN. It aims to provide an end-user experience similar to Skype or Google Hangouts, while implementing a receiver-driven architecture that takes advantage of the NDN architecture’s name-based forwarding, data signatures, caching, and request aggregation. As implemented, NDN-RTC employs widely used open source components, including the WebRTC library, VP9 codec, and OpenFEC for forward error correction. This paper also presents the design, implementation in C++, and testing of NDN-RTC on the NDN testbed using a demonstration GUI conferencing application, NdnCon, which provides “real-time” HD video communication over NDN to end-users.

1. INTRODUCTION

Named Data Networking (NDN) is a proposed future Internet architecture that shifts the “thin waist” of the Internet from the current host-centered paradigm of IP to data-centered communication. In NDN, every chunk of data has a hierarchical name, which can include human-readable components, and a cryptographic signature binding name, data, and the key of the publisher. Consumers of data issue “Interest” packets for these “Data” packets by name. Signed, named Data packets matching the Interest can be returned by any node on the network, including routers. NDN’s intrinsic caching can be leveraged by content distribution applications and significantly help to reduce the load on data publishers in multi-consumer scenarios [7]. Additionally, duplicate Interests for the same content can be aggregated in routers, further reducing the load on those publishers and the network. For more about NDN, please see the publications on the project website¹, including [20, 21, 9].

¹<http://named-data.net>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ICN’15, September 30–October 2, 2015, San Francisco, CA, USA.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3855-4/15/09 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/2810156.2810176>.

Efficient content distribution has long been a driver application for NDN research, as well as for the broader field of Information Centric Networking (ICN). Prior work in this area, including our own, is covered briefly in Section 4. However, *low-latency* applications such as “real-time” videoconferencing present particular design and implementation issues that have been less widely explored in publicly available prototypes or the NDN and ICN literature. For example, obtaining the “latest data” from a network with pervasive caching, without relying on direct consumer-producer communication (which impacts scaling potential) and while trying to keep application-level latency low, is a significant challenge.

The NDN-RTC library was created to explore real-time communications experimentally. The project is ongoing; this paper presents the current design and initial evaluation. We have been working towards the goal of using NDN-RTC in NDN project-related videoconferences and meetings. To be useful for project communications, we needed reasonable CPU and bandwidth efficiency, echo cancellation, and modern video coding performance. To meet these goals, NDN-RTC is built in C++ for performance and leverages the widely-used WebRTC library, incorporating its existing audio pipeline and video codec.

The remainder of this paper is organized as follows: Section 2 briefly describes potential NDN benefits for real-time communication. Section 3 details main project goals. Section 4 covers background and prior work. Section 5 describes the architecture of the library, designed namespace, data structures and algorithms. Section 6 discusses implementation details. Section 7 evaluates main outcomes. Finally, Section 8 provides a conclusion and explains future work.

2. WHY USE NDN FOR REAL-TIME COMMUNICATION?

The NDN project team uses application-driven research to explore NDN’s affordances for modern applications and to refine the architecture itself. Though it is based on what was learned from our NDNVideo project [7], NDN-RTC is a clean slate design with new goals.

A complete exploration of potential benefits for using NDN for real-time communication is outside the scope of this paper. Given that NDN proposes a general internet architecture, we are motivated on one hand by simply showing its viability for applications beyond the content distribution examples that are most often discussed in the literature. However, we can briefly present a few potential benefits of using NDN for RTC that are exciting for us:

- By using names rather than IP addresses for routing and forwarding, as with any other NDN application, RTC applications stand to inherit the benefits for mobility, scalability, and simplifications of network infrastructure that are currently being researched in the ICN community. With NDN-RTC, we use straightforward naming and communication scheme, leveraging conventions where possible, and build on the current libraries, forwarder, and testbed to increase the likelihood that these benefits can be inherited (or at least explored) in future work.
- Receiver-driven architectures with minimal publisher coordination can gain scalability from the network, as we have previously shown is viable in past work discussed below. Early tests of NDN-RTC, described in later sections, demonstrate network-supported scalability at a limited scale. In the future, by using broadcast or group encryption schemes, suggest that NDN could be highly efficient in supporting secure one-to-many or few-to-many low-latency broadcasts with very little additional application infrastructure - whether of entertainment content, presentations, closed-circuit cameras, computer vision sources, etc.
- Finally, because we are building protocols close to the thin waist of the NDN architecture, what is learned from exploring low-latency transmission of time series data here is beginning to provide directly applicable insight into handling other high rate and/or low latency time series, such as sensor data feeds.

3. DESIGN OBJECTIVES

As discussed above, NDN-RTC aims to explore low-latency audio/video communication over NDN, and provide a working multi-party conferencing application that can be used by NDN project team members across the existing NDN testbed. We are also attempting to preserve network-supported scalability by avoiding direct consumer-producer communication (e.g., Interests that require new Data to be generated by the producer for each request). Further, it explores if “real-time” communication can be achieved in a manner consistent with most other NDN data dissemination applications, described as follows. Bidirectional communication is achieved by having each node participate as a publisher and consumer. The publisher 1) acquires and transforms media data, 2) names, packetizes and signs it, and 3) then passes the packets to an internal or external component that responds to Interests received from the “black box” of the NDN network with signed, named data chunks. The consumer issues Interests with appropriate names and selectors to that “black box” of the NDN network, at the rate necessary to achieve its objectives and be a good citizen of the network²—as informed by the performance of the network it observes in response to its requests. It reassembles and renders them. Rate adaptation can be handled at the consumer by publishing in multiple namespaces corresponding to multiple bitrates (or to layers of a scalable video stream) and enabling the consumer to select them on-the-fly.

Once the namespace is defined, the publishing problem in this scenario (and in practice, at least so far) is relatively straightforward. Complexity is at the the consumer, which

must determine what names to issue at what rate, to get the best quality of experience for the application. For real-time conferencing, this means low-latency access to the freshest data that the “black box” of the NDN network can deliver to a given consumer.

Conference setup and multi-party chat could be handled by applying techniques such as those developed in ChronoChat [18], which uses a set reconciliation based synchronization protocols to exchange messages in a many-to-many scenario.

Based on this high-level concept, specific design goals were developed for the initial version described in this paper:

- **Low-latency audio/video communication.** The library should be capable of maintaining low-latency (approx. 250-750ms) communication for audio and video, similar to consumer videoconferencing applications.
- **Passive consumer & cacheability.** There should be no explicit negotiation or coordination between publisher and consumer for the media transmission itself, to enable explorations of highly scalable consumer to producer ratios.
- **Multiple bitrates.** The library and namespace should support multiple bitrates (from which consumers will select), enabling near future work on adaptive rate control.
- **Multi-party conferencing.** Publishing and fetching several media streams simultaneously should be straightforward, enabling near future work on multi-party conferencing.
- **Data verification.** The library should provide content verification using existing NDN features, as a building block for trust management and encryption-based access control.

Our design approach was further influenced by the relatively young state of NDN research:

- **Segment-level control.** NDN-RTC works directly with data segments rather than video frames, group-of-pictures (GOP) blocks, or other higher-level constructs. This was done because interest-level metadata is piggy-backed to each segment and may be exploited by the consumer to adjust fetching behaviour. This was done because most approaches explored in current research do not handle deadline-driven playout, assume large buffer sizes relative to network latency variations, and because frame sizes often exceed current NDN data object maximum sizes. While in the future, successful fetching and buffering patterns may be abstracted to a lower-level library, this approach enabled us to experiment with interest expression and buffering at fine granularity. Note that where appropriate, application-level operations are indeed performed on frames, such as forward error correction based recovery.
- **Assumptions about the network.** We make a number of assumptions about caching performance and other network behavior, presented throughout the paper. We understand that future networks may not behave in this manner, but test out the performance of simpler assumptions on the NDN testbed prior to trying more complex schemes.

In these ways we hope that NDN-RTC presents not the endgame for RTC over NDN but rather offers an experimental platform for developing more refined approaches.

²While work on congestion control and defining proper behaviour of such applications on the network is underway, it is future work with respect to this paper.

4. BACKGROUND AND PRIOR WORK

Most video streaming work on ICN has focused on play-out without the constraints of supporting interactive conversations. For example, UCLA’s NDNVideo, which supported live video and playback, was tested and deployed over NDN [7], scaling to approximately one thousand consumers from a single, simple publisher over plain-vanilla NDN. [6] Its data chunks were named sequentially according to NDN naming conventions [19], with a second time-based namespace mapping sequential chunks to a timeline to enable efficient random access by consumers. Though the project worked well for live and pre-recorded media streaming, it did not meet requirements for low-latency communication, and its architecture was not immediately extensible for use as a conferencing solution. This work was extended by others in NDNlive and NDNtube [17], which were designed to demonstrate a new API for application developers, the Consumer/Producer API [13], which works with higher-level “Application Data Units” rather than Interests/Data packets. They functioned over with the latest NFD (NDN Forwarder Daemon) and use repo-ng [1] for offline media storage. At the time of publication of current work NDNlive was still under development, and the project does not aim to address real-time audio/video conferencing.

Other streaming video work includes [12] and [15], which explore the advantages of using ICN networks for MPEG Dynamic Adaptive Streaming over HTTP. Although not directly related to low-latency streaming, these works also leverage ICN networks’ caching ability for serving chunks of video files efficiently to multiple consumers. In contrast with all of the above, NDN-RTC was developed from the ground up with low-latency objective and interactive conversation in mind, providing an alternative design motivated by these requirements.

The Voice-over-CCN project [8] was an early exploration of real-time communication over ICN, providing a similar level of quality compared to VoIP solutions with much greater scalability potential and simpler, more flexible architecture. VoCCN solved lock-step flow of data in ICN network by pipelining some number of interests for future data. This idea is widely exploited and developed in NDN-RTC.

Early in the NDN project, an audio conferencing application was developed in [22]. It leveraged use of Mumble VoIP software and successfully used NDN as a transport. Initial effort for conference and user discovery was made in this work as well. However, echo cancellation quality was poor, which made it difficult to use in real world scenarios, and only preliminary work in video was performed. Therefore, we elected to build NDN-RTC on top of the WebRTC library, despite the additional, significant implementation complexity, in order to use its audio-processing capabilities and video codecs, and potentially give an opportunity for easier integration with supported web browsers.³ Finally, the most recent related work in ICN-based real-time communication that we are aware of is [10]; however, it handles interest retransmission and buffering at the GOP level and thus does not (yet) meet the latency requirements laid out in our own design goals.

³No modifications are needed to WebRTC as used in NDN-RTC, which we believe will enable us to make comparisons between IP and NDN implementations of the library in the future.

5. APPLICATION ARCHITECTURE

There are two roles in the NDN-RTC library: producer and consumer. In bidirectional communication, applications use the library to play both roles, but a variety of other multi-party and one-to-many scenarios can be achieved. The library is built on a consumer-driven approach following NDN’s own Interest-Data exchange model. In contrast to the sender-driven approach of typical IP RTC, where the producer writes data to a destination host and the consumer reads it as it arrives, the producer publishes data to network-connected storage at its own pace, while the consumer requests data as needed and manages the relationship between outgoing Interests, incoming data segments, and buffer fill.

5.1 Producer

The producer’s main tasks are to acquire video and audio data from media inputs, encode them, *name them*, marshal the named data into network packets, sign the packets, and store them in an application-level cache⁴ that will asynchronously respond to incoming Interests. In this way, flow control responsibility shifts to the consumer, and scaling can be supported by network caches downstream rather than publishing infrastructure at the application level.

5.2 Namespace

A primary design question is how the data should be named so it can be retrieved by the consumer with desired properties. The NDN-RTC namespace defines names for media (segmented video frames and bundled audio samples), error correction data, and metadata, as shown in Figure 1. As there is no direct consumer-producer communication, the namespace is designed to efficiently support the type of fetching operations performed by the consumer, as introduced in Section 5 and detailed in Section 5.4. Note that we handle audio and video streams independently, which enables the library to support audio-only streaming as well as for applications to prioritize audio over video for increased quality of experience for interactive conversations in bursty or low-bitrate scenarios.

5.2.1 Media

The NDN-RTC producer abstracts the published media for a given source as a collection of *media streams*. A media stream represents a flow of media data, such as video frames or audio samples, coming from a source—currently, an input device on the producer. (For now, names for streams are derived from their corresponding device information.) A typical publisher will publish several media streams simultaneously—e.g., a camera and a microphone, but also a separate stream for screenshots. The data from a stream is encoded at one or more bitrates configured on the producer, so in the name hierarchy, each stream has children corresponding to different encoder instances called *media threads*. Media threads allow the producer to, for example, provide the same media stream in several quality levels, such as low, medium and high quality, so that the consumer can choose the media thread suitable for its requirements and current network conditions.

NDN-RTC packetizes the WebRTC video encoder output directly. Encoded media segments published under the hierarchical names described above, with video frames further

⁴Currently, this cache is provided to the application by the NDN-CCL library.

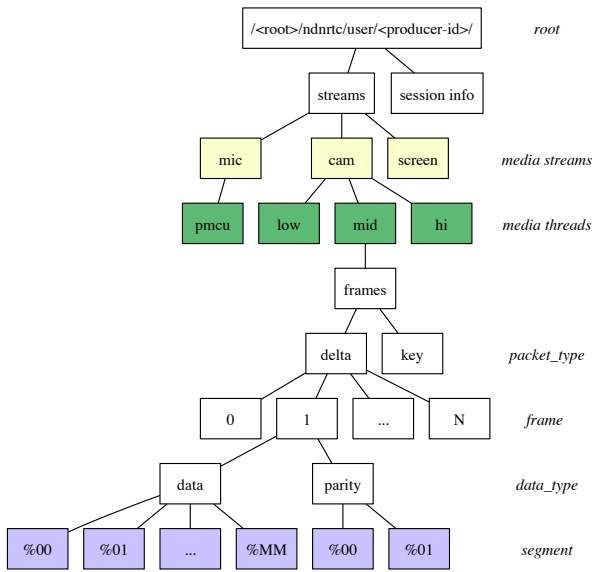


Figure 1: NDN-RTC namespace

separated into two namespaces per frame type, **delta** and **key**, each numbered sequentially and independently. Unlike Key frames, Delta frames decoding depends on previous Key frame and all Delta frames that come after it. Section 7 elaborates more on the reasons why this separation is needed. As needed, frame data is segmented with continuous sequence numbers per frame type.⁵

The next level in the tree separates data by type, either media or parity. Parity data for forward error correction, if the producer opts to publish it, can be used by a consumer to recover frames that miss one or more segments. In the case of both parity and regular frames, the deepest level of the namespace defines individual data segments. These segments are numbered sequentially, and their names conform to NDN naming conventions [19].

Audio streams are handled more simply, as their samples are much smaller than the maximum payload size, and there is no equivalent to the key/delta frame distinction in the audio codecs in use. All audio packets are published under the **delta** namespace. Multiple audio samples are bundled into one data packet, until the size of one data segment is reached, and published only after that.

5.2.2 Metadata

NDN-RTC uses both stream-level and packet-level metadata. Consumers need to know the producer’s publishing configuration and to save them from traversing the producer’s namespace, the producer publishes meta-information

⁵For example, the average sizes of frames for 1000 kbps stream using VP8/VP9: key frames are ≈ 30 KB, and delta frames are ≈ 3 -7KB. Therefore, depending on the underlying transport’s performance for delivering objects of this size, the producer may need to segment encoded frames into smaller chunks and name them in a way that makes reassembly straightforward. Based on our current observations of performance and the prevalence of UDP as a transport for the NDN testbed, NDN-RTC currently packetizes media into segments that are less than the typical 1500 byte MTU.

about current streams under the **session info** and updates it whenever configuration has changed.

Additionally, data names carry further metadata as part of each packet, which can be used by consumers regardless of which frame segment was received first. Four components are added at the end of every data segment name:

`.../delta/frame_no/data/seg_no/nseg/pl/prseq/parnum`

nseg - total number of segments for this frame;

pl - absolute playback position for current frame; this is different from the *frame*, which is a sequence number for the frame in its domain (i.e. **key** or **delta**);

prseq - sequence number of the corresponding frame from other domain (i.e., for delta frames, it is the sequence number of the corresponding key frame required for decoding);

parnum - number of parity segments for the frame.

Metadata in the name, rather than in the packet, is expected to be useful for application components or services that may not need to understand the packet payload.

5.3 Data objects

The producer generates signed data objects from input media streams and places them in an in-memory, application-level cache. These objects contain stream data and metadata.

5.3.1 Media stream

Each media packet payload consists of two types of data – a chunk from the media stream and metadata that describes it. Video stream data contains raw bytes received from the WebRTC library’s video encoder, which represent the encoded frame. For audio, NDN-RTC captures and encapsulates RTP and RTCP packets coming from the WebRTC audio processing pipeline, in order to obtain echo cancellation, gain control and other features, which are then fed into a similar pipeline on the consumer side for proper rendering and corrections.⁶

5.3.2 Metadata

The receiver-driven architecture of NDN-RTC and our experimental approach’s deliberate avoidance of explicit consumer-producer synchronization (to explore network scaling support) have suggested the importance of providing sufficient meta information from the producer. Apart from metadata provided in the namespace as was described earlier, there is also metadata supplied in the data objects themselves.

Packet-level metadata is applied to video as a header prepended to segment #0 (see Figure 2(a)). Each audio bundle (packet) is prepended by similar frame header, as seen in Figure 2(b). The frame header contains media-specific information such as frame size, timestamp, current rate and Unix timestamp⁷ (see Figure 3).

Additionally, segment-level metadata (see Figure 3) carries the producer’s observations of Interest arrival that is used by consumers to adjust fetching and playback mechanisms. The latter makes use of the Interest nonce value, and thus may not be as useful in larger multi-party calls.

⁶This is an artifact of the current implementation to benefit from the full audio pipeline of WebRTC; features of these protocols are not used, would be eliminated in the future.

⁷Producer timestamp is *not* required but can be used for calculating actual delay between NTP-synchronized producers and consumers)

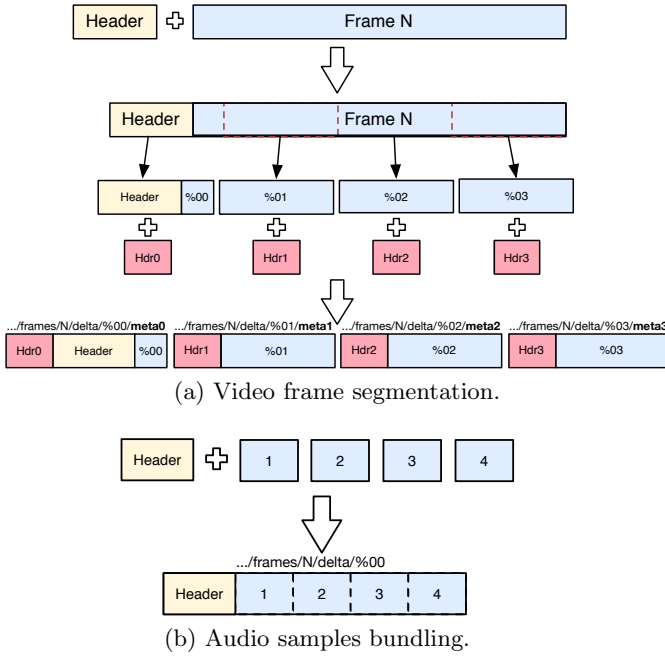


Figure 2: Segmentation and bundling

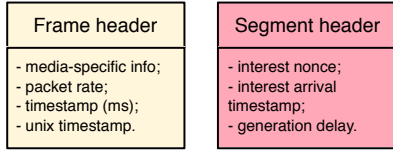


Figure 3: Frame and (experimental) segment headers.

For this reason, we consider these primarily for experimental purposes at this time:

Interest nonce: Nonce of the Interest which *first* requested this particular segment. Example interpretations include the following: 1) Value belongs to an Interest issued previously: consumer received non-cached data requested by previously issued Interest; 2) Value is non-zero, but it does not belong to any of the previously issued Interests: consumer received data requested by some other consumer; data may be cached; 3) Value is zero: data requested after it has been published; data may be cached.

Interest arrival timestamp: Timestamp of the Interest arrival. Monitoring publisher arrival timestamps may give consumers information about how long it takes for Interests to reach the producer. This value is only valid when the nonce value belongs to one of a given consumer's Interests.

Generation delay: Time interval in milliseconds between Interest arrival and segment publishing. A consumer can use this value in order to control the number of outstanding Interests. This value is again only valid when the nonce value belongs to one of the consumer's Interests.

5.4 Consumer

In NDN-RTC's receiver-driven architecture, the consumer aims to 1) choose the most appropriate media stream band-

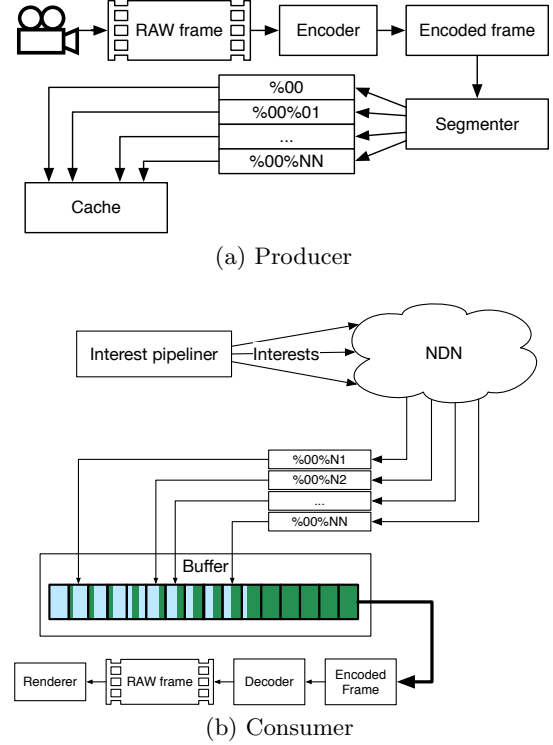


Figure 4: NDN-RTC producer and consumer operation.

width from those provided by the producer (by monitoring network conditions); 2) fetch (and, if necessary, reassemble) media in the correct order for playback; 3) mitigate, as far as possible, the impact of network latency and packet drops on the viewer's quality of experience. The consumer implements Interest pipelining and data buffering, as shown in Figure 4(b). An asynchronous Interest pipeline issues Interests for individual segments. Independently, a frame buffer handles re-ordering of packets, and informs the pipeline of its status to prompt interest reexpression.

5.4.1 Frame fetching

The consumer uses an estimate of the number of segments it must fetch for a given frame, issuing M initial Interests, as illustrated in Figure 5. If Interests arrive too early, they will be held in the producer's PIT and stay there until the frame is captured and packetized. We call the delay between Interest arrival and availability of the media data the **generation delay**, d_{gen} . Conceptually, we wish to keep this interval low, to avoid accumulating outstanding Interests with short lifetimes, but we also do not wish to have Interests arrive after data is published, as this increases latency from the end-user's perspective. Once the encoded frame is segmented into N segments and published, Interests $0 - M$ are answered and the Data returns to the requestor(s).

Upon receiving the first data segment, the consumer knows from the metadata the exact number of segments for the current frame, and issues $N - M$ more Interests for the missing segments, if any. These segments will be satisfied by data with no generation delay, as the frame has been published already by the producer. The time interval between receiving the very first segment and when the frame is fully assembled

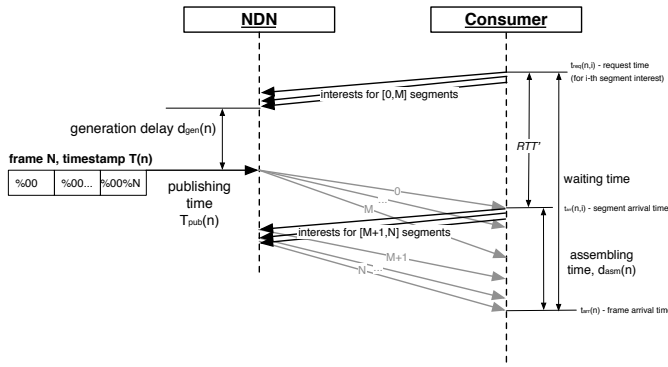


Figure 5: Fetching frame

is represented by d_{asm} and called **assembly time**. Note that for frames that are smaller than the estimate, some Interests may go unanswered; this is currently a tradeoff made to try to keep latency low for the frame as a whole.

Additional round-trips to refetch missing data can be avoided if the consumer is able to make more accurate estimates of the number of initial Interests. The library estimates the average number of segments for each frame type for a given bitrate, and uses this evolving estimate to calculate the number of initial Interests to issue for the next frame in each namespace. A similar process is used for fetching audio, though for now, audio bundles are represented by just one segment.

5.4.2 Buffering

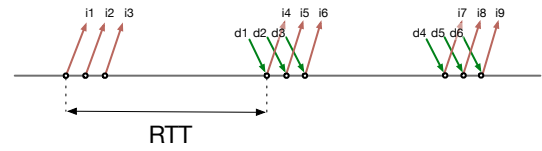
The consumer uses a “jitter buffer” to manage out-of-order data arrivals and variations in network delay, and as a place to assemble segments into frames (see Figure 9). Our receiver-based paradigm requires the consumer to request data by name explicitly, and organize it by frame as well as segment. Outstanding Interests are represented in the buffer by “reserved slots” - those that have partial frame data or no data at all. The NDN-RTC jitter buffer’s size is expressed in terms of two values measured in milliseconds at any given point in time: its *playback size* is the playback duration in milliseconds of all complete ordered frames; its *estimated size* is *playback size* + *number of reserved slots* \times $1/\text{producer rate}$ which reflects an estimated size of the buffer in case if all reserved slots have data. Each frame-level slot has an associated set of interests.

The difference between estimated buffer size and playback size corresponds to the effective RTT, which we call RTT' (this cannot be smaller than the actual network RTT value). Monitoring this value over times provides the consumer information on possible “sync” status with the producer as will be described further below.

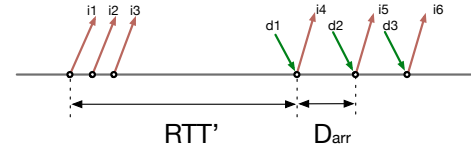
Playout progress of the jitter buffer is used for retransmission control. At some point inside the buffer (J milliseconds from the buffer end) there is a checkpoint before it is estimated to be too late for another roundtrip, where each frame is being checked for completeness. At the checkpoint, if incomplete frame can not be recovered using available parity data, Interests for the missing segments are re-issued.

5.4.3 Interest expression control

A key challenge in a consumer-driven model for videoconferencing in a caching network appears to be how to en-



(a) Bursty arrival of cached data, which reflects Interests expression pattern and indicates that the data is not the latest.



(b) Periodic arrival of fresh data, reflects publishing pattern and sample rate.

Figure 6: Getting the latest data: arrival patterns for the cached and most recent data

sure the consumer gets the latest data, without (per our design goal) resorting to direct producer-consumer communication. To get fresh data, the consumer cannot rely on using such flags in the protocol as *AnswerOriginKind* and *RightMostChild*. The frame period for streaming video is of the same order of magnitude of network round trip time, suggesting there is no guarantee that the data satisfying those flags received by a consumer will be the most recent one. Instead, it is necessary to use other indicators to ensure that the consumer is requesting and receiving the most up-to-date stream data possible given its (potentially evolving) network connectivity.

Our current solution is to leverage the known segment publishing rate, which is available in stream-level metadata, and note that, under normal operation, old, cached samples are likely to be retrieved more quickly than new data.⁸ We define the **interarrival delay** (d_{arr}) as the time between receipt of successive samples by a given consumer.

Our current assumption is that delays in the most recent samples follow the publishers’ generation pattern, but cached data will follow the pattern of Interest expression.⁹ Therefore, by monitoring inter-arrival delays of consecutive media samples and comparing them to the timing of its own Interest expression, which is distinct from the expected generation pattern, consumers can estimate whether they are receiving fresh data or cached data (see Figure 6). Note that the consumer’s objective is to obtain fresh data at a consistent rate from the network as a black box, not to try to somehow have Interests “reach” the producer directly.

NDN-RTC interest expression is managed in two modes, *bootstrapping* and *playback*. During bootstrapping, the consumer “chases” the producer and aims to exhaust network cache of historical (non-real time) segments. By increasing

⁸If the consumer is the *only* consumer of the stream, its Interests will go directly to the publisher, which also yields the correct behavior. A more complex challenge, for further study, is when segments are inconsistently cached in different ways along the path(s) that Interests take.

⁹Even though this assumption has proved its’ viability for low-scale tests, it has to be more carefully explored in future work for more complex scenarios

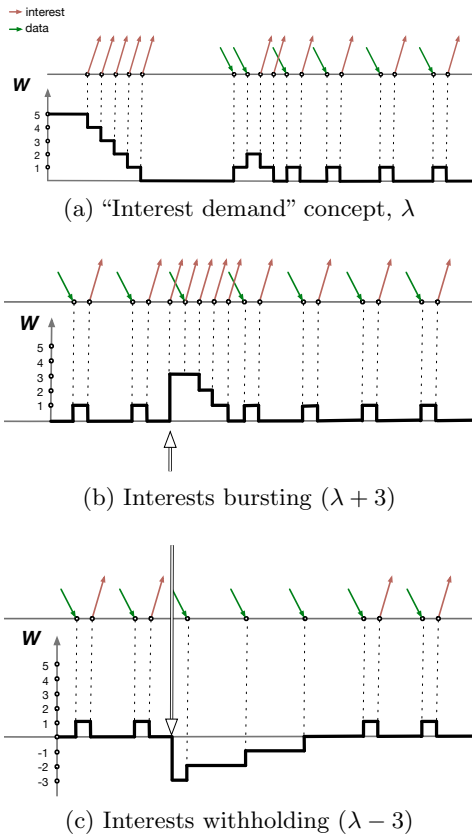


Figure 7: Managing Interest expression

the number of outstanding Interests, the consumer “pulls cached data” out of the network, unless the freshest data begin to arrive. In order to control Interest expression, the NDN-RTC consumer tracks a quantity called “Interest demand”, λ , which can be interpreted as how many outstanding Interests should be sent at the current time (see Figure 7). The consumer expresses new Interests when the demand is positive, i.e. $\lambda > 0$. For example, before the bootstrapping phase, the consumer initializes λ with a value which reflects the consumer’s estimate of how many Interests are needed in order to exhaust network cache and reach the most recent data. In playback, every time a new Interest is expressed, λ is decremented, and when new data arrives, λ is incremented, thus enabling the consumer to issue more Interests.¹⁰

Bootstrapping. In current design, there are two experimentally determined indicators that are used by the consumer to adjust λ : effective RTT (RTT') and inter-arrival delay d_{arr} . As described above, at bootstrapping (and re-acquisition), the consumer interprets d_{arr} stabilization around a relatively constant period, that means that consumer receives the freshest data available from the network. However, it does not necessarily mean that the consumer issues Interests efficiently. Figure 10(a) shows that even though the consumer has exhausted cache rather quickly,

RTT' is three times larger than the actual RTT for the network (100ms), which means that the majority of the issued Interests remain pending while waiting for the requested data to be produced.

The consumer makes several iterative attempts to adjust λ during bootstrapping, which can be described as follows:

1. The consumer initializes Interest demand with λ_d , and initiates Interests expression.
2. If the consumer did not receive freshest data during allocated time¹¹, increase Interest demand: $\lambda = \lambda + 0.5\lambda_d$; $\lambda_d = \lambda_d + 0.5\lambda_d$.
3. Whenever consumer receives freshest data (cache exhausted), decrease Interest demand: $\lambda = \lambda - 0.5\lambda_d$; $\lambda_d = \lambda_d - 0.5\lambda_d$ and wait for either of the two possible outcomes: a) RTT' decreases and the consumer still receives the freshest data – go to step 3; b) the consumer starts to get cached data (d_{arr} fluctuates greatly) – restore previous value for λ_d , increase λ accordingly and stop any further adjustments as the consumer has achieved sufficient synchronization with the producer.

Note that λ is a counter of how many Interests can be issued more whereas λ_d represents total number of outstanding Interests allowed. The value $\lambda_d - \lambda$ shows how many outstanding Interests consumer has issued at any given point in time. The way λ and λ_d are adjusted was determined empirically and may be a good topic for further research.

Bootstrapping begins with issuing an Interest with the enabled *RightMostChild* selector, in **delta** namespace for audio and **key** namespace for video (video decoding process can start only with a Key frame). Once an initial data segment of a sample with number S_{seed} has been received, the consumer initializes λ with initial value λ_d and asks for the next sample data $S_{seed} + 1$ in the appropriate namespace. Upon receiving the first segments of sample $S_{seed} + 1$, the consumer initiates the fetching process (described above) for all namespaces (**delta** and **key**, if available). The bootstrapping phase stops when the consumer finds the minimal value of λ which still allows receiving the most recent data and the consumer switches to the playback mode.

Interest demand provides a manageable mechanism to speed up or slow down Interest expression, coupling the asynchronous interest expression mechanism with the status of the playback buffer. An increase in λ value makes the consumer issue more Interests (Figure 7(b)), whereas any decrease in λ holds the consumer back from sending any new Interests (Figure 7(c)). Larger values of λ make the consumer reach a synchronized state with the producer more quickly. However, a larger value means a larger number of outstanding Interests and larger RTT' because of longer generation delays d_{gen} for each media sample. By adjusting the value of λ and observing inter-arrival delays d_{arr} , the consumer can find minimal RTT' value while still getting non-cached data, adapting towards a loose synchronization with the producer.

Playback. During playback, the consumer continues to observe RTT' and d_{arr} . Whenever d_{arr} indicates that no fresh data is being received, the consumer increases Interest demand and starts the adjusting process over again to find minimal RTT' for the new conditions. Such approach helps the consumers to adjust in cases when data may suddenly

¹⁰While inspired by the TCP congestion window, the Interest demand, as currently employed in NDN-RTC, may play a different role in ICN networks, which we are exploring experimentally in this application.

¹¹In the current implementation, 1000ms.

start to arrive from different network hub which introduces new network *RTT*.

Interest batches. Practically, for video, the consumer controls expression of “batches” of Interests rather than individual Interests, because video frames are composed of several segments. λ is adjusted on a per-frame basis, rather than per-segment.

6. IMPLEMENTATION

NDN-RTC is implemented as a library written in C++, which is available at <https://github.com/remap/ndnrtc>. It provides a publisher API for publishing an arbitrary number of media streams (audio or video) and a consumer API with a callback for rendering decoded video frames in a host application. NDN-RTC builds on functionality from several third-party libraries with which it is linked. NDN-CPP [11] is used for NDN connectivity and to provide in-memory storage for the application. The WebRTC framework [3] is used in two ways: 1) incorporation of the existing video codec; 2) full incorporation of the existing WebRTC audio pipeline, including echo cancellation. OpenFEC [2] is used for forward error correction support.

To demonstrate and evaluate the library, a desktop NDN videoconferencing application, NdnCon, [4] was implemented on top of NDN-RTC. It provides a convenient user interface for publishing and fetching media streams, text chat, and organizing multi-party audio/video conferences. It was used, along with a command-line interface for the evaluation below. The NDN-RTC library does not provide conference call setup functionality. This task was intentionally left out to be solved by a host app, and we are exploring it currently in the NdnCon demo app. The OS X platform is currently supported; Linux build instructions will be added soon.

7. EVALUATION & ITERATIVE REFINEMENT

Over the course of NDN-RTC development, we run numerous tests across existing testbed as well as in isolated environment in order to explore different library design patterns and implementations. These tests also helped us to understand the nature of low-latency communication over NDN and which requirements it raises for the NFD. Even though, for now we are still in the process of establishing well-defined metrics and test scenarios, there are some useful results described further in this section. There were several design iterations that each introduced improvements in the overall quality of experience for the end user, as well as in application efficiency in terms of bandwidth and computation. Each iteration tackled problems that were revealed during tests. These motivated namespace, application packet format and other revisions, which are reflected in the design detailed above.

7.1 Video streaming performance

A series of tests were conducted in order to assess bandwidth usage and perceived quality compared to Skype video calls. Each test was comprised of six runs of two-person, five-minute conference talks using NdnCon, the graphical conferencing application built on top of NDN-RTC: a) three runs of audio+video with low, medium and high video bandwidths settings (0.5, 0.7 and 1.5 Mbit/s accordingly); b) one run of audio-only conference; c) one run of Skype au-

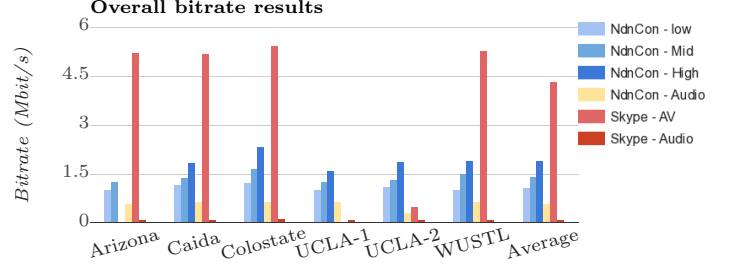


Figure 8: 2-peer conference tests compared to Skype

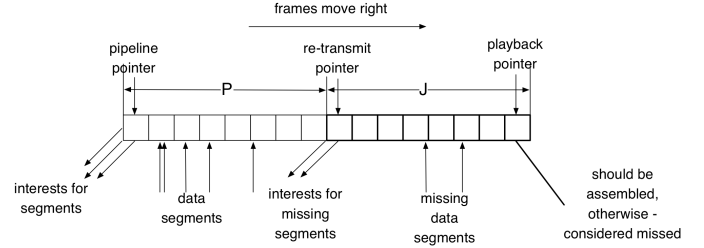


Figure 9: Frame buffer

dio+video conference; d) one run of Skype audio-only conference. Tests were conducted across the existing NDN testbed, between the UCLA REMAP hub and six other hubs. These tests covered both one-hop and multi-hop (with several intermediate hubs) paths.

Figure 8 shows overall bitrate usage results. Whereas Skype has adapted to use link capacity between peers, and delivered higher bitrate videos, NdnCon did not adjust to the current network conditions which make adaptive rate control features highly desirable as future work. Early in these tests, audio sample bundling was quickly introduced in NDN-RTC to reduce audio bandwidth (and the number of Interests on the consumer side), making it comparable to Skype audio bandwidths.

The results of such testing influenced iterative updates to the design, two of which are described below.

Separation of key and delta frame namespaces. Video streaming performance in early versions of NDN-RTC suffered from video “hiccups”, even when being tested on trivial one-hop topologies. The cause of this problem turned out to be an inefficient frame fetching process. In early NDN-RTC versions, the difference in size, and thus segments, of key frames and delta frames was not reflected in the producer’s namespace and consumers were forced to issue equal number of initial Interests (M on Figure 5) regardless of the frame type. This resulted in additional round trips of missing Interests and, consequently, larger assembling times (d_{asm}) for key frames that resulted in missed play-out deadlines. Having a separate namespace for key frames enables consumers to maintain separate Interest pipelines per frame type and collect historical data on the average number of Interests required to retrieve one frame of each type in one round trip.

7.2 Consumer-Producer synchronization

Bootstrap behavior. In initial library versions based on the approach taken in NDNVideo, the consumer “chased” the producer’s time-series data by exhausting cached data via issuing a large number of outstanding Interests. However, there was no mechanism for the consumer to figure out when to issue those Interests and whether Interest expression should be postponed in order to reduce timeouts. For two similar test runs (one-hop topology), the number of timed out Interests and re-transmissions varied greatly (either $\approx 1\%$ or $\approx 50\%$). One was due to an incorrect consumer’s synchronization with the producer; Interests were issued too early, so they timed out before any data had been produced. This problem was addressed by increasing the Interests’ lifetime.

However, for previous library versions, the mechanism for buffering (see Figure 9) dictated the Interests’ lifetime. In fact, in order to maintain the re-transmission checkpoint, all Interests entering the buffer had a lifetime equal to half of the current buffer size. This approach resulted in unavoidable Interest timeouts, in the cases when the consumer issued Interests far too early, before the actual data was produced. This was further complicated by forwarding strategies in NFD that did not handle consumer-initiated retransmission over short periods.

For the current version of the library, the re-transmission checkpoint is placed at a time estimated to be the effective RTT from the end of the buffer ($J = RTT$ on the Figure 9). This, together with an updated NFD re-transmission strategy [5], allows for larger Interests’ lifetimes. In fact, current implementation does not rely on Interests’ lifetimes anymore and every Interest is set to have 2000 ms lifetime.

Moreover, the problem described above can not occur if the consumer knows that it is issuing Interests too early. The chasing algorithm in older library versions was exhausting the network cache too aggressively; Interests were issued constantly until they filled up the buffer. With the introduction of the λ concept, the consumer has more useful control of the Interest expression. Figure 10 shows how a larger value of λ helps to exhaust the cache more quickly. The number of outstanding Interests is controlled by a consumer and directly influences how fast consumer can “chase” the producer. Thus, the consumer can control the “aggressiveness” of cache exhaustion.

7.3 Multi-party use

In one of many experiments, NdnCon was used to stream an NDN seminar over the existing NDN testbed. An audio/video bridge was set up using third-party tools, such as Soundflower and CamTwist Studio, allowing captured screen and audio feeds from existing IP-based conferencing tools to be simulcast. Figure 11(a) shows an example of instantaneous NDN testbed utilization during the one-hour conference call. It is estimated that media streams were consumed by five to eight people. Overall quality was satisfying, as reported by users.

Other tests of multi-party conferencing ability included four peers, each publishing three video streams and one audio stream and fetching one video and one audio stream from each of the other participants. Participants were distributed across four NDN testbed hubs - UCLA, REMAP, CAIDA and WUSTL (Figure 11(b)). These results were also generally satisfying from a quality of experience perspectives. However, for one user, audio cutoffs occurred more often

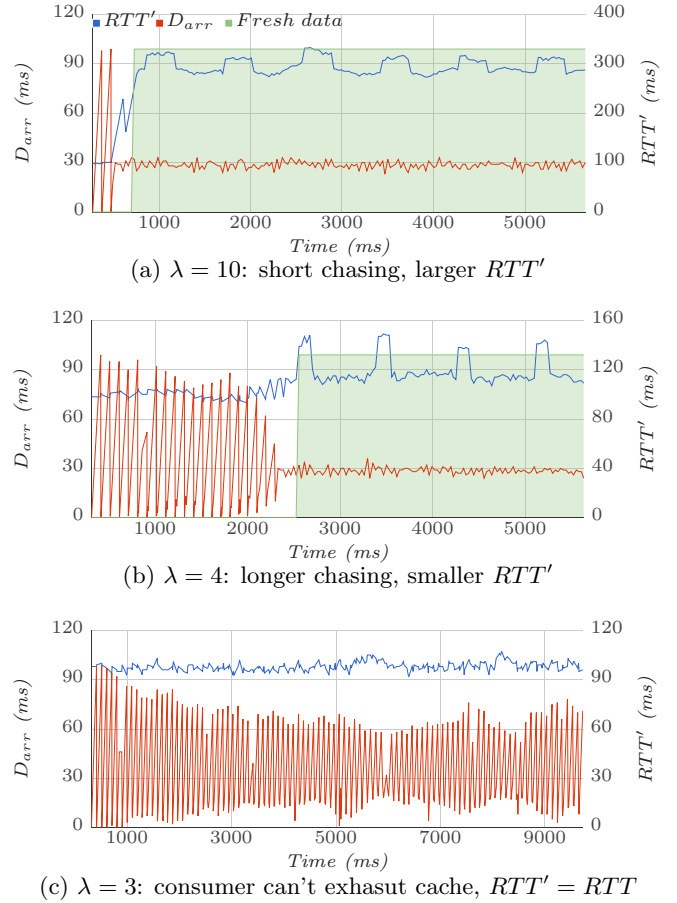


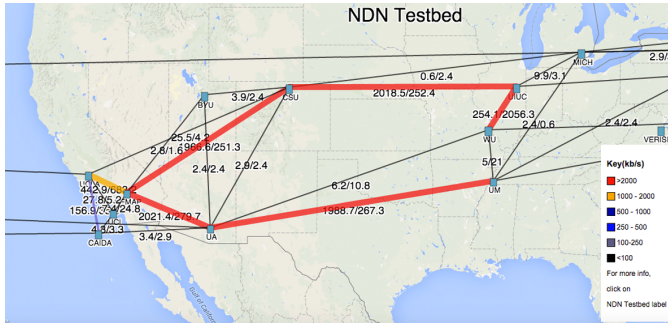
Figure 10: Larger λ decreases “chasing” phase, but increases RTT' for the same network configuration ($RTT \approx 100ms$)

than for the other participants. Many-to-many test scenarios are complex and are ongoing.

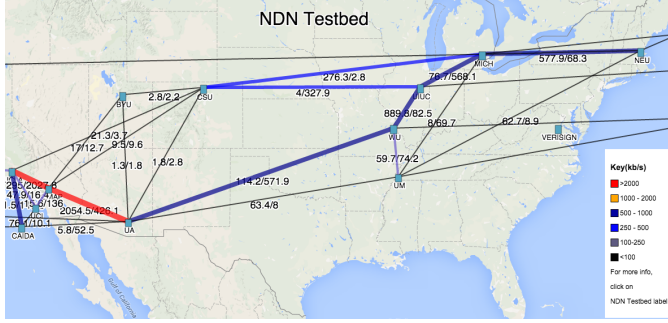
8. CONCLUSION AND FUTURE WORK

This paper presents the design, implementation, and initial experimental evaluation of NDN-RTC, a low-latency videoconferencing library built on Named Data Networking that is intended to support future experimentation. Our approach to this project has been experimentally driven so far, and has generated a functional low-latency streaming tool that we can now use a platform for exploring important design challenges in real-time media over NDN. This is a rich area, and some of the future work that we have identified includes:

Scalability tests. NDN-RTC has shown that real-time communication using NDN is viable with the current open source implementation and on the current NDN testbed; we are in the process of evaluating its performance in a variety of traffic and path scenarios. Both appropriate schemes for determining the “latest data” the network can provide at the correct rate and approaches to congestion control algorithms are an open challenge that requires collaboration between application developers, architecture researchers, and testbed operators.



(a) NDN testbed utilization during biweekly NDN seminar using NdnCon for simulcast.



(b) NDN testbed utilization during 4-peer call between UCLA, REMAP, WUSTL and CAIDA hubs.

Figure 11: NDN testbed utilization during one-to-many and many-to-many scenarios.

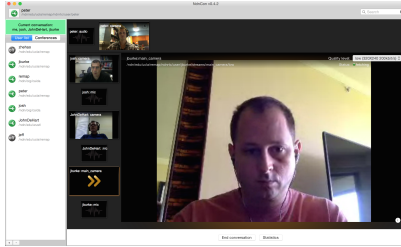


Figure 12: NdnCon screenshot.

Adaptive rate control. In the current design, the producer can choose to publish the same video stream at different bitrates, thus requiring the consuming user to manually select the best stream but laying the groundwork for adaptive rate control as a near-term effort. Such work will need to determine if monitoring of d_{arr} and other approaches described above can address challenges of such adaptation over ICN, as suggested in papers such as [16]. We are currently in the process of co-developing an adaptative rate control solution.

Scalable video coding. A more efficient way to relieve the producer from having to publish multiple copies of the same content at different bandwidths may be to use scalable video coding. By reflecting SVC layers in the namespace, the consumer will have more freedom for adapting media streams to the current network. Just as with audio, the SVC base layer may need to be prioritized.

Audio prioritization. For quality of experience in typical audio/videoconferencing applications, audio should be prioritized over video. This can be done at the application level, and we may provide such support in the future.

Encryption-based access control. The current NDN-RTC design supports basic content signing and verification. However, a prominent requirement of many types of videoconferencing is confidentiality, which we expect can be supported through encryption-based access control. While it may appear that encryption could limit the gains offered by caching, recent work exploring that application of advanced cryptographic techniques such as attribute-based encryption to multimedia in ICN [14] suggest interesting new directions for balancing benefits of ICN with security requirements.

Inter-consumer synchronization. The absence of direct consumer-producer coordination shifted the complexity of “RTC-over-NDN” streaming to the consumer. Another related aspect of teleconferencing which was not covered by this work is to ensure media playback synchronism across different consumers. This points more generally to the need for time synchronization implementations over NDN.

9. ACKNOWLEDGEMENTS

This project was partially supported by the National Science Foundation (award CNS-1345318 and others) and a grant from Cisco. The authors thank Lixia Zhang, Van Jacobson, and David Oran, as well as Eiichi Muramoto, Takahiro Yoneda, and Ryota Ohnishi, for their input and feedback. John DeHart, Josh Polterock, Jeff Thompson, Zhehao Wang and others on the NDN team provided invaluable testing of NdnCon. The initial forward error correction approach in NDN-RTC was by Daisuke Ando.

10. REFERENCES

- [1] Repo-ng. <https://github.com/named-data/repo-ng>.
- [2] OpenFEC Library. <http://openfec.org>.
- [3] WebRTC Project. <http://www.webrtc.org>.
- [4] NdnCon GitHub Repository. <https://github.com/remap/ndncon>, September 2014.
- [5] NFD version 0.2.0 release notes. http://named-data.net/doc/NFD/0.3.1/RELEASE_NOTES.html, August 2014.
- [6] P. Crowley. Named data networking: Presentation and demo. In *China-America Frontiers of Engineering Symposium*, Frontiers of Engineering, 2013.
- [7] J. B. Derek Kulinski. NDNVideo: random-access live and pre-recorded streaming using ndn. Technical report, UCLA, September 2012.
- [8] V. Jacobson, D. K. Smetters, N. H. Briggs, M. F. Plass, P. Stewart, J. D. Thornton, and R. L. Braynard. Voccn: voice-over content-centric networks. In *Proceedings of the 2009 workshop on Re-architecting the internet*, pages 1–6. ACM, 2009.
- [9] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard. Networking named content. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pages 1–12. ACM, 2009.
- [10] A. Jangam, R. Ravindran, A. Chakraborti, X. Wan, and G. Wang. Realtime multi-party video conferencing service over information centric network. In *Proceedings of Workshop on Multimedia Streaming in Information Centric Networks (MUSIC) at ICME*, 2015.

- [11] J. B. Jeff Thompson. NDN Common Client Libraries. *NDN, Technical Report NDN-0007*, September 2012.
- [12] S. Lederer, C. Mueller, and B. Rainer. Adaptive streaming over content centric networks in mobile networks using multiple links. *ÅÏ (ICC)*, 2013.
- [13] I. Moiseenko and L. Zhang. Consumer-producer api for named data networking. In *Proceedings of the 1st international conference on Information-centric networking*, pages 177–178. ACM, 2014.
- [14] J. P. Papanis, S. I. Papapanagiotou, A. S. Mousas, G. V. Lioudakis, D. I. Kaklamani, and I. S. Venieris. On the use of attribute-based encryption for multimedia content protection over information-centric networks. *Transactions on Emerging Telecommunications Technologies*, 25(4):422–435, 2014.
- [15] D. Posch, C. Kreuzberger, and B. Rainer. Client starvation: a shortcoming of client-driven adaptive streaming in named data networking. *Proceedings of the 1st ÅÏ*, 2014.
- [16] D. Posch, C. Kreuzberger, B. Rainer, and H. Hellwagner. Client starvation: a shortcoming of client-driven adaptive streaming in named data networking. In *Proceedings of the 1st international conference on Information-centric networking*, pages 183–184. ACM, 2014.
- [17] L. Wang, I. Moiseenko, and L. Zhang. Ndnlive and ndntube: Live and prerecorded video streaming over ndn. Technical report, UCLA, 2015.
- [18] Y. Yu. ChronoChat. <https://github.com/named-data/ChronoChat>.
- [19] Y. Yu, A. Afanasyev, Z. Zhu, and L. Zhang. Ndn technical memo: Naming conventions. Technical report, UCLA, July 2014.
- [20] L. Zhang, A. Afanasyev, J. Burke, V. Jacobson, K. Claffy, P. Crowley, C. Papadopoulos, L. Wang, and B. Zhang. Named data networking. Technical report, 2014.
- [21] L. Zhang, D. Estrin, J. Burke, V. Jacobson, J. Thorton, D. K. Smetters, B. Zhang, G. Tsudik, K. Claffy, D. Krioukov, D. Massey, C. Papadopoulos, T. Abdelzaher, L. Wang, P. Crowley, and E. Yeh. Named data networking tech report 001. Technical report, 2010.
- [22] Z. Zhu, S. Wang, X. Yang, V. Jacobson, and L. Zhang. Act: audio conference tool over named data networking. pages 68–73, 2011.