Understanding the Role of In-Network Content Caching in Future Internet Design: A Vertical Slice Approach

Master's Thesis Proposal by *Harshad Shirwadkar*

May 13, 2015

1 Introduction and Background

Information Centric Networking(ICN) is an approach of redesigning the Internet as a network architecture that focuses on *content* (or named data) rather than *hosts*(as in today's Internet). Various ongoing future Internet projects like XIA[1], NDN[2] incorporate the idea of ICN. The reason for this shift of focus is that most of the applications running over the Internet today are content oriented. For example, when you access youtube.com in your browser window, you are more interested in watching the video than connecting to youtube's video server. The nodes in ICN talk in terms of content. So, these architectures have notions of *Content Request* and *Content Response* (See table 1).

1.1 Content Caching in ICN

One of the key features of these architectures is "in-network caching". The *Content Response* objects enable in-network caching on routers (or even on dedicated cache servers). In-network caching optimizes bandwidth consumption, reduces network congestion and provides fast fetching for the popular content.

1.2 Open Problems related to Content Caching in ICN

In the context of in-network-caching, much has been researched about content popularity determination, cache eviction policies, cache topologies. But, there is not much literature on the effect of prevalence of content caches on other components of the Internet (transport protocols / naming service). In other words, we do not exactly know how the existing

Architecture	Content Request	Content Response
NDN	Interest Packet	Data Packet
XIA	CID Request packet	CID Response Packet

Table 1: Notion of Content Request and Content Response in various FIA proposals

network protocols should be modified to effectively utilize the caching infrastructure. Section 2.1 talks about the impact on DNS and the questions that need to be answered. Section 2.2 talks about trade-offs between privacy and caching performance. Section 2.3 looks at what APIs underlying caching architecture can provide to the applications programmers to effectively use caches. Note that in order to limit the proposal size, I have only included 3 problem areas. I intend to look at many more.

As this study involves understanding the impact of content caching on layers from top to bottom, we call it a "Vertical Slice Approach". However, given the time constraints, it might not be possible to study each and every layer. In the initial period, I will spend considerable amount of time on identifying which problems to study (See section 5 for details).

2 Goals

As mentioned in the previous section, the impact of content caches on network architecture design has not been well studied. Studying this would be the ultimate goal of my thesis. In my thesis, I not only wish to find answers to some of the challenging questions but also come up with interesting research challenges that can be posed to the ICN community. In order to establish some context to the questions mentioned in 2.1, 2.2 and 2.3, I will try to study these problems with respect to *eXpressive Internet Architecture*[1] which is an ongoing research project at CMU. I don't want to restrict myself to the questions mentioned in this proposal. Rather, I am excited to come up with new research questions. Following sections list a few candidate questions that can be looked at.

2.1 DNS and ICN

Traditionally, DNS has been responsible for providing IP addresses for names. Note that these names have been identifying hosts (typically servers running that service). However, in ICN, DNS has much larger role to play: it must provide network addresses for content names as well as host names. Since content names outnumber host names by several orders of magnitude, we most probably will need to redesign existing DNS design. For example, to save DNS overhead, wherever possible, content publisher might provide Network IDs for some content objects (e.g., in case of web pages where publisher needs to embed links to other web pages). However, this means we are essentially moving name look up from client to the publisher. This is quite opposite to today's DNS mechanism and hence needs in-depth study.

2.2 Privacy and Performance

Privacy comes at a cost and that cost has been well captured in [3]. In context of ICN caching, the cost of privacy might be as severe as not being able to cache content at all. For example, caching becomes impossible when client opens a secure tunnel with the end server (as in TLS). Also, some of the recent proposals in the networking community demand the use of HTTPS by default. Such proposals can prove to be detrimental to caching in ICN.

Since security can not be compromised, ICN caching needs to accommodate the existence of TLS like security protocols. One possible way achieve that by letting clients establish tunnels with in-network caches instead of final server. But, what might its implications be? Aren't we adding extra processing overhead? In the context of privacy, we need to find answers to questions like these.

2.3 Caching APIs

Today, we have the notion of sockets as end points of communication. But, is that sufficient in ICN since we care about data more? For example, in ICN, since caches exist everywhere, we could give abilities to the applications to pre-allocate a cache *slice* on router. But, there are many pros and cons in doing this. So, the problem of defining Caching APIs aims at identifying the trade-offs and come up with a set of APIs that we could provide to applications.

3 Evaluation

3.1 Periodic Checkpoints

The evaluation will be based on periodic checkpoints and the final outcome. In section 5, I have sketched out a rough timetable. At every step, we will have a checkpoint.

3.2 Weekly XIA Meetings

The XIA group organizes meetings every week on Fridays. I will present my progress in those meetings periodically.

3.3 Final Grading

The final grading will consist of following factors primarily:

- **Novelty of Ideas:** This captures the quality of arguments made in the thesis.
- **Contribution to XIA project:** I will make every effort to contribute to the XIA project by getting my prototypes approved and finally merged in the XIA code base.
- **Final Deliverables:** The final grade will also depend upon the final deliverables such as a possible SIGCOMM submission.
- **Periodic Checkpoints:** Last but not the least, periodic checkpoints will play an important role in the final grade.

4 Deliverables

- Conference Paper
  At the end of the Fall semester, I hope to come up with a draft for SIGCOMM 2016.
- Contribution to XIA
  My prototypes would get merged into CMU's XIA project.

5 Expected Timetable

Time Period	**Target**
So far	Have become familiar with XIA concepts and XIA implementation. Through independent research, I have also implemented Content Cache for XIA.
June 2015	Read more papers and focus on other problems. Meanwhile, also take one existing problem (Caching APIs) and come up with a solution to it.
July 2015	Come up with solution to at least 3 other problems apart from Caching APIs.
August - September 2015	Implement solutions and get reviewed from XIA team. Perform experiments and gather results to evaluate the implemented solutions.
October 2015	Start writing final thesis report and SIGCOMM paper for the work done so far.
November 2015	By the end of November, defend the final thesis.

Table 2: Expected timetable

6 Thesis Committee

I am extremely happy to be advised by Prof. Peter Steenkiste for my thesis. Prof. Srini Seshan has agreed to serve as my thesis reader.

References

[1] A. Anand, F. Dogar, D. Han, B. Li, H. Lim, M. Machado, W. Wu, A. Akella, D. G. Andersen, J. W. Byers, S. Seshan, and P. Steenkiste. Xia: An architecture for an evolvable and trustworthy internet. In *Proceedings of the 10th ACM Workshop on Hot Topics in Networks*, HotNets-X, pages 2:1–2:6, New York, NY, USA, 2011. ACM.

[2] 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*, CoNEXT '09, pages 1–12, New York, NY, USA, 2009. ACM.

[3] D. Naylor, A. Finamore, I. Leontiadis, Y. Grunenberger, M. Mellia, M. Munafò, K. Papagiannaki, and P. Steenkiste. The cost of the "s" in https. In *Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies*, CoNEXT '14, pages 133–140, New York, NY, USA, 2014. ACM.