Data-Centric Networking: advanced modeling and design

Ilaria Cianci
Telematics Lab - DEE - Politecnico di Bari
i.cianci@poliba.it

Michele Tortelli
Telematics Lab - DEE - Politecnico di Bari
m.tortelli@poliba.it

Luigi Alfredo Grieco
Telematics Lab - DEE - Politecnico di Bari
a.grieco@poliba.it

Gennaro Boggia
Telematics Lab - DEE - Politecnico di Bari
g.boggia@poliba.it

1. Introduction

In the last years users have changed the way they use Internet. They are interested in contents or, in other words, in "what" the network offers. For historical reasons, the technology is based on classic host-to-host communications, namely on "where" contents are.

Different solutions for Future Internet have been proposed to satisfy this new trend and to solve several problems that afflict the current TCP/IP architecture, like traffic explosion, network congestion, weak support for mobility and broadcasting, depletion of IP addresses and security flaws. Many of them are based on a "data-centric" approach, i.e. users can retrieve contents without any information about their physical location.

A possible way to build a data-centric network is to fully replace the TCP/IP stack, i.e., the clean-state approach. Examples of these networks are the PSIRP (Publish Subscribe Internet Routing Paradigm) architecture [4], the 4WARD NetInf project [5], and the Cache-and-Forward Network Architecture [1].

Another class of data-centric networks is known as "overlay networks". Differently from the clean-state one, this approach aims to preserve the main advantages of the current architecture, building an overlay that adds the advantages of the data-centric approach. Examples of overlay networks are DONA (Data-Oriented Network Architecture) [3] and CCN (Content-centric Networking) [2].

In this work we focus on CCN, a new architecture proposed by the Palo Alto research Center team [2], because it allows to improve performance and to solve the main problems of current Internet. In a CCN each content becomes an entity including all the mechanisms useful to verify integrity and validity; users can retrieve them regardless of their physical locations. This allows to achieve traffic and congestion reduction.

At the present, several research groups in the world, like the French INRIA, the Advanced Computer Science Research Group at Cambridge University, the Department of Computer Science at the University of Helsinki, the Department of Electronic & Electrical Engineering at the University College London, ...) are considering different ways to use the Content-Centric Networks and their performance limits. However, in our opinion, the lack of a unique method for the analysis of CCN poses serious obstacles to the design of innovative applications.

Our research group at the Politecnico di Bari has recently launched a line of research that aims to develop innovative models for characterizing the performance of a CCN. In the next section we briefly outline our ideas proposing an analytical model that investigates CCN proprieties with particular focus on fairness in the cache usage.

2. CCN Modeling

We proposed in [6] a discrete-time model to analyze the fairness in cache usage of CCNs. The model captures the distribution of content replicas among nodes of the network by taking into account the network topology, the number of contents, the cache size, and the content popularity. In particular, we consider the time axes splitted in slots: during the k-th slot all request made in the previous one are served.

The following basic notation is used:

- S: number of nodes;
- M: number of contents;
- B: cache size;
- H: average path length;
- A: total cache space;
- π_i: popularity of content i;
- $p_i(k)$: probability that a node does not store a copy of the i-th content;

In our model, during the k-th timeslot, the average number of requests for the i-th content is equal to $\pi_i\,[S - n_i(k)]$. Each one of these requests will generate $n_i(k)$ replies and each reply will pass through a path with H hops length. The max number of nodes that can cache the i-th content during the k-th slot is $H \cdot n_i(k)$. Remembering that $p_i(k)$ is the probability that a node does not store a copy of the i-th content, we can express the average number of new copies of the i-th content during the k-th slot as:

$$D_i(k) = \pi_i[S - n_i(k)] \cdot H \cdot n_i(k) \cdot p_i(k) \tag{1}$$

The total number of maximum new copies will be then $D_T(k) = \sum_{i=1}^{M} D_i(k)$.

Now we can distinguish two different scenarios: in the first one the overall cache availability is not enough for the total number of new copies ($D_T(k) > A$); in the second one, instead, results $D_T(k) \leq A$.

We assume that $\sum_{i=1}^{M} n_i(k) = B \cdot S$, since we expect that at each new request the number of copies of each content will grow mathematically up to the saturation of caches.

Analyzing the timemathematicallythe system, we found the equilibrium point, by setting $n_i(k+1) = n_i(k) = n_i^{eq}$, and we obtain:

$$n_i^{eq} = \begin{cases} 1 + \frac{A}{D_T^{eq}}(p_i^{eq})^2 n_i^{eq}\pi_i \cdot H \cdot S, \text{ if } D_T^{eq} > A \\ 1 + (p_i^{eq})^2 n_i^{eq}\pi_i \cdot H \cdot S + n_i^{eq}\frac{A - D_T^{eq}}{B \cdot S}, \text{ otherwise.} \end{cases} \quad (2)$$

After few steps (see [6] for all details), we obtained a closed form expression in both cases. For $D_T^{eq} > A$:

$$n_i^{eq} = \begin{cases} g_1(\pi_i), \text{ if } \pi_i < \pi_0 \\ g_2(\pi_i), \text{ if } \pi_i \geq \pi_0 \end{cases}, \quad (3)$$

where

$$g_1(\pi_i) = 1 + \frac{-\pi_i \cdot \left(\frac{2}{S} - 1 + \frac{\alpha}{\pi_i}\right) + \sqrt{\pi_i^2 \cdot \left(\frac{2}{S} - 1 + \frac{\alpha}{\pi_i}\right)^2 + 8\frac{\pi_i^2}{S}}}{4\frac{\pi_i}{S}} \quad (4)$$

$$g_2(\pi_i) = S \cdot \left(1 - \sqrt{\frac{\alpha}{\pi_i}}\right) \quad (5)$$

and π_0 is chosen so that $g_1(\pi_0) = g_2(\pi_0)$.

For the second case ($D_T^{eq} \leq A$), instead, results:

$$n_i^{eq} = \begin{cases} l_1(\pi_i), & \text{if } \pi_i < \pi_0' \\ l_2(\pi_i), & \text{if } \pi_i \geq \pi_0' \end{cases}, \quad (6)$$

where

$$l_1(\pi_i) = \frac{-1 \pm \sqrt{1 + 4\pi_i H^2 S^2 \cdot (\pi_i - \alpha')}}{2HS(\pi_i - \alpha')} \quad (7)$$

$$l_2(\pi_i) = S \cdot \left(1 - \sqrt{\frac{\alpha'}{\pi_i}}\right) \quad (8)$$

In figure 1 it is shown both the functions g_1 and g_2 compared to a reference proportional trend (i.e. the number of content replicas is proportionally to their popularity). We can observe that contents with small popularity track the proportional trend. The number of content replicas for contents with intermediate popularity is more than linear compared with the proportional trend, while for the most popular contents the number of copies in the network is sub-linear.

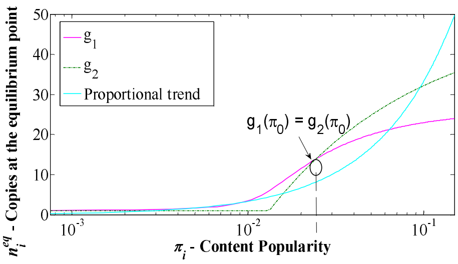

Figure 1. Comparison between g_1, g_2 and the proportional trend

3. Validation and results

To validate the model presented in the previous section we used a simple simulator developed in Matlab environment. In particular, we analyze a scenario with 50 nodes (see figure 2) and 200 contents. We assigned to each content a different popularity, uniformly distributed in the interval $[7.5 \cdot 10^{-5}, 0.15]$.

In figure 3 it is shown the theoretical trend and the simulated one for three different scenarios. In all cases, the Mean Absolute Relative error between the two trend is less then 28%.

4. Conclusion and ongoing works

Our analysis has brought to light pros and cons of a CCN network. Evaluating the fairness, we found that CCN overlay privileged contents having intermediate popularities, despite of the common idea that high popularity contents are more diffused. This behavior could be seen both from a positive and a negative point of view. The network behavior could limit the desired diffusion of some contents, but on other hand the negative effect of an Interest Flooding attack can be avoided. Our ongoing works aim to validate the developed framework with an emulated network, composed by physical and virtual hosts. Our goal is also to investigate about others performance indexes of CCN, to evaluate the impact of contents with different sizes and to develop killer CCN applications.

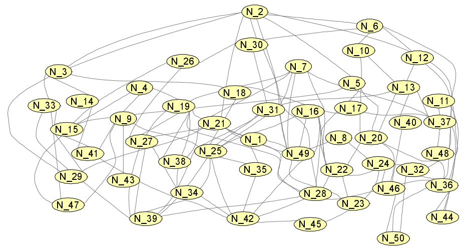

Figure 2. Simulation scenario

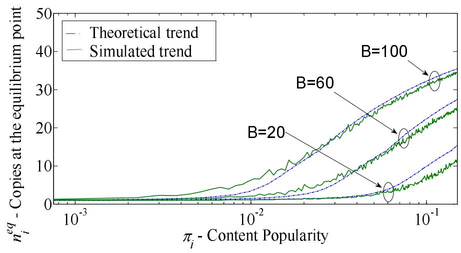

Figure 3. Number of content replicas in the network

References

[1] Gopinath, S., Jain, S., Makharia, S., and Raychaudhuri, D. An experimental study of the cache-and-forward network architecture in multi-hop wireless scenarios. In *Local and Metropolitan Area Networks (LANMAN), 2010 17th IEEE Workshop on* (may 2010), pp. 1 –6.

[2] Jacobson, V., Smetters, D. K., Thornton, J. D., Plass, M. F., Briggs, N. H., and Braynard, R. L. Networking named content. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies* (2009), CoNEXT '09.

[3] Koponen, T., Chawla, M., Chun, B., Ermolinskiy, A., Kim, K., Shenker, S., and Stoica, I. A data-oriented (and beyond) network architecture. *SIGCOMM Comput. Commun. Rev. 37* (Aug. 2007).

[4] Lagutin, D., Visala, K., and Zhang, L. Publish/subscribe for internet: Psirp perspective. In *Towards the Future Internet* (2010), IOS Press.

[5] Ohlman, B., et al. *First Netinf Architecture Description*, technical report ed. fp7-ict-2007-1-216041-4ward, Jan. 2009.

[6] Tortelli, M., Cianci, I., Grieco, L. A., Boggia, G., and Camarda, P. A fairness analysis of content centric networks. In *IFIP International Conference Network of the Future, Paris.* (Nov. 2011).