TRANSACTIONS ON EMERGING TELECOMMUNICATIONS TECHNOLOGIES
Trans. Emerging Tel. Tech. 2012; **00**:1–8
DOI: 10.1002/ett

SatCache: a profile-aware caching strategy for information-centric satellite networks

Salvatore D'Oro, Laura Galluccio, Giacomo Morabito, Sergio Palazzo

DIEEI – University of Catania

ABSTRACT

Information-centric networking (ICN) is a new networking paradigm which is attracting increasing attention from the scientific community. ICN focuses on content handling and distribution rather than host-to-host communications. Satellite systems have always played a key role in content distribution and therefore, they are expected to be fundamental components of the information-centric networking. However, it is not obvious whether solutions that are being developed for terrestrial ICNs are effective in satellite scenarios as well. More specifically, focus of this paper is on *in-network content caching*, which is an important feature characterizing all ICN proposals but, as we will show, should be re-designed to address the specific characteristics of satellite networks. Indeed, in this paper a novel caching scheme named *SatCache* is proposed which exploits the broadcast nature of the satellite communication medium and creates a profile of the preferences of network users in order to estimate their potential interest in a given content. The proposed approach is based on a simple model of the user behavior. By aggregating the content requests generated by a large number of users, a collective behavior can be observed which fits the output of large measurement campaigns discussed in the literature. Simulation results given in the paper show the effectiveness of the proposed scheme. Copyright © 2012 John Wiley & Sons, Ltd.

1. INTRODUCTION

In the last few years there has been an increasing interest in the so called *information-centric*[*] networking (ICN) paradigm, which is expected to be a key component of the Future Internet [14, 12].

Indeed, this interest is witnessed by the large number of research projects active on this topic worldwide and the number of successful research events and workshops held in the last few months (see [22] for a survey).

Motivation of the ICN paradigm is related to a simple evidence [12]: *"people value the Internet for* what *content it contains, but communication is still in terms of* where*"*. In other words the Internet was designed to support data exchange between pairs of hosts, while most users are interested in retrieving the desired data regardless of the specific hosts that store it.

Accordingly, there is a mismatch between the current architecture and protocols run by the Internet and how people use the Internet most of the time. The amazing success of the Internet and the importance of its role in our daily life are a clear evidence that such mismatch has been handled quite well; nevertheless, there are a few issues that are becoming more and more critical as user mobility increases and the importance of data authenticity increases [13].

Objective of the ICN paradigm is to make the network-layer aware of *what* it is handling, e.g., which contents users are requesting, which contents a node may provide, which contents a node is forwarding. Since its introduction a huge amount of research effort has been provided to the study of appropriate network layer solutions and several results are available in the literature [13, 12, 5].

Proposed solutions differ in several ways, e.g., how name should be constructed, how data request and data itself should be routed, the approach to be utilized at the transport layer, etc.; however, they all recognize the importance of *in-network caching*, that is the possibility to store (or "cache") a popular content *inside* the network to make it available near the potential users.

In this paper we focus on the problems of in-network caching for information-centric networks in a very interesting scenario: satellite networks. Indeed, satellite systems are expected to play a crucial role in the Future Internet thanks to their well known characteristics [7], especially in the content distribution domain where they have always been of primary importance [10, 9]. In this context, traditional in-network caching solutions are not effective and therefore, new solutions are needed.

[†]This is the peer reviewed version of the following article [23], which has been published in final form at http://onlinelibrary.wiley.com/doi/10.1002/ett.2723/full. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving

[*] Also referred to as *content-centric networking*.

Note that the opportunities and challenges arising when satellite links are included in ICNs have been the focus of some interesting works, i.e., [6, 19]. However, in-network caching has not been investigated to date.

In this paper we propose a new caching approach called *SatCache* for information-centric satellite networks. This exploits the broadcast nature of satellite communications and *users' profiles* to decide where caching any piece of content is more convenient. More specifically, we assume that each user is characterized by an interest profile which describes her level of interest on a certain topic. Also we assume that each content can be characterized through an interest profile and that users decide the contents they require depending on the difference between their interest profile and those of the content items. We will show that such a policy can be easily translated into a model of the behavior of the single user and that the resulting behavior of a large group of users accurately mimics what can be obtained in several measurement campaigns.

Effectiveness of the proposed approach for caching is assessed through numerical examples.

More specifically, the rest of this paper is organized as follows. In Section 2 we will briefly describe how in-network caching is executed in ICN systems and why it is ineffective in satellite networks. Then, in Section 3 we will introduce *SatCache*, our caching strategy for information-centric satellite networks, and in Section 4 we will detail the system model. SatCache will be assessed in Section 5. Finally, in Section 6 some concluding remarks will be drawn.

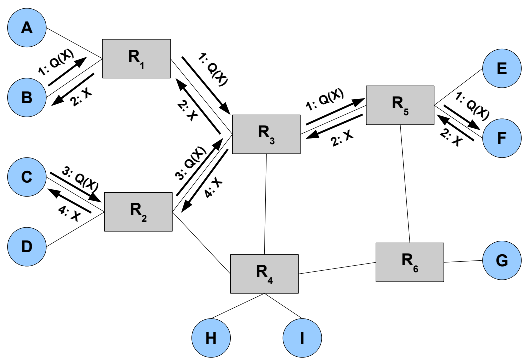

Figure 1. Caching in terrestrial networks: an example.

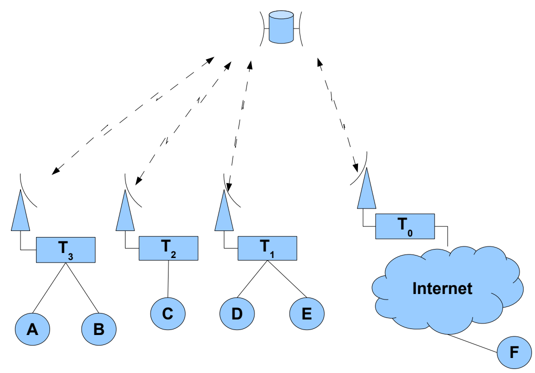

Figure 2. Exemplary satellite network configuration.

2. CACHING IN SATELLITE NETWORKS

In all ICN solutions, routers store a fraction of the data flowing through them in a local cache. Before forwarding messages requesting a certain data to the next hop, they check whether such data is currently stored in the local cache. If this is the case, then they send the requested data to the user and drop the data request. Otherwise they pass the data request to the next hop towards the destination of the requested data.

For example, consider the exemplary terrestrial network shown in Figure 1. Node B issues a request for a given content X. Such content has been published by node F and therefore, the request message $Q(X)$ traverses routers R_1, R_3, and R_5. Node F receives the request issued by node B and sends it the requested content X. Content X will traverse nodes R_5, R_3 and R_1 backward. These will store local copies of such data in their cache.

Then, suppose that node C generates a request for the same content X. The related request message $Q(X)$ will traverse routers R_2 and R_3. The latter – router R_3 – realizes that there is a copy of X in its cache and therefore, it does not forward the request to the next hop towards F. Instead R_3 sends the content X to node C.

Observe that exploiting the ICN approach, node C can authenticate X even if it is receiving the data from a node different from the publisher of the content, F.

It is obvious that exploiting in-network caching, network performance may be improved significantly in terms of both network resource utilization, data availability, and data delivery delay†.

In satellite networking scenarios traditional in-network caching solutions are ineffective. For example in Figure 2 we show a typical satellite network configuration. The satellite communication system provides a certain number of user nodes, that is, A, B, C, D, and E, with Internet access. Indeed, each user node is connected to a satellite terminal and one of such satellite terminals, T_0, is connected to the rest of the Internet through some broadband communication technology.

Now suppose that, as in the terrestrial case described in Figure 1, node B and then node C request the same content X which has been published by node F. If this is the case,

†Interesting studies on the feasibility and utility of in-network caching can be found in [17] and [18]

Trans. Emerging Tel. Tech. 2012; **00**:1–8 © 2012 John Wiley & Sons, Ltd.
DOI: 10.1002/ett
Prepared using ettauth.cls

then node C can retrieve the content X from the cache located in terminal T_0. However, observe that:

- *The increase in network resource efficiency is negligible*. In fact, content X must be transmitted through the satellite communication system again and therefore, the utilization of satellite communication resources, which are the scarcest and, thus, the most precious, does not decrease.
- *The delay reduction is negligible*. In fact, the cached copy of content X must traverse the satellite communication segment which, in typical network configurations, introduces most of the delay.

Accordingly, traditional caching solutions in satellite networks are ineffective and novel appropriate solutions are needed.

3. SATCACHE

In this section we introduce *SatCache* a caching scheme for information centric satellite networks. More specifically, in Section 3.1 we provide the major guidelines that should be considered in the design of the caching strategy. Then, in Section 3.2 we describe the operations executed by SatCache.

3.1. Design requirements

The most obvious solution for caching in satellite networks is to make terminals store all the data transmitted by the satellite in a local cache. Although, in principle, this is feasible thanks to the broadcast nature of satellite communications, in practice this cannot be done because it would require very sophisticated and fast cache management schemes, which are not available with current technologies. Accordingly, novel solutions are required that support the following design requirements:

- Data should be cached as close as possible to users that are potentially interested in it. Obviously such requirement can be easily supported by caching data in the satellite terminals.
- The caching strategy should take *information locality* into account. Indeed, there are experimental evidences that the probability that two users are interested in the same information increases as their geographical distance decreases [4, 21].
- Data that is most likely used by interested users should be cached. To achieve such a goal, satellite terminals will take into account the interests of the users they are providing with network access.
- The proposed solution should not require modifications in the utilized ICN approach and technology. In fact, users want to use the same applications regardless of the access network they are exploiting.

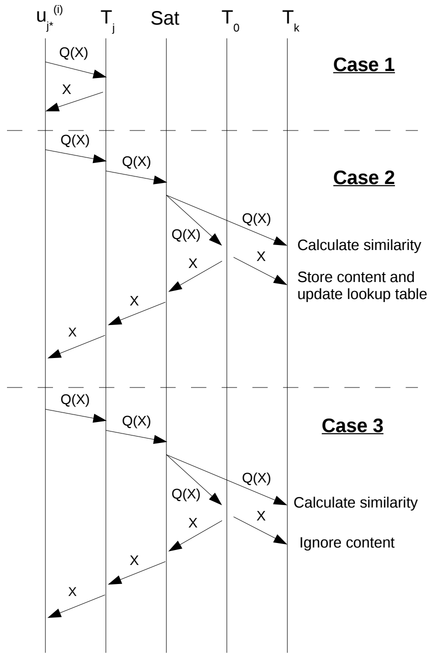

Figure 3. Exchange of signaling associated to the SatCache scheme.

3.2. SatCache scheme

In this section we describe the SatCache scheme more in detail. In fact, in Section 3.2.1 we show how each terminal evaluates the degree of interest of its users on a given content. In Section 3.2.2 we describe how the cache is managed and how messages are exchanged in SatCache.

We now provide some definitions and notation that will be used in the following of this section.

Consider a satellite access network consisting of $(M + 1)$ satellite terminals, which we denote as T_0, T_1, ..., T_M. We assume that T_0 is connected to the Internet as shown in Figure 2. Each of the other satellite terminals gives access to a certain number of user hosts. We denote the set of N_i users connected to T_i as $U_i = \{u_1^{(i)}, u_2^{(i)}, ..., u_{N_i}^{(i)}\}$. Observe that usually all the users connected to the same satellite terminal are close to each other and therefore, we identify their positions with the position of the satellite terminal $p(T_i)$.

Let us suppose that L pieces of content are available in the network and let $\Gamma = \{X_1, X_2, ..., X_L\}$ denote the set of such contents. Observe that a content X can be also characterized by a reference *position*, $p(X)$, which represents the geographical position that is the most relevant to X. As an example if content X is "weather forecast for New York city", $p(X)$ will be the geographical position of the city of New York to which the content X refers to.

Furthermore, as in [15] and [16], we suppose that both the users of the satellite network and the contents are characterized by the so-called *interest profile*:

- The **interest profile of a user** u is denoted as $ip(u)$ and represents the interest of such user in certain topics or her belonging to some communities or interest groups.
- The **interest profile of a content** X is denoted as $ip(X)$ and represents the relevance of a certain content to a given topic or to some community or interest group.

Accordingly, as we will explain better in the following section, we can represent both users and contents as elements in the same space denoted as *interest space* and detailed below.

Definition (Interest space) – The *interest space* is an m-dimensional unit cube $C = [0, 1]^m$, where m is the total number of interests defined in the network.

Interests can both refer to certain topics – e.g., cinema, literature, football, etc. – or be associated to a physical or virtual community – such as, for example, a Facebook interest group, the supporters of a given political party, etc. Note that the definition of the interest space in an actual implementation may be modified over time.

Accordingly, the interest profile of a user u can be represented as an m-dimensional vector in the interest space, that is $ip(u) \in C$, reporting in each of the m dimensions the degree of interest of user u in a certain topic.

It is obvious that the more similar are the interest profiles of a user u and a content X, the higher is the degree of interest of u on X. It is also obvious that the smaller is the distance between the position of user u, that is $p(u)$, and the reference position of X, that is $p(X)$, the higher becomes the degree of interest of u in X – for example, the closer is a certain geographical location to a user u, the higher is the interest of u in the weather forecast for such location.

In the following we will explain how the SatCache scheme exploits the above concepts.

3.2.1. Similarity evaluation

The most obvious way to evaluate the potential interest of a user u to a given content X would be to insert a few *meta tags* in the content itself – which describe the topics covered by X – and then to check whether such meta tags match with the interest profile $ip(u)$. However, this requires the involvement of content publishers, which may lead to two problems. In fact, on the one hand publishers may not be willing to mark their content just to support the SatCache scheme – actually they already do this to support search engines. On the other hand, such approach is vulnerable to misbehavior of content publishers. Indeed, in order to increase the probability that a given content is requested by users, they could insert meta tags which do not represent the actual content but are indeed very popular.

As a consequence, we use a different approach based on the fact that people with similar profiles are likely to be interested in the same contents [15]. Indeed, a satellite terminal decides whether to store a certain content X in the cache or not based on

- the similarity between the interest profile of its users and the user that requested X,
- its distance from the user that requested X.

It follows that we need to define an appropriate parameter that is able to quantify the above aspects together. To this purpose, we extend the concept of interest profile to include the position of the user (or the reference position of a content). In other words, given a user u with interest profile $ip(u)$ and located in $p(u)$ we define[‡] $ip^*(u)$ as the Cartesian product of $ip(u)$ and $p(u)$, that is $ip^*(u) = ip(u) \times p(u)$.

It follows that $ip^*(u)$ is an *extended interest profile* that can be represented with an $(m + 2)$-dimensional vector.

Given such definitions, the satellite terminal can use the *similarity* between the extended interest profiles of its users and the extended interest profile of the user that requested X to estimate the potential interest of its users in content X. Therefore, a methodology is needed to measure the similarity between two extended interest profiles. To this purpose we use the so called *cosine similarity* [20]:

Definition (Cosine similarity) – Given two extended interest profiles $ip^*(u_1)$ and $ip^*(u_2)$, we estimate their similarity as follows:

$$\Theta\left(ip^*(u_1), ip^*(u_2)\right) = \frac{ip^*(u_1) \cdot ip^*(u_2)}{|ip^*(u_1)|\ |ip^*(u_2)|} \quad (1)$$

Note that with eq. (1) we can take both the similarity between the interests of different users and the distance between their positions into account through a unique parameter.

3.2.2. SatCache operations

The operations performed by SatCache are summarized in Figure 3.

Let us assume that a user $u_{j^*}^{(i)}$, whose position is $p(u_{j^*}^{(i)})$ and interest profile is $ip(u_{j^*}^{(i)})$ accesses the network through the satellite terminal T_i. To this purpose she issues a request for content X. Such request, identified as $Q(X)$, contains the user identifier, its position, her interest profile and the content identifier X and reaches the satellite terminal T_i. Accordingly, three different cases should be considered. In *Case 1* the satellite terminal T_i has a copy of content X in its cache and therefore, as shown on top of Figure 3, drops $Q(X)$ and transmits content X to $u_{j^*}^{(i)}$.

In the other cases, the satellite terminal forwards the message to the satellite transponder which sends it back to the earth station using the downlink. Such $Q(X)$ is directed to the satellite terminal T_0 which connects the satellite communication system with the network

‡ The same definition applies in case we focus on a content rather than on a user.

infrastructure; however all the other satellite terminals can overhear. Let us focus on the operations run by one of them, say T_k.

Upon receiving the request $Q(X)$, the satellite terminal T_k reconstructs the extended interest profile of $u_{j^*}^{(i)}$ and evaluates the maximum similarity, $\Theta_k^{(\max)}(u_{j^*}^{(i)})$, between $ip^*(u_{j^*}^{(i)})$ and the extended interest profile of its users, that is

$$\Theta_k^{(\max)}(u_{j^*}^{(i)}) = \max_{u \in U_k} \left\{ \Theta \left(ip^*(u_{j^*}^{(i)}), ip^*(u) \right) \right\} \quad (2)$$

Then, based on the value of $\Theta_k^{(\max)}(u_{j^*}^{(i)})$ we can further distinguish two different cases. In fact, if such value is higher than a given threshold, that is $\Theta_k^{(\max)}(u_{j^*}^{(i)}) > \Theta_{\mathrm{Thr}}$ then we are in *Case 2* and the satellite terminal T_k will execute the operations required to store the content X. Indeed, it will update the cache lookup table according to the *Least Recently Used* (LRU) mechanism [1] and then store content X when this reaches the earth after it is transmitted by the satellite terminal T_0 and relayed by the satellite.

The last case, that is *Case 3*, occurs when $\Theta_k^{(\max)}(u_{j^*}^{(i)}) \leq \Theta_{\mathrm{Thr}}$. If this the case, then the content X should not be stored in the cache and therefore, no specific actions are taken by the satellite terminal T_k when it receives content X.

4. SYSTEM MODEL: DESCRIPTION AND VALIDATION

In this section we present a simple model that generates the pattern of multimedia content requests issued by users characterized by the interest profile discussed above in Section 3.2. The model, which will be described in Section 4.1, is based on a simple intuition on user behavior: people tend to ask content items which are popular, in line with their interests, and which they have not requested in the most recent past. By implementing this simple rule, our model is able to capture the collective behavior of multimedia content users that has been already described in several scientific contributions.

In Section 4.2 we will validate our model by comparing the statistical characteristics of its output to the statistical characteristics which have been observed in previous works.

4.1. Model description

A model can be considered a *good* one to characterize a certain phenomenon if it is able to capture the key characteristics of such phenomenon by using only a few parameters. Accordingly, it is the result of a tradeoff between accuracy and simplicity.

Our model of the multimedia content requests issued by users is based on the following facts:

1. Different users have different tastes, which are reflected in their interest profile, and the content they decide to enjoy depends on such profile.
2. Multimedia industry produces contents which are likely to be of interest of users.
3. Users can select content items that they are aware of; the probability that a user is aware of a given content item is strongly related to the current popularity of such content item. This popularity changes over time.
4. Users refrain from requesting the same content item several times within a short time frame, if the cost for enjoying such content (in terms of money or required time) is not negligible.

More specifically, we consider a model in which time is slotted. Our model assumes a certain number of users, N, characterized by an interest profile which is distributed in an n-dimensional interest space according to the probability density function $f_U(\mathbf{x})$, with $\mathbf{x} \in [0,1]^n$. We assume that, at each time slot, all users generate a content request.

Objective of the multimedia content producers is to publish content items which are likely to meet the tastes of the public. Therefore, we suppose that there are K multimedia content items characterized by an interest profile which is distributed within the interest space based on a probability density function $f_X(\mathbf{x})$, with $\mathbf{x} \in [0,1]^n$, which is equal to $f_U(\mathbf{x})$, i.e,

$$f_X(\mathbf{x}) = f_U(\mathbf{x}) \qquad \forall \mathbf{x} \in [0,1]^n \quad (3)$$

At any time slot t, a certain number of content items K_P are *popular*. We denote the set of popular content items at time slot t as $\Psi(t)$. Initially the set of popular content items is generated at random. Later on, in this section, we will explain how the set of popular content items is updated over time.

Furthermore, each user u maintains an updated list of the last L contents it has requested. We call such list the *recently requested content list*, $\Lambda_u(t)$.

At a given time slot, t, a user u randomly selects v_P content items from the list of popular contents and v_R content items from the remaining content items. We denote the set of such $(v_P + v_R)$ content items as $\Gamma_u(t)$. Note that by choosing the above parameters in such a way that $v_P/(v_P + v_R) >> K_P/K$, popular content items are requested more frequently than others, as obvious. A user is not expected to request a content which it has requested recently; accordingly, it removes from $\Gamma_u(t)$ the elements which are included in the recently requested content list $\Lambda_u(t)$. Finally, user u requests the content item belonging to $\Gamma_u(t) - \Lambda_u(t)$ which maximizes the cosine similarity, defined as given in eq. (1). In other words, at time slot t user u requests the content item $\xi_u(t)$ selected as follows:

$$\xi_u(t) = \arg \max_{x \in \Gamma_u(t) - \Lambda_u(t)} \{\Theta\left(ip(u), ip(x)\right)\} \quad (4)$$

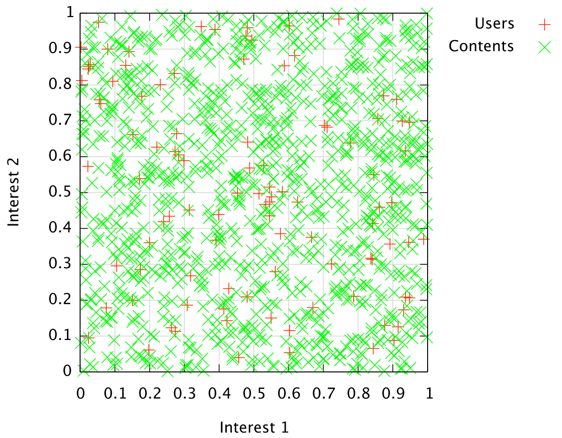

Figure 4. Interest profiles of $N = 100$ users and $K = 1000$ content items in a two-dimension interest space.

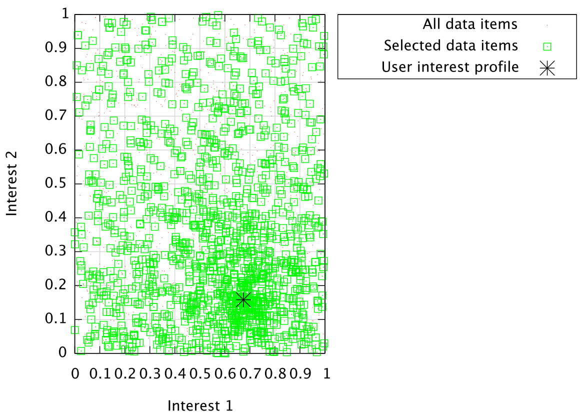

Figure 5. Representation of the values of Interest 1 and Interest 2 of user $i = 1$ and of the first 1000 contents that she requests.

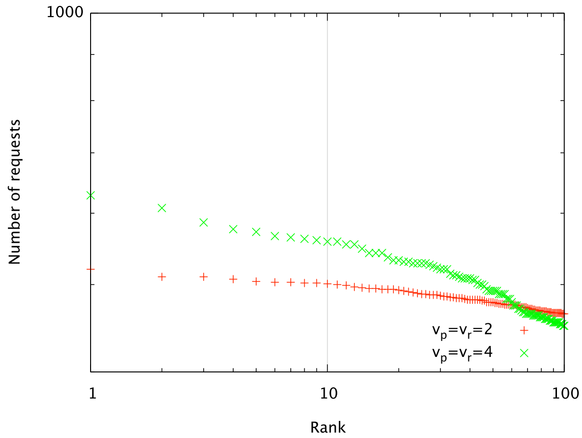

Figure 6. Number of requests of a given content vs its ranking for the 100 most requested content items.

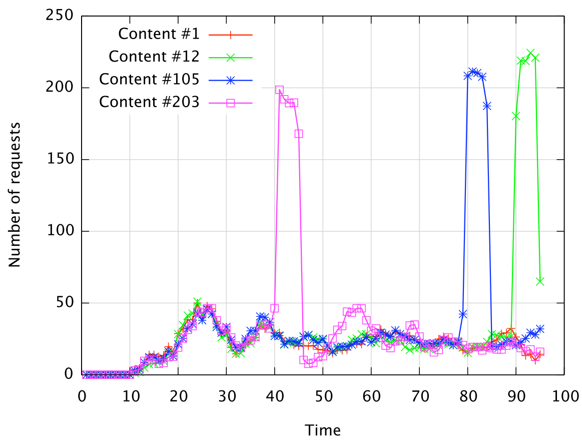

Figure 7. Number of requests vs. time for four different content items.

If the selected content is not a popular content, i.e., $\xi_u(t) \notin \Psi(t)$, we assume there is always the chance that it becomes popular later on. Indeed, we assume that $\xi_u(t)$ enters in the list of popular contents $\Psi(t)$ with a certain probability p_Ψ. If this is the case, then the *least recently used* (LRU) content item is removed from the list of popular contents $\Psi(t)$.

4.2. Model validation

In this section we validate the model proposed in the previous section. We consider the case in which there are N =100 users and $K = 1000$ contents uniformly distributed in a two-dimensional interest space. In Figure 4, we represent the interest profile of users and contents in the two-dimensional interest space.

Also we suppose that the number of popular contents is $K_P = 200$ while the number of contents included in the recently requested content list is $L = 50$.

For example, in Figure 5 we show the interest profiles of the first 1000 contents that are requested by the user 1 assuming $v_p = v_r = 4$. In the same figure we represent the interest profile of user 1. As expected user 1 will request contents with interest profiles similar to hers, mostly.

In Figure 6 we show the overall request rate for the 100 most popular content items. The two curves represented in the figure have been obtained by assuming two different settings for the parameters v_p and v_r. Note that in the logarithmic scale the curves decrease linearly, which means they exhibit a Zipf distribution which is in line with a large number of measurement campaigns [2, 3].

Finally, in Figure 7 we show the number of requests over time for four content items. We have shown the number of requests for content 1 which do never enter the set of popular contents $\Psi(t)$, and three other content items which are selected in the set $\Psi(t)$ at different time instants. Observe that the pattern of requests when a content item enters the set of popular contents is characterized by a rapid increase in the number of requests and a rapid decrease after most users enjoy the content. Such behavior has been observed experimentally in several measurement campaigns [3, 8].

Trans. Emerging Tel. Tech. 2012; **00**:1–8 © 2012 John Wiley & Sons, Ltd.
DOI: 10.1002/ett
Prepared using ettauth.cls

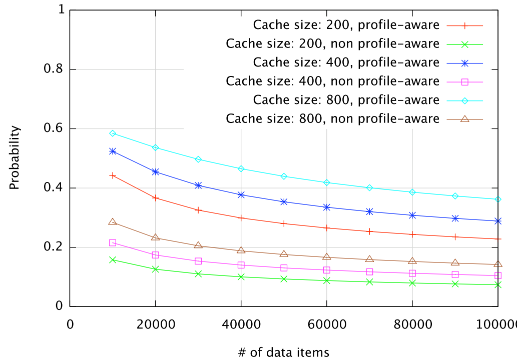

Figure 8. Hit probability versus the number of contents in the network for different values of the cache size and for both the case in which the cache management is aware of the interest profile of users and the case when it is interest profile oblivious.

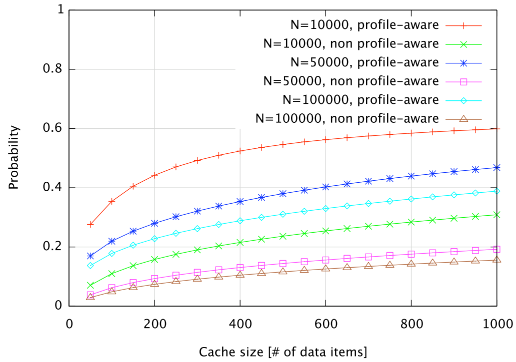

Figure 9. Hit probability versus the cache size for different values of the number of contents available in the network and for both the case in which the cache management is aware of the interest profile of users and the case when it is interest profile oblivious.

5. NUMERICAL EXAMPLES

In this section we assess the effectiveness of SatCache through some numerical examples. More specifically we consider a system consisting of $M = 100$ satellite terminals, each of which is connected to a user only, that is $N_i = 1$ for all $i \leq M$. Furthermore, we consider that the positions of users are uniformly distributed on a squared area with edge 1000 km. As far as the interest space is concerned, we will assume that the number of dimensions of the interest space is equal to $m = 6$ and that the interest profiles of the users are randomly distributed in the interest space.

Regarding the content request generation process, we assume that users generate requests according to an exponential distribution. Furthermore, we assume that each user generating a request selects the specific content item according to the model described in Section 4.1.

In Figure 8 we show the *hit probability*, that is, the probability that a user requests a content which is currently stored in the cache of her terminal versus the value of the number of contents in the network. In this figure we represent the hit probability obtained by SatCache for different sizes of the cache. Furthermore, in order to estimate the performance improvement obtained thanks to social awareness, we represent the hit probability for both the case in which the caching strategy takes user profiles into account and the case in which the caching strategy is oblivious with respect to the social profile of users (therefore, it is a simple Least Recently Used mechanism [1]). As expected, performance improves as the cache size increases and becomes worse as the number of contents available in the network increases.

In order to achieve a better understanding of the impact of cache size into the performance of SatCache, in Figure 9 we plot the hit probability versus the number of data items that can be stored in the cache. We have represented different curves for different values of the number of contents available in the network, and again we have considered both the case in which the cache management strategy is aware of users' social profile and the case in which it is not.

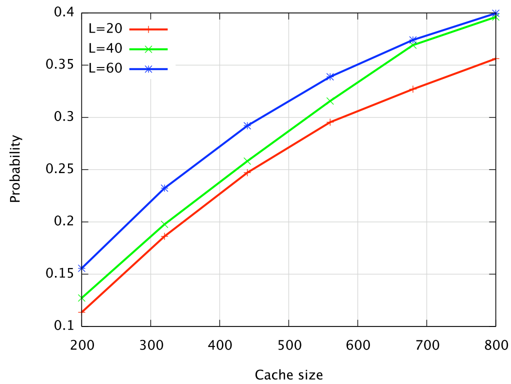

Figure 10. Hit probability versus the cache size for different values of the number of contents in the recently requested content list $\Lambda_i(t)$.

Observe that the hit probability results more than doubled by simply adding into the system awareness on the user profile.

Finally, in Figure 10 we represent the hit probability versus the cache size for different values of L, that is the size of the recently requested content list, $\Lambda_u(t)$. In the figure as expected we observe that the hit probability increases when the value of L increases.

6. CONCLUSION

In this paper we have introduced SatCache, an in-network caching scheme for information-centric satellite networks which exploits the broadcast nature of the satellite channel, information locality and the, so called, interest profile of users and contents to improve communication resource efficiency. Our study shows that the use of appropriate similarity estimation between interest profiles of users and contents can significantly help to speed up the content delivery, thus reducing the cost in terms of satellite bandwidth resource waste and the delay as compared to non interest profile aware approaches.

Another contribution of this paper is that it introduces a simple model for characterizing the behavior of users which provides patterns of content requests fitting with the results of large measurement campaigns available in the literature. This paper thus provides interesting guidelines for the design of caching systems to be used in information-centric networks.

ACKNOWLEDGMENT

The work of Giacomo Morabito was partially supported by the FP7/NICT EU-JAPAN GreenICN project.

REFERENCES

1. R. Alonso, D. Barbara, and H. Garcia-Molina. Data caching issues in an information retrieval system. *ACM Transactions on Database Systems*. Vol. 15, No. 3. Mar. 1990.
2. L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web Caching and Zipf-like Distributions: Evidence and Implication. *Proc. of IEEE Infocom*. Apr. 1999.
3. M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. *Proc. of ACM IMC*. Oct. 2007.
4. L. Cherkasova and M. Gupta. Characterizing locality, evolution, and life span of accesses in enterprise media server workloads. *Proc. of ACM NOSSDAV*. May 2002.
5. A. Detti, N. Blefari Melazzi, S. Salsano and M. Pomposini. CONET: A Content Centric Inter-Networking Architecture. *Proc. of ACM Sigcomm - ICN*. Aug. 2011.
6. A. Detti and A. Caponi and N. Blefari-Melazzi. Exploitation of information-centric networking principles in satellite networks. *Proc. of IEEE ESTEL 2012* . Oct. 2012.
7. G. Fairhurst, L. Caviglione, and B. Collini-Nocker. FIRST: Future Internet: A Role for Satellite Technology. *Proc. of IEEE IWSSC*. Oct. 2008.
8. F. Figueiredo, F. Benevenuto, and J. M. Almeida. The Tube over Time: Characterizing Popularity Growth of YouTube Videos. *Proc. of ACM WSDM*. Feb. 2011.
9. L. Galluccio, G. Morabito, and G. Schembra. Transmission of Adaptive MPEG Video over Time-Varying Wireless Channels: Modeling and Performance Evaluation. *IEEE Trans. on Wireless Communications.* Vol. 4, No. 6. Nov. 2005.
10. L. Galluccio, S. Palazzo, A. Panto'. Support of Delayed Real-Time Services over Geo Satellite Networks through a Recovery Service. *Proc. of IEEE Globecom*. Nov. 2007.
11. L. Galluccio, G. Morabito, and S. Palazzo. Caching in information-centric satellite networks. *Proc. of IEEE ICC*. Jun. 2012.
12. V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard. Networking Named Content. *ACM CoNEXT* Dec. 2009.
13. T. Koponen, M. Chawla, B.-G. Chun, A. Ermolinskiy, K. H. Kim, S. Shenker, and I. Stoica. A Data-Oriented (and Beyond) Network Architecture. *Proc. of ACM Sigcomm*. Aug. 2007.
14. D. Kutscher, H. Flinck, and H. Karl. Information-Centric Networking: A position paper. *Proc. of 6th GI/ITG KuVS Workshop on Future Internet* Nov. 2010.
15. A. Mei, G. Morabito, P. Santi, and J. Stefa. Show me your friends and I'll tell you what you like. *Proc. of Extremecomm*. Sep. 2010.
16. A. Mei, G. Morabito, P. Santi, and J. Stefa. Social-aware stateless forwarding in pocket switched networks. *IEEE Infocom*. Apr. 2011.
17. L. Muscariello, G. Carofiglio and M. Gallo. Bandwidth and storage sharing performance in information centric networking. *Proc. of ACM Sigcomm - ICN*. Aug. 2011.
18. D. Perino and M. Varvello. A Reality Check for Content Centric Networking. *Proc. of ACM Sigcomm - ICN*. Aug. 2011.
19. V. A. Siris and C. N. Ververidis and G. C. Polyzos and K. P. Liolis. Information-Centric Networking (ICN) architectures for integration of satellites into the Future Internet. *Proc. of IEEE ESTEL 2012* . Oct. 2012.
20. P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison-Wesley. 2005.
21. H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng. Understanding user behavior in large-scale video-on-demand systems. *ACM SIGOPS Operating System Review*. Vol. 40, No. 4. Apr. 2006.
22. http://trac.tools.ietf.org/group/irtf/trac/wiki/icnrg
23. D'Oro, Salvatore and Galluccio, Laura and Morabito, Giacomo and Palazzo, Sergio. SatCache: a profile-aware caching strategy for information-centric satellite networks. *Transactions on Emerging Telecommunications Technologies*. Vol. 25, No. 4. 2014