TECHNICAL REPORT

IIT TR-09/2023

Analytical Model for N-dimensional Multi-Channel Data Dissemination in Opportunistic Network

A. Pinizzotto

Analytical Model for N-dimensional Multi-Channel Data Dissemination in Opportunistic Networks

Antonio Pinizzotto
Institute for Informatics and Telematics, National Research Council of Italy

Abstract

In opportunistic networks data dissemination is an important, although not widely explored, topic. Since opportunistic networks topologies are very challenged and unstable, data-centric approaches are an interesting directions to pursue. Data should be proactively and cooperatively disseminated from sources towards possibly interested receivers, as sources and receivers might not be aware of each other, and never get in touch directly. In this work we consider a utility-based cooperative data dissemination system in which the utility of data is based on the social relationships between users. Specifically, we study the performance of this system through an analytical model. Our model allows us to characterise the data dissemination process, as it describes both its stationary and transient regimes.

This work aims to extend an analytical model previously obtained for a scenario of 2 channels, to the case of n channels. A case study for 3 channels is first introduced to better understand the n-channel general case analysis.

Index Terms

opportunistic networks, data dissemination, delay tolerant networks.

I. Introduction

In [1] a data dissemination system is analyzed and modeled assuming that nodes can at most subscribe to two channel[1]. In other words, each node is supposed to have chosen a channel of interest and to maintain this choice along the time. Each node has a buffer that can store only one unit data and the node can decide which data to store each time it meets another node. The decision is based on a utility function whose variables are the initial distribution of the node preferences, and the current distribution of the number of nodes actually storing each channel data.

The purpose of the present work is to generalize the model in [1] from the simple case of two channel to the n channels case. To perform a direct comparison with the simulation results presented in [1] for the 2 channels case, we used the same assumptions for the cost function and the distribution of nodes' interests.

The rest of the report is organised as follows. Section II overviews the data dissemination system model. Section III introduces a study on a 3-channel case to familiarise in a visual way with the problem. In Section IV for the general case of N channels a comprehensive analytic study is presented with the most important results. Finally In Section V we verify the analytical model by simulation and in Section VI a summary of the results is reported.

II. The Data Dissemination System Model

The reference scenario we deal with in this work is similar to the one used in PodNet [2], named "podcasting for ad hoc networks". As in the paradigm of a typical opportunistic network, we consider a number of mobile users whose devices cannot be continuously connected to the Internet. Communication is possible by opportunistically exploiting pair-wise intermittent contacts between users to exchange messages, and bring them towards their final destinations. Sporadic contacts of users with point of access to the Internet are also possible. In podcasting applications, data objects (e.g., software updates, music or video files, advertisements, ...) are organised in different *channels* to which users can subscribe. Data objects might be generated both from within the Internet or by the mobile users. The data dissemination system defined in [3] is responsible for managing subscriptions, and bringing data objects to subscribed users.

A. Utility-based data dissemination system

A data dissemination system has to specify mechanisms for managing subscriptions and delivering data to subscribed users. In our reference framework each node advertises the channels its user is subscribed to upon making contact with any other node, and the subscription to a channel is performed just at the beginning of the events.

The typical form of the utility function is the product of the access probability to the data object (p_{ac}) by a measure of the retrieval cost (c), normalised by the object's size (s). The rationale of this definition is that the utility of an object should be high if it is very popular and costly to be retrieved. Normalising by the size is usually just a way to simplify the approximate

The activity of this report was carried out in collaboration with Marco Conti, Chiara Boldrini and Andrea Passarella, from Institute for Informatics and Telematics, National Research Council of Italy.

[1]A channel is set of data regarding a particular topic such as software update, podcasting, music and video files

solution of the resulting knapsack problem. In our framework we use exactly this kind of utility function, and we define the cost c as a monotonically decreasing function (denoted as $f_c()$) of the object's availability in the network (hereafter referred to as p_{av}). Specifically, p_{av} is defined as the probability of finding the object in the cache of any node. Clearly, the higher p_{av}, the lower the cost to retrieve the object, the lesser its value. Different types of functions can be used for $f_c()$ such as exponential or linear ones.

In our opportunistic networking scenario, the users of any node cache are i) the local user, and ii) the users of the other encountered nodes. Our utility function takes this into consideration by defining multiple components, one for the local user and one for any social community the user is in contact with. In this paper, we consider a simplified, yet significant, version, made up of two components only, one related to the local user ($u^{(l)}$), the other (the *social* component, $u^{(s)}$) aggregating the utility for all the other users encountered in the past:

$$U = u^{(l)} + u^{(s)} = p_{ac}^{(l)} f_c(p_{av}^{(l)}) + p_{ac}^{(s)} f_c(p_{av}^{(s)}). \tag{1}$$

In Equation 1: i) $p_{ac}^{(l)}$ represents the probability that the local user is interested in the data object; ii) $p_{ac}^{(s)}$ represents the probability of meeting nodes interested in the data object; iii) $p_{av}^{(l)}$ represents the probability that the local node "sees" the data object in the caches of encountered peers; and iv) $p_{av}^{(s)}$ represents the average probability (over encountered nodes), i.e. the probability that those nodes "see" the data object in the caches of nodes *they* encounter.

B. Assumptions in the model

In our model we assume that the time is slotted, and nodes compute utilities at the beginning of each time slot, storing then the data from the channel with the higher utility. Furthermore we assume that nodes have global knowledge of the state of the network, in particular of the number of nodes storing each group i data in their buffer. In addiction the nodes are supposed to be able to access to any data without taking into account any mobility model. In this way we can model the evolution of the data distribution process with a Markov chain whose status is the vector $\mathbf{n} = (n_1, \ldots, n_N)$, where N is the number of channels, and n_i the number of nodes storing objects of channel i. This Markov chain completely describes the data distribution process in the network. Specifically, since the chain is finite, stationary distributions always exist.

For the sake of explanation, let us focus on a user subscribed to channel j, and let us evaluate its utility parameters with respect to channel i. The local access probability to is $p_{ac}^{(l)} = \mathbf{1}_{ij}$ where $\mathbf{1}_{ij}$ is the standard indicator function. The social access probability is the probability of meeting a node subscribed to channel i. Under our assumptions, this is the probability that any given node subscribes to channel i (throughout referred to as z_i, while $\mathbf{z} = (z_1, \ldots, z_N)$ is the distribution vector for all the z_i) [2]. Since we assume that nodes can compute exact p_{av} parameters, $p_{av}^{(l)}$ and $p_{av}^{(s)}$ are both equal to n_i/M, where M is the number of nodes in the system. Therefore, the utility of channel i computed by any node subscribed to j is

$$U_{ij} = (\mathbf{1}_{ij} + z_i) \cdot f_c(\frac{n_i}{M}). \tag{2}$$

The properties of the Markov chain can be analysed by exploiting the fundamental observation that, all nodes subscribed to j store data objects of the channel $\hat{\imath}$ such that:

$$\hat{\imath} = \arg\max_i \{U_{ij}\}. \tag{3}$$

Without loosing in generality, we make the assumption:

$$f_c() > 0 \tag{4}$$

The reason will be clear in the following, but we can anticipate that it will simplify the model. In the following we will denote with p_i the probability that a node initially stores a data item of channel i.

III. Analysis of the 3 channels case

Before approaching the general case of N channels, we analyze in details the 3 channels case. The reason is that for $N = 3$ it is possible to have a visual interpretation of the space state and this allows a deeper insight in the system behaviour. Such a visualization is difficult for $N = 4$ and not possible for $N > 4$. In the following there is then a detailed analysis of the system behavior in the case of 3 channels, in terms of stability, when varying some key parameters.

[2]Note that we are assuming that the probability distribution of subscribing to channels is the same for all nodes.

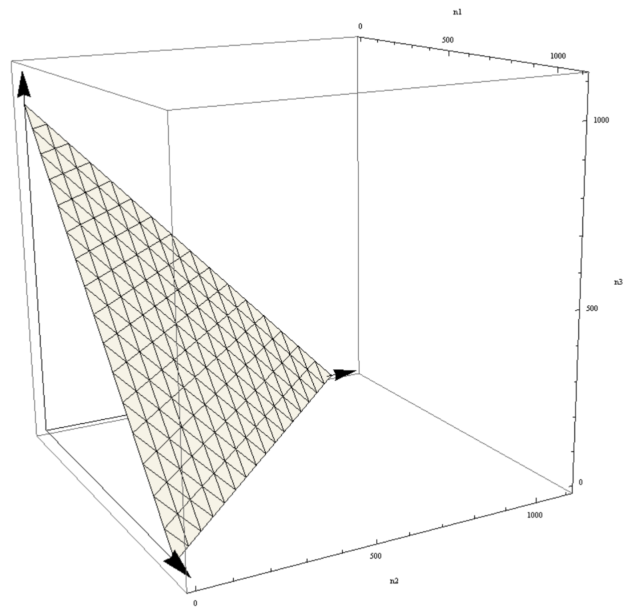

Fig. 1. Domain of states in the case $N = 3$: all the possible states are on the triangle.

A. Space of States

In case of 2 channels the state space of the system is very simple. We have two groups of nodes, and at each step, each node can decide to store one of two types of data. So the total number of possible states is given by the combined choice made by each group of users. As each group can decide indipendently what channel to store, the number of possible states is N^N, where N is the number of channels. For example, for $N = 2$, there are 4 possible states: $(M, 0)$, $(0, M)$, (Mz_1, Mz_2) and (Mz_2, Mz_1). Actually, with the assumption $f_c() > 0$ it can be proved that the state (Mz_2, Mz_1) is not possible. In the case $N = 3$ the number of possible states is then $3^3 = 27$. Let us consider another condition on the space state, that is:

$$\sum_{k=1}^{N} n_k = M \tag{5}$$

where N is the number of channels. Exploiting this constraint, it is possible to represent the states belonging to the $\mathbb{R}^N$ space, in the projected subspace $\mathbb{R}^{N-1}$. In fact, for example, in the case $N = 3$ all the states belong to the plain depicted in the Figure 1, that is the 27 possible states belong to the depicted triangle. It means that we can consider their projection on the plane (n_1, n_2) without losing any information, because for each of them we can obtain the 3^{rd} component, knowing that $n_3 = M - n_1 - n_2$. In the following we will identify each of these states using either the notations $\mathbf{n} = (n_1, n_2, n_3)$ where n_i is the number of nodes storing data from group i, or $\mathbf{c} = (c_1, c_2, c_3)$ where $c_i \in \{1, 2, 3\}$ and represents the index of the data channel stored by the user from group i. For example the state $\mathbf{c} = (1, 1, 2)$ corresponds to the state $\mathbf{n} = (2, 1, 0)$. It is worth nothing that the notation $\mathbf{c}$ gives more detailed information, while in the notation $\mathbf{n}$ the groups nodes choices are implicit. In other words to a state $\mathbf{c}$ can correspond more states $\mathbf{n}$.

Another notation that will be used is G_j to represent the nodes/users of group j, that is the nodes whose users have subscribed to the channel j.

Moreover, given a state $\mathbf{c}$, let's define $\Psi(\mathbf{c})$ the next state for $\mathbf{c}$, and similarly $\Psi(\mathbf{n})$ the next state for $\mathbf{n}$.

B. Zones of decision per single user group

Altough the number of states is 27, each of them can be taken into account only if there is another state from which the system can reach them. To evaluate this condition let us consider the decision zones of each group of users. Figure 2 depicts the 3 utility $U_{i1}(n_1, n_2)$ functions of the nodes G_1 for an exponential $f_c()$. Each utility function is a surface, and has an intersection with the other utility function. The intersections are lines, and the three lines have a common point. Let us define the following variables:

$$H_{ij}^{*(k)} = U_{ik} \cap U_{jk} \tag{6}$$

$$H^{(k)} = U_{1k} \cap U_{2k} \cap U_{3k} \tag{7}$$

All of these objects have to be considered on their projection in the plane (n_1, n_2). Each of the three lines $H_{ij}^{(k)}$ divide the plane into two half-planes corresponding to a different choice for the user G_k. Combining all these half-planes, with mutual

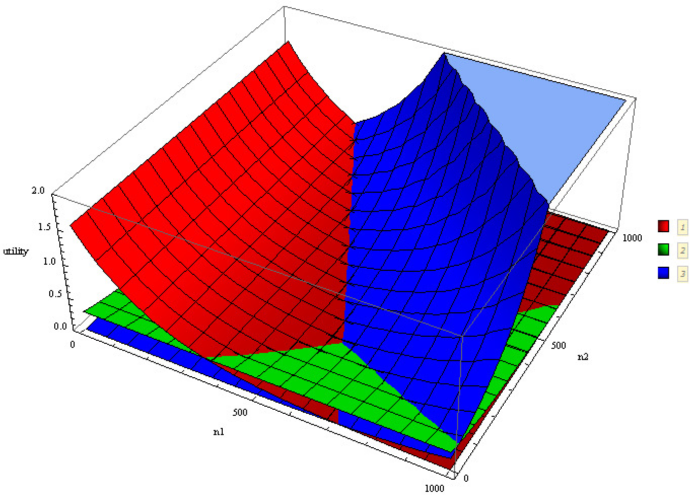

Fig. 2. Utility functions in the case $N = 3$ (the state space is in $\mathbb{R}^2$ (n_1, n_2) as a projection of the complete space $\mathbb{R}^3$)

intersections, we obtain three convex polytopes [4] which can be unambiguously determined by proper lines $H_{ij}^{(k)}$ obtained by a proper truncation of the curves $H_{ij}^{*(k)}$ to the point $H^{(k)}$. It is trivial to prove that each one of the three $H^{(k)}$ points has three decision zone oriented in the same way. The next step is now to find if there exist possible boundaries for curves and lines defined in Equations 6 and 7.

C. Constraints for lines $H_{ij}^{(k)}$ and points $H^{(k)}$

In this section we want to find possible boundaries for the existence domain of the curves $H_{ij}^{(k)}$ and consequently for the points $H^{(k)}$. Let us introduce another assumption, without losing any generality in the model. Let z_i be ordered in a descending way, that is $i < j \Rightarrow z_i > z_j$. Let us start by considering the curve $H_{1;2}^{(1)}(n_1, n_2)$. It is the locus of points satisfing the condition $U_{1;1} = U_{2;1}$ that is $(1 + z_1) f_c(\frac{n_1}{M}) = z_2 f_c(\frac{n_2}{M})$. For $z_2 \neq 0$ and $f_c(\frac{n_1}{M}) \neq 0$ we can write:

$$\frac{1 + z_1}{z_2} = \frac{f_c(\frac{n_2}{M})}{f_c(\frac{n_1}{M})} > 1 \tag{8}$$

Exploiting the condition 4 we can write $f_c(\frac{n_2}{M}) \geq f_c(\frac{n_1}{M})$. Finally, being $f_c()$ strictly monotone decreasing, we obtain the following condition for the existing domain of the curve:

$$\exists H_{1;2}^{(1)}(n_1, n_2) : n_2 \leq n_1 \tag{9}$$

Repeating a similar reasoning for the other two curves of G_1 we obtain a set of three half-planes, whose intersection is the existence domain for the point $H^{(1)}$, which is the open unbounded convex polytope named A_1 in Figure 3. Applying an analogous procedure to the other curves, we get the domain constraints A_2 and A_3 respectively for $H^{(2)}$ and $H^{(3)}$. This leads to a possible relative position of decision boundaries curves $H_{i;j^{(k)}}$ as shown in Figure 4. It is worth to note that the curves don't need to be straight lines, like in the figure but can have a curved shape.

D. Relative position of curves $H_{i;j}^{(k)}$

According to the constraints found until now, there could be other three degrees of freedom for the relative position of the curves $H_{i;j}^{(k)}$, such as the one among $H_{1;2}^{(1)}$ and $H_{1;2}^{(3)}$. We can prove that the relative position depicted in Figure 4 is the only one possible. More in general we want to prove that, fixed an arbitrary value $n_2 = n_2^*$ and leaving only n_1 as a variable, the following inequalities are true:

$$\begin{cases} H_{1;2}^{(2)} < H_{1;2}^{(0)} < H_{1;2}^{(3)} < H_{1;2}^{(1)} \\ H_{2;3}^{(3)} < H_{2;3}^{(0)} < H_{2;3}^{(1)} < H_{2;3}^{(2)} \\ H_{1;3}^{(3)} < H_{1;3}^{(0)} < H_{1;3}^{(2)} < H_{1;3}^{(1)} \end{cases} \tag{10}$$

where $H_{1;2}^{(0)}$ is the straight line $n_2 = n_1$, $H_{2;3}^{(0)}$ the one passing through the points $(M, 0)$ and $(0, M/2)$, and $H_{1;3}^{(0)}$ the one through the points $(M/2, 0)$ and $(0, M)$. To prove that $H_{1;2}^{(3)} < H_{1;2}^{(1)}$ along the direction of n_1 we first prove that there is no

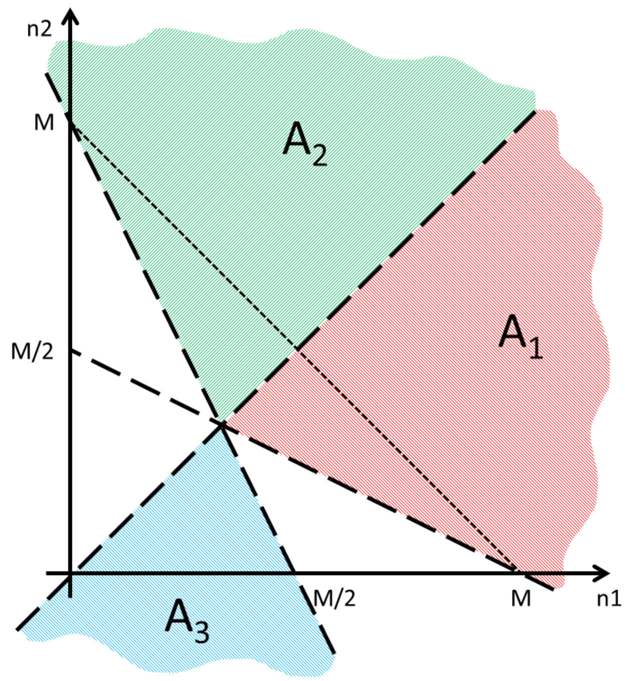

Fig. 3. Constraint boundaries A_k for $H^{(k)}$ points.

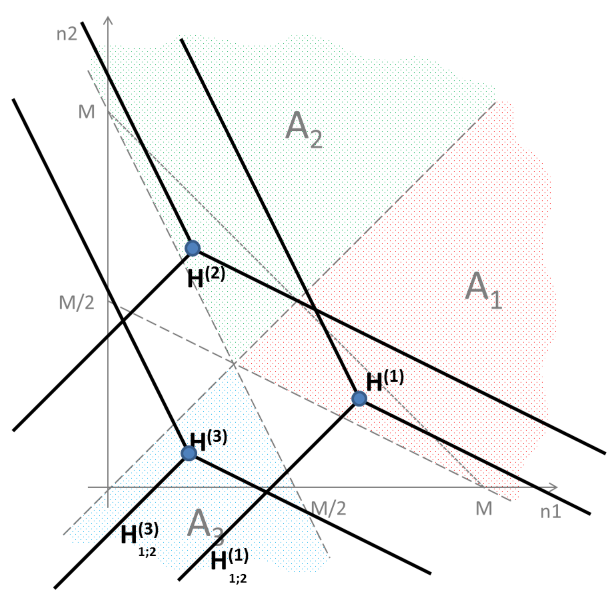

Fig. 4. Decision boundaries lines and $H^{(k)}$ points

intersection among the two lines. Let us suppose by contraddiction that a point of intersection exists, $P^* = (n_1^*, n_2^*)$. So the point must belong to both the lines, that is:

$$\begin{cases} z_1 f_c(\frac{n_1^*}{M}) = z_2 f_c(\frac{n_2^*}{M}) \\ (1+z_1) f_c(\frac{n_1^*}{M}) = z_2 f_c(\frac{n_2^*}{M}) \end{cases} \tag{11}$$

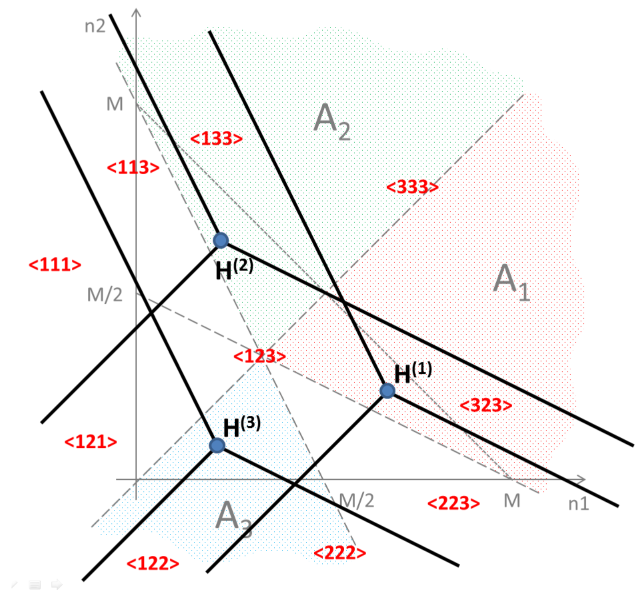

Fig. 5. The only possible zones from which to reach states of the system

Dividing the first equation by the second one we get the contraddiction $0 = 1$. So there is no intersection. To evaluate the relative position along n_1, let us fix a value $n_2 = \bar{n}_2$ for both the curves:

$$\begin{cases} H_{1;2}^{(3)}: & z_1 f_c(\frac{n_1^{(3)}}{M}) = z_2 f_c(\frac{\bar{n}_2}{M}) \\ H_{1;2}^{(1)}: & (1+z_1) f_c(\frac{n_1^{(1)}}{M}) = z_2 f_c(\frac{\bar{n}_2}{M}) \end{cases} \tag{12}$$

The second term in Equation (12) is the same, so also the first terms have to be identical. We have:

$$z_1 f_c(\frac{n_1^{(3)}}{M}) = (1+z_1) f_c(\frac{n_1^{(1)}}{M}) \tag{13}$$

then

$$\frac{f_c(\frac{n_1^{(3)}}{M})}{f_c(\frac{n_1^{(1)}}{M})} = \frac{(1+z_1)}{z_1} > 1 \tag{14}$$

and then

$$f_c(\frac{n_1^{(3)}}{M}) > f_c(\frac{n_1^{(1)}}{M}) \tag{15}$$

and as $f_c()$ is strictly monotonical decreasing we have $n1^{(3)} < n1^{(1)}$ which means that:

$$H_{1;2}^{(3)} < H_{1;2}^{(1)} \tag{16}$$

and this concludes the proof. With analog reasoning the other inequalities can be proved.

E. Pruning the Space of States by reachability

As a consequence of what discussed and proved in the previous sections, the relative position of the $H_{i;j}^{(k)}$ boundaries is only one, and this generates only 10 distinct decision zones, as shown in Figure 5. It means that whatever the initial state of the system is, after the first step, the system will be only in one of these 10 states out of 27. Figure 6 shows the position of the states after pruning. Some considerations can be done about them. Almost all of the states are on the boundary of the zone of the possible states; only one is inside. Moreover there is a sort of central symmetry among the decision zones and the corresponding states. E.g., the state $< 1, 1, 1 >$ $(n_1 = M,\ n_2 = 0,\ n_3 = 0)$ is reached by states bolonging to the corresponding zone that is at the opposide side. On the other hand, the state $< 1, 2, 3 >$ $(n_1 = Mz_1,\ n_2 = Mz_2,\ n_3 = Mz_3)$ is very close or inside its corresponding decision zone, depending on the $f_c()$, its parameters, and the z_i distribution, which determines the position of the boundary lines $H_{i;j}^{(k)}$. It is worth to noting that the state $< 1, 2, 3 >$ is the one in which every node stores the data unit to which it has subscribed.

According to the considerations and results of the previous sections, some important properties can be highlighted.

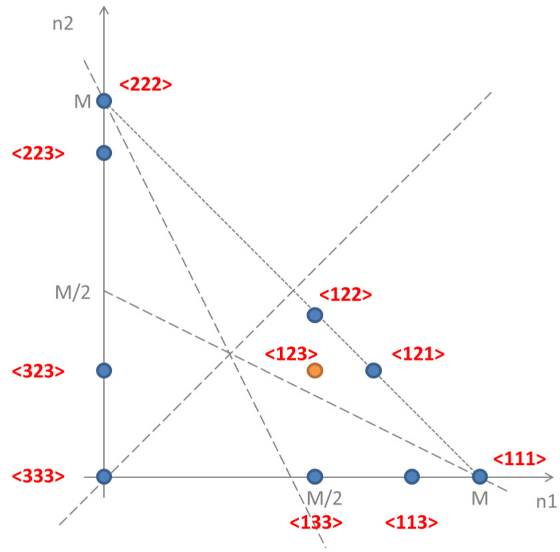

Fig. 6. The only states that are reachable from any other state

Lemma 1. *The states on the boundary of the State Space cannot be absorbing.*

Proof: Let us consider the states on the n_1 axes (Figure 6). They are characterized[3] by having $n_2 = 0$. So the states on axes n_1 can have $n_3 \neq 0$. If one of these states is absorbing, it must belong to a decision zone where, at the next step, no node will chose to store data belonging to group 2, so to have $n_2 = 0$. But that can never happen because the boundaries lines $H^{(2)}_{1;2}$ and $H^{(2)}_{2;3}$ are constrained to stay above all the states on the n_1 axes. In other words, starting from these states, the nodes of group 2 will choose to store data from channel 2. ■

Theorem 1. *There is a threshold condition which defines two cases:*

- *if the condition is met then there is a single absorbing state (Mz_1,Mz_2,Mz_3) and there could be also recurrent classes;*
- *if the condition is not met, there is no absorbing state but only one or more recurrent classes.*

The condition to be met is the following:

$$\begin{cases} \frac{f_c(z_2)}{f_c(z_1)} \leq \frac{1+z_1}{z_2} \\ \frac{f_c(z_3)}{f_c(z_1)} \leq \frac{1+z_1}{z_3} \\ \frac{f_c(z_3)}{f_c(z_2)} \leq \frac{1+z_2}{z_3} \end{cases} \tag{17}$$

Proof: It is trivial to prove that the point $P_z = (n^*_1, n^*_2, n^*_3) = (Mz_1, Mz_2, Mz_3) \in B$, as shown if Figure 7, if $z_1 + z_2 + z_3 = 1$ (z_i is a probability distribution) and $z_1 > z_2 > z_3$ (by assumption). If $P_z \in B$ the only way for P_z to go out from the central polytope (pseudo-hexagon) is through at least one of the three boundary lines: $H^{(1)}_{1;2}$, $H^{(1)}_{1;3}$ and/or $H^{(2)}_{2;3}$. So the conditions to meet are:

$$\begin{cases} H^{(1)}_{1;2}: & (1+z_1)f_c(\frac{n^*_1}{M}) \geq z_2 f_c(\frac{n^*_2}{M}) \\ H^{(1)}_{1;3}: & (1+z_1)f_c(\frac{n^*_1}{M}) \geq z_3 f_c(\frac{n^*_3}{M}) \\ H^{(2)}_{2;3}: & (1+z_2)f_c(\frac{n^*_2}{M}) \geq z_3 f_c(\frac{n^*_3}{M}) \end{cases} \tag{18}$$

Substituting $n^*_i = Mz_i$ we get the Equations (17). This concludes the proof. ■

It is worth noting that all the results obtained until this section are valid for any cost function $f_c()$ and any distribution z_i, provided the assumptions on them in previous sections.

IV. Analysis of the general case of N channels

In the following sections, the general case of N channel is explored. In particular some general proprieties are stated and proved. These proprieties are useful to characterize the behaviour of the system.

[3]It is worth to recall that Figure 6 represents the projection of the states that actually stay on the triangle depicted in Figure 1

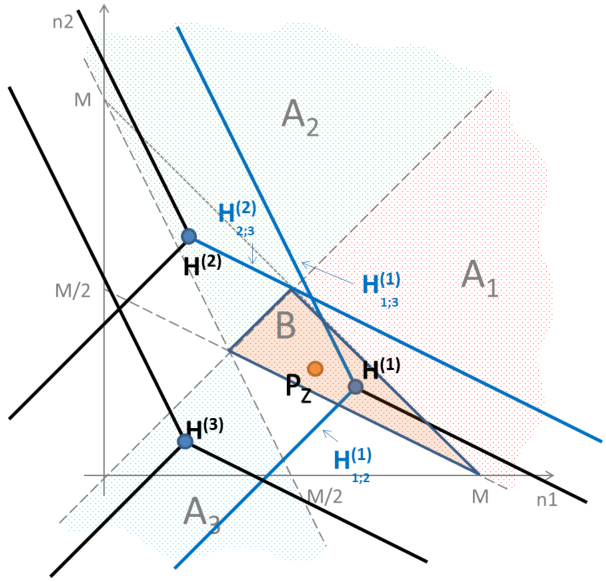

Fig. 7. The system has an absorbing state P_z if it is included in the central polytope (pseudo-hexagon)

A. Assumptions

As described in Section II the cost function $f_c()$ used in the utility function U_{ij} is assumed to belong to a class (Φ_c) with some constraints: the function is strictly positive and monotonically decreasing. Formally:

$$\Phi_c = \left\{ f_c(x) : \forall x \in \mathbb{R} \Rightarrow f_c(x) > 0, \frac{\partial}{\partial x} f_c(x) < 0 \right\} \tag{19}$$

Moreover, the the probabilities z_i for nodes to subscribe to channels i decrease in reverse order:

$$i < j \Rightarrow z_i > z_j \tag{20}$$

with, obviously:

$$\sum_{i=1}^{N} z_i = 1 \tag{21}$$

B. Reachable states

The system state, at any time, is completely described by the vector $\mathbf{c} = (c_1, c_2, \ldots, c_N)$, where N is the number of channels and c_i is the channel chosen by the nodes in group G_i (nodes subscribed to channel i). Without any constrain for the choses channel, the possible number of states of the system would be then N^N. We will show in the following that using the utility function U_{ij}, defined in Equation (2), the actual number of reachable states has an upper bound much lower than N^N, and represents a specific subset of all possible states, with a nice geometric interpretation in the space representation given by $\mathbf{n} = (n_1, n_2, \ldots, n_N)$. Let us start considering what happens if a node in group G_i does not choose to save data from its subscribed channel i.

Lemma 2. *If nodes subscribed to channel i do not save data from channel i, then no other node can save data from channel i.*

Proof: Let us consider the case in which nodes G_i do not save data from channel i, but save data from a different channel j. This happens if and only if:

$$\exists j : z_j f_c(\frac{n_j}{M}) > (1 + z_i) f_c(\frac{n_i}{M}) \tag{22}$$

Let us suppose by contradiction the hypothetical existence of a node within the group G_k that selects channel i, with $k \neq i$. There are two cases: $k \neq j$ and $k = j$. We want to prove that both cases lead to a contradiction.

Let us consider the case $k \neq j$. This would imply:

$$z_i f_c(\frac{n_i}{M}) > z_j f_c(\frac{n_j}{M}) \tag{23}$$

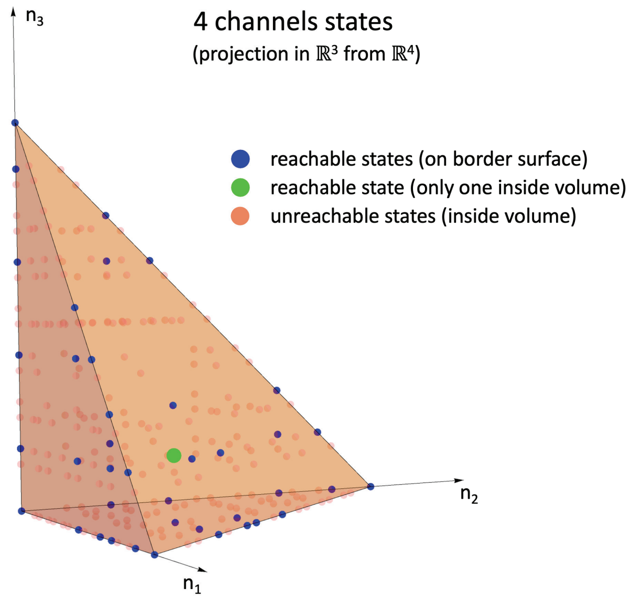

Fig. 8. The states of a 4-channels system: their set is represented as a projection from $\mathbb{R}^4$ (n_1, n_2, n_3, n_4) to $\mathbb{R}^3$ (n_1, n_2, n_3).

Using the Equations (22), the right-hand side of Equation (23) can be bounded below:

$$z_i f_c(\frac{n_i}{M}) > z_j f_c(\frac{n_j}{M}) > (1 + z_i) f_c(\frac{n_i}{M}) \tag{24}$$

This implies:

$$z_i f_c(\frac{n_i}{M}) > (1 + z_i) f_c(\frac{n_i}{M}) \tag{25}$$

Subtracting $z_i f_c(\frac{n_i}{M})$ from both sides we finally get:

$$0 > f_c(\frac{n_i}{M}) \tag{26}$$

But the Equation (26) contradicts the assumption $f_c(x) > 0$ in Equation (19).

The case $k = j$ is more immediate. If we bound below and above the inequality in Equation (22) we get:

$$(1 + z_j) f_c(\frac{n_j}{M}) > z_j f_c(\frac{n_j}{M}) > (1 + z_i) f_c(\frac{n_i}{M}) > f_c(\frac{n_i}{M}) \tag{27}$$

from which we get:

$$(1 + z_j) f_c(\frac{n_j}{M}) > f_c(\frac{n_i}{M}) \tag{28}$$

that is, if $k = j$ the nodes G_k do not save the i channel data. This concludes the proof by contradiction. ■

A consequence of Lemma 2 is that the only possible state in which all channel data are saved in the system at the same time, is the one in which every node saves data to the channel to which it is subcribed to: let's define this state $\hat{\mathbf{c}}^* = (1, \dots, N)$, corresponding to $\hat{\mathbf{n}}^* = M(z_1, \dots, z_N)$. Let us name this state the *"all channels"* state. It is worth nothing that the definition of $\hat{\mathbf{n}}^*$ does not represent an actual stochastic realization of the chose channels by the nodes, for which the element n_i must obsviously be integer numbers. Actually it is an average of all possible choices, according to the $\mathbf{z}$ probability distribution. In other words, a possible realization is $\hat{\mathbf{n}}^{*'} = M(p_1, \dots, p_N) = M\mathbf{p}$ where $\mathbf{z}$ is the average of all possible $\mathbf{p}$.

For all the other reachable states, except $\hat{\mathbf{n}}^*$, at least one channel data is not saved in the system. That is there is at least an index i for which $n_i = 0$ in state $\mathbf{n^b}$. If we represent the set of all possible states using a projection of $\mathbb{R}^N$ on $\mathbb{R}^{N-1}$, such as in Figure 6 or in Figure 8, we have a portion of space corresponding to a convex polytope [4]. In this space, the only reachable states are on the boundary surface (let us name them *"boundary"* states, $\mathbf{n^b}$), while just one, the *"all channels"* state, is inside.

Given the Lemma 2 and its consequences discussed above, it is possible to evaluate the number of states that can be actually reached from ayn other state.

[4]The states' space can be represented in a subspace of dimension $N-1$ because for each state represented by $\mathbf{n} = (n_1, \dots, n_{N-1})$ the value of n_N is obtained by the constraint $\sum_{k=1}^{N} n_k = M$. In other words, the states' space in $\mathbb{R}^N$ belonh to an hyperplane of dimension $N-1$.

Lemma 3. *The number of reachable states can be bounded above by:*

$$R_N = \sum_{k=0}^{N-1} \binom{N}{k} (N-k)^k \tag{29}$$

Proof: In the $\mathbf{c} = (c_1, c_2, \ldots, c_N)$ state representation, let us consider a set a states where there are k, and only k, groups of user who have chosen to save the data from the channel they are subscribed to. Let us call them *dominant* groups. E.g.: for $k = 3$ and $N = 6$, one of these possible groups is $\mathbf{c} = (*, 2, 3, *, *, 6)$, where 2, 3 and 6 are the *dominant* chosen channels. The number of all of these groups can be easily computed as the number of ways to choose k elements from N, without repetition, that is $\binom{N}{k} = \frac{N!}{k!(N-k)!}$, and this is the first factor in the sum in Equation (29). For each one of these sets of states, the remaining $N-k$ groups of users can only chose one of the k *dominant* channels: this is a consequence of Lemma 2. As there are no more *dominant* groups, except the k already mentioned, the other groups must chose one of the k *dominant* channel data, and repetitions are allowed. As a consequence, the number of possibile states for each group is $(N-k)^k$, that is the second factor in the sum in Equation (29). We must then sum on all possible number of *dominant* groups, that is k from 1 to $N-1$. Moreover we have excluded by this sum the *"all channel"* state, where all the groups are *dominant*. Note that we can not extend the sum in Equation (29) to $k = N$ because the second factor does not work correctly with zero remaining groups, and would add a zero contribution. But we note that the formula in the sum gives a contribution of 1 for $k = 0$. So we can extend the sum from $k = 0$ to $k = N-1$ to get the total number of reachable states. This concludes the proof. ∎

To have an idea of what is the order of reduction in the number of reachable states as compared with the theoretical number N^N we note that formula for R_N in Equation (29) does not change the result if we extend the sum for k from 0 to N and that it generates a sequence known in mathematics as *EIS A000248* [5]. An asymptotic estimate can be derived either from the Laplace method or from the saddle-point method expounded in [6]:

$$R_N \simeq \frac{N!}{\sqrt{2\pi N\mu}} \mu^{-N} e^{\frac{N+1}{\mu+1}} \tag{30}$$

where μ is the positive solution of:

$$\mu(\mu+1)e^{\mu} = N+1 \tag{31}$$

Using the Stirling approximation $N! \simeq \sqrt{2\pi N}\frac{N^N}{e^N}$, and $\mu \simeq log(N)$ as an approximated solution of the Equation (31), with some mathematical manipulation a rough bound for large values of N can be found: $R_N < \frac{N^N}{e^N}$.

C. Boundary states features

An important goal of this study is to identify what are the possible absorbing states. In the following we prove that the next step of a *boundary* state can not be itself, so it can not be absorbing.

Theorem 2. *"boundary" states can not be absorbing.*

Proof: Given a *"boundary"* state, $\mathbf{n^b} = (n_1, n_2, \ldots, n_N)$, at least one component must be 0, that is: $\exists i : n_i = 0$. Let us suppose by contradiction that $\mathbf{n^b}$ is absorbing, that is, $\Psi(\mathbf{n^b}) = \mathbf{n^b}$. In other words, if $n_i = 0$, at the next step it must be $n_i = 0$. But it can be verified that:

$$n_i = 0 \Rightarrow U_{ii} > U_{ji}, \forall j \neq i \tag{32}$$

This implies that the node G_i will choose to store data from channel i and it will be $n_i \neq 0$. To prove the statement in Equation (32) consider the following equations:

$$U_{ii} = (1+z_i)f_c(0) \tag{33}$$

$$U_{ji} = z_j f_c(\frac{n_j}{M}) < z_j f_c(0) \leq (1+z_i)f_c(0) = U_{ii} \tag{34}$$

From Equation (34) we get $U_{ji} < U_{ii}$ which proves the Equation (32). This concludes the prof. ∎

D. One possible absorbing state

We can now state what is the only possible absorbing state.

Theorem 3. *The only potential absorbing state is the* "all channel" *state* $\hat{\mathbf{n}}^* = M(z_1, \ldots, z_N)$.

Proof: A consequence of the Lemma 2 and Theorem 2 is that the only reachable states are the *"boundary"* states and the *"all channels"* state. As the *"boundary"* states can not be absorbing, the only candidate is the *"all channels"* state. ∎

Let us verify if this $\hat{\mathbf{n}}^*$ state can be absorbing and under what conditions.

E. All Channels state conditions to be absorbing state

In this section we investigate what are the conditions for the *"all channels"* state to be absorbing, if any. Let us consider the matrix U_{ij}:

$$U_{ij} = \begin{bmatrix} U_{11} & U_{12} & \dots & U_{1M} \\ U_{21} & U_{22} & \dots & U_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ U_{M1} & U_{M2} & \dots & U_{MM} \end{bmatrix} \tag{35}$$

For the specific state $\hat{\mathbf{n}}^* = M\mathbf{z}$ it is:

$$\begin{bmatrix} (1+z_1)f_c(z_1) & z_1 f_c(z_1) & \dots & z_1 f_c(z_1) \\ z_2 f_c(z_2) & (1+z_2)f_c(z_2) & \dots & z_2 f_c(z_2) \\ \vdots & \vdots & \ddots & \vdots \\ z_M f_c(z_M) & z_M f_c(z_M) & \dots & (1+z_M)f_c(z_M) \end{bmatrix} \tag{36}$$

In order for the state $M\mathbf{z}$ to be absorbing, each diagonal element in matrix in Equation (36) must be greater than the other elements of the same column. This would require $M(M-1)$ comparisons to verify if it is absorbing. Actually the number of comparisons can be reduced, as proved in the following Lemma and Theorem.

Lemma 4. *In* "all channels" *state* $\hat{\mathbf{n}}^*$*,* $U_{hh} > Ulh$*, for* $h > l$*.*

Proof: Given the assumptions in Equations (19), (20) and (21), we want to prove that in $\hat{\mathbf{n}}^* = M\mathbf{z}$ it holds $U_{hh} > U_{lh}$, for $h > l$, that is $(1+z_h)f_c(z_h) > z_l f_c(z_l)$ is always true. In other words, we want to prove that the diagonal elements in Equation (36) are always greater than the elements above. Let us consider a point $\mathbf{z} = (z_1, z_2, \dots, z_N)$ in an N-dimensional space Ψ ($\mathbb{R}^N$), subject to the constraints in Equations (20) and (21). In particular $\mathbf{z}$ belongs to an hyperplane of dimension $N-1$, that divides Ψ into 2 semi-spaces. Le us define the function $\hat{U}_{jk}(\mathbf{x}) \triangleq (1_{jk} + z_j)f_c(x_j)$, where x_j is the *j-th* component of the variable $\mathbf{x}$ while z_j is a constant given by the *j-th* component of $\mathbf{z}$. Let us then consider the locus of points $\mathbf{x}^* \in \Psi$ that satisfy the equation $U_{hh}(\mathbf{x}_*) = U_{lh}(\mathbf{x}_*)$, that is:

$$(1+z_h)f_c(x_h^*) = z_l f_c(x_l^*) \tag{37}$$

Let us name H_h^{lh} this locus of points. It is worth nothing that we do not impose any condition on the other x_i^* component of $\mathbf{x}^*$ ($i \neq l, h$). From the Equation (37) we get:

$$\frac{f_c(x_h^*)}{f_c(x_l^*)} = \frac{z_l}{(1+z_h)} < 1 \tag{38}$$

This implies $f_c(x_h^*) < f_c(x_l^*)$ and being $f_c^{'}() < 0$ we get:

$$x_h^* > x_l^* \tag{39}$$

that means that the locus of points $\mathbf{x}^* \in \Psi$ satisfying the condition $U_{hh}(\mathbf{x}_*) = U_{lh}(\mathbf{x}_*)$ are subject to the constraint in Equation (39). As for $\mathbf{z}$ it must be $z_h < z_l$, then we get:

$$\mathbf{z} \notin H_h^{lh} \tag{40}$$

It means that z belongs to one of the 2 semi-spaces in which H_h^{lh} divides Ψ. Let us now consider a point $\bar{\mathbf{x}^*} \in H_h^{lh}$ with the following constraints on its coordinates:

$$\begin{cases} \bar{x}_h^* = z_h & \\ \bar{x}_l^* : & (1+z_h)f_c(z_h) = z_l f_c(x_l^*) \\ x_i^* & \text{any for } i \neq h, l \end{cases} \tag{41}$$

From the Equations (20), (39) and (41) we can write:

$$\bar{x_l^*} < \bar{x_h^*} = z_h < z_l \tag{42}$$

then $x_l^* < z_l$ and, being $f_c()$ monotonically decreasing, $f_c(x_l^*) > f_c(z_l)$. We can then write $(1+z_h)f_c(z_h) = z_l f_c(x_l^*) > z_l f_c(z_l)$, from which:

$$(1+z_h)f_c(z_h) > z_l f_c(z_l) \tag{43}$$

This concludes the proof. ■

We can finally get the condition for the *"all channels"* state to be absorbing.

Theorem 4. *The "all channels" state $M\mathbf{z}$ is absorbing if the following condition is met:*

$$\bigcap_{1<i<j<N} \left\{ \frac{f(z_j)}{f(z_i)} < \frac{1+z_i}{z_j} \right\} \tag{44}$$

Proof: If $M\mathbf{z}$ state is absorbing, every node G_i must continue to choose to buffer the channel data it is subscribed to. That is, the following conditions must be met:

$$U_{ii} > U_{ji} \quad \forall i, j \,:\, i \neq j \tag{45}$$

According to Theorem 4, some of these conditions are always met for $M\mathbf{z}$ state. In particular $U_{ii} > U_{ji} \, \forall i, j \,:\, i > j$. So the only conditions to verify are $U_{ii} > U_{ji} \forall i, j \,:\, i < j$. That is $(1 + z_i)f_c(z_i) > z_j f_c(z_j) \forall i, j \,:\, 1 < i < j < N$. From which we get the Equation (44). ■

Summarising, with the previous theorems, we have proved, for an N-channels system, what are the reachable states and under what conditions absorbing states can exists. We must now investigate what are the possible recurrent classes of the system. In the following section some preliminary study about it is reported.

F. Behaviour of "single channel" states

Let us consider the class of states for which all the subscribers (to different channels) have chosen to save data from the same channel. There are N (number of channels) of these possible states which we name $\hat{\mathbf{n}}^{\mathbf{k}} = (n_{1,k}, \ldots, n_{N,k})$ where

$$n_{i,k} = \begin{cases} M & for \quad i = k \\ 0 & for \quad i \neq k \end{cases} \tag{46}$$

with $k = 1, \ldots, N$. It is worth to note that these states are on the vertexes of the convex polytope representing the domain of the feasable states. In the equivalent representation, let's name these states $\hat{\mathbf{c}}^{\mathbf{k}} = (k, \ldots, k)$.

For example, in the case $N = 3$, the "single channel" states are $\mathbf{n} = (M, 0, 0)$, $(0, M, 0)$, and $(0, 0, M)$, or, in the equivalent representation, $\mathbf{c} = (1, 1, 1)$, $(2, 2, 2)$, and $(3, 3, 3)$ (see Figure 6).

Let's define two more states' classes. One is the single state class $\hat{\mathbf{c}}^* = (1, \ldots, N)$ in which every node saves data from the channel it is subscribed to. The other one is composed by the set of states $\hat{\mathbf{c}}^{*,\mathbf{h}} = (1, \ldots, h-1, c_h, h+1, \ldots, N)$ where

$$c_h = \begin{cases} 2 & for \quad h = 1 \\ 1 & for \quad h > 1 \end{cases} \tag{47}$$

with $h = 1, \ldots, N$. In other words, $\hat{\mathbf{c}}^{*,\mathbf{h}}$ is the set of states in which any node saves data from the channel it is subscribed to, except the nodes subscribed to channel h that saves data from the channel with the lowest index different from h. E.g., for 3 channels, $\hat{\mathbf{c}}^* = (1, 2, 3)$, $\hat{\mathbf{c}}^{*,\mathbf{1}} = (2, 2, 3)$, $\hat{\mathbf{c}}^{*,\mathbf{2}} = (1, 1, 3)$, $\hat{\mathbf{c}}^{*,\mathbf{3}} = (1, 2, 1)$.

In the following we will use equivalently both the notations $\mathbf{n}$ and $\mathbf{c}$ for the states of the system.

Lemma 5. *A "single channel" state $\hat{\mathbf{c}}^{\mathbf{k}}$ can have only two possible next states: i) $\Psi(\hat{\mathbf{c}}^{\mathbf{k}}) = \hat{\mathbf{c}}^*$ or ii) $\Psi(\hat{\mathbf{c}}^{\mathbf{k}}) = \hat{\mathbf{c}}^{*,\mathbf{k}}$.*

Proof: If $\mathbf{c} = \hat{\mathbf{c}}^{\mathbf{k}}$ then for any node G_j subscribed to a channel $j \neq k$ there are three different cases for the utility function: $U_{jj}(\hat{\mathbf{n}}^{\mathbf{k}}) = (1 + z_j)f_c(0)$, $U_{hj}(\hat{\mathbf{n}}^{\mathbf{k}}) = z_h f_c(0)$ (for $h \neq k$) and $U_{kj}(\hat{\mathbf{n}}^{\mathbf{k}}) = z_k f_c(1)$. Using the assumptions made on $f_c()$, it holds always true that $U_{jj}(\hat{\mathbf{n}}^{\mathbf{k}}) > U_{hj}(\hat{\mathbf{n}}^{\mathbf{k}})$ and $U_{jj}(\hat{\mathbf{n}}^{\mathbf{k}}) > U_{kj}(\hat{\mathbf{n}}^{\mathbf{k}})$. As a consequence, any node G_j will save data from channel j. On the other hand, for any node subscribed to the channel k we have: $U_{kk}(\hat{\mathbf{n}}^{\mathbf{k}}) = (1 + z_k)f_c(1)$ and, for $h \neq k$, $U_{hk}(\hat{\mathbf{n}}^{\mathbf{k}}) = z_h f_c(0)$. Let us consider the greatest value among $U_{hk}(\hat{\mathbf{n}}^{\mathbf{k}})$ that is always $\hat{U} = U_{\hat{h}k}(\hat{\mathbf{n}}^{\mathbf{k}})$ with $\hat{h} = \min_h \{h\}$, according to the assumptions on z_h. There are two possible cases: if $k = 1$ then $\hat{U} = U_{2,1}(\hat{\mathbf{n}}^{\mathbf{1}}) = z_2 f_c(0)$ corresponding to $\hat{h} = 2$, otherwise if $k > 1$ then $\hat{U} = U_{1k}(\hat{\mathbf{n}}^{\mathbf{k}}) = z_1 f_c(0)$ and $\hat{h} = 1$. This implies that nodes subscribed to channel k will choose the channel, to save data from, according to the $\max\{U_{kk}, \hat{U}\}$, resulting then in choosing the channel k or one between the channels 1 (if $k > 1$) and 2 (if $k = 1$). In the first case the next state will be $\hat{\mathbf{c}}^*$, in the second one it will be $\hat{\mathbf{c}}^{*,\mathbf{k}}$. ■

To better understand the dependance of the system from the cost function, let us restrict $f_c()$ to a class of functions, named H_λ, introducing a parameter λ in $f_c(\frac{n_i}{M}) = f_c(\frac{n_i}{M}, \lambda)$. To specify the class we define a support function

$$h(\lambda) = \frac{f_c(0, \lambda)}{f_c(1, \lambda)} \tag{48}$$

We impose that $f_c()$ must be such that $h(\lambda)$ is monotonically increasing, that is $\frac{\partial}{\partial \lambda} h(\lambda) > 0$. In other words, increasing λ the cost of a fully diffused channel will decrease if compared with the cost of a channel not spread at all. For the sake of clarity $H_\lambda \subset \Psi_c$, that is $f_c()$ is subject to the assumptions described in Section II-B. An example of $f_c() \in H_\lambda$ is $e^{-\lambda \frac{n_i}{M}}$.

Lemma 6. $\forall f_c() \in \Phi_c$, $\exists \, \bar{k} \in \mathbb{N}$ *such that*

$$\Psi(\hat{\mathbf{c}}^{\mathbf{k}}) = \begin{cases} \hat{\mathbf{c}}^* & for \quad k \leq \bar{k} \\ \hat{\mathbf{c}}^{*,\mathbf{k}} & for \quad k > \bar{k} \end{cases} \tag{49}$$

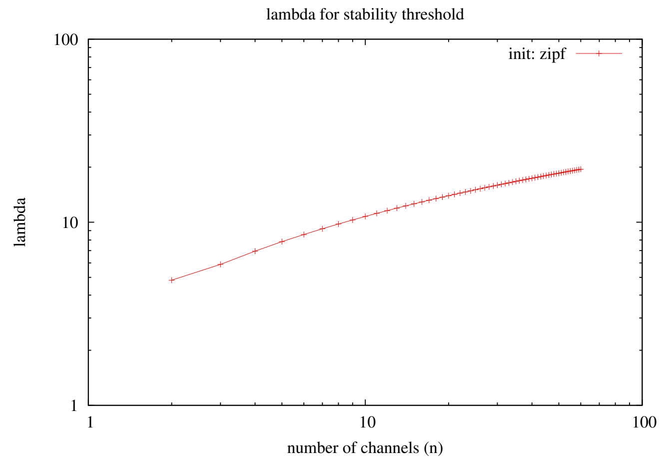

Fig. 9. Simulation results for exponential cost function (the graph is the λ boundary for an absorbing state)

with $k \in [1, \dots, N]$.

Proof: As seen in Lemma 5, if the system is in the state $\hat{\mathbf{c}}^{\mathbf{k}}$, there are only two possible next states, $\hat{\mathbf{c}}^*$ and $\hat{\mathbf{c}}^{*,\mathbf{k}}$, where the difference between them depends only on the choice of the G_k users (subscribed to channel k): they can decide to save data from channel k or from a specific different channel, according to the $\hat{\mathbf{c}}^{*,\mathbf{k}}$ definition. For $k = 1$ the choice of G_1 users is $\hat{\imath} = \arg\max_i \{U_{i,1}\}$ which can be $\hat{\imath} = 1$ (which is k, in this case) if $\frac{f_c(0)}{f_c(1)} < \frac{1+z_1}{z_2} \triangleq \alpha_1$, or $\hat{\imath} = 2$, otherwise. For all the other G_k users, with $k > 1$, the choice $\hat{\imath} = \arg\max_i \{U_{i,k}\}$ is $\hat{\imath} = k$ if $\frac{f_c(0)}{f_c(1)} < \frac{1+z_k}{z_1} \triangleq \alpha_k$, and $\hat{\imath} = 1$, otherwise. The thresholds α_k are monotonically decreasing with k, so for any fixed value $\frac{f_c(0)}{f_c(1)} \triangleq \bar{h}$ there is a $\bar{k}$ for which the following inequations hold true: $\alpha_N < \alpha_{N-1} < \dots < \alpha_{\bar{k}+1} < \bar{h} < \alpha_{\bar{k}} < \dots < \alpha_2 < \alpha_1$. In other words there is a lower contiguous set of integers $\left[1, \dots, \bar{k}\right] \triangleq S^-_{\bar{k}}$ such that, for $k \in S^-_{\bar{k}}$, $\bar{h} < \alpha_k$ holds true, which implies $\Psi(\hat{\mathbf{c}}^{\mathbf{k}}) = \hat{\mathbf{c}}^*$, and there is an upper contiguous set of integers $\left[\bar{k}+1, \dots, N\right] \triangleq S^+_{\bar{k}}$ such that, for $k \in S^+_{\bar{k}}$, $\alpha_k < \bar{h}$ holds true, which implies $\Psi(\hat{\mathbf{c}}^{\mathbf{k}}) = \hat{\mathbf{c}}^{*,\mathbf{k}}$. It is worth to noting that one of the two sets could be empty. This concludes the proof. ■

Theorem 5. *If $f_c(\frac{n_i}{M}, \lambda) \in H_\lambda$, and if, varying the value of λ, there is a change in any $\Psi(\hat{\mathbf{c}}^{\mathbf{k}})$, then it will happen in such a way that increasing λ the change will affect a sequence of states $\hat{\mathbf{c}}^{\mathbf{k}}$, with consecutive decreasing k. Increasing λ, the change in affected states' transitions will be from $\hat{\mathbf{c}}^*$ to $\hat{\mathbf{c}}^{*,\mathbf{k}}$.*

Proof: By definition of H_λ, $h(\lambda) = \frac{f_c(0,\lambda)}{f_c(1,\lambda)}$ is monotonically decreasing with λ. ■

This completely describes the behaviour of *"single channel"* states, that is their next step.

V. Validation by simulation

In order to verify the results obtained so far we performed a simulation for a specific case of the cost function $f_c()$ and preference distribution. As the present work aims to generalize to N channels the 2 channels study in [1], we used the same assumptions for the cost function and the distribution of nodes' interests by channel subscriptions. In particular we used an exponential cost function, and a zipf interest distributions. Then for each value of n (number of subscribing channels) ranging from 2 to 60, for the cost function $f_c(n_i) = e^{-\lambda \frac{n_i}{M}}$ we run simulations in which the initial subscription of node (z_i) coincides with their preference to store (p_i). Then we found the threshold value for λ for which the *"all channels"* state is absorbing. All the values match the theoretical results described in Theorem 4.

Figure 9 shows the results. For each n (number of channels), for any λ value below the value of the graph the system has an absorbing state.

Just to consider a specific case, for exponential cost function $f_c(n_i) = e^{-\lambda \frac{n_i}{M}}$ the conditions of Theorem 1, that can be obtained also from the Theorem 4, reduces to two (one of them is redundant) and we get:

$$\begin{cases} \lambda \leq \frac{ln(\frac{1+z_1}{z_2})}{z_1 - z_2} \\ \lambda \leq \frac{ln(\frac{1+z_1}{1-(z_1+z_2)})}{2z_1 + z_2 - 1} \end{cases} \tag{50}$$

If z_i has a zipf distribution, then the first condition is sufficient, and we get:

$$\lambda \leq 5.88518 \tag{51}$$

The value matches the simulation results.

VI. Conclusion and Future Work

In this work we have described part of the extension of the analytical model presented in [1]. The previous model completely describes the behaviour of a 2 channel data dissemination system in opportunistic networks. In the present work we have initially analysed the system in the case of 3 channels, to familiarise with the problem with the help of a visual representation of the state space, not possible for a higher number of channels. Then the general case of N channels is studied providing a general model to identify the conditions for the stability of the system with any cost function and any channel interest distribution. In particular we have found what are the reachable states of the system, that are a subset of all theoretical states. We have identified two classes of states: *"boundary* states" and one *"all channels"* state. Then we have proved that the *"all channels"* state is the only possible absorbing state and have found the condition to be satisfied by the parameters of the system for the absorbing state to be possible. Finally we have investigated some partial behaviours in order to identify the recurrent classes of the system. This last topic is object of future developments.

References

[1] Boldrini, C. and Conti, M. and Passarella, A., "Modelling data dissemination in opportunistic networks," in *Proceedings of the third ACM workshop on Challenged networks*. ACM, 2008, pp. 89–96.

[2] V. Lenders, G. Karlsson, and M. May, "Wireless Ad hoc Podcasting," in *Proc. of IEEE SECON*, 2007.

[3] C. Boldrini, M. Conti, and A. Passarella, "Context and resource awareness in opportunistic network data dissemination," in *Proc. of IEEE AOC*, 2008.

[4] B. Grünbaum, *Convex Polytopes Second Edition prepared by Volker Kaibel, Victor Klee, and Günter Ziegler Graduate Texts in Mathematics vol 221 POD*. Springer, 2003.

[5] "The on-line Encyclopedia of integer sequences EIS A000248," https://oeis.org/A000248.

[6] R. S. Philippe Flajolet, *Analytic Combinatorics, VIII.13 pages 131 and 571*. Cambridge University Press, 2009.