Evaluation of Advanced Routing Strategies with Information-Theoretic Complexity Measures

Michele Amoretti[(✉)] and Stefano Cagnoni

University of Parma, Parma, Italy
{michele.amoretti,stefano.cagnoni}@unipr.it

Abstract. Based on hierarchy and recursion (shortly, HR), recursive networking has evolved to become a possible architecture for the future Internet. In this paper, we advance the study of HR-based routing by means of the Gershenson-Fernandez information-theoretic framework, which provides four different complexity measures. Then, we introduce a novel and general approach for computing the information associated to a known or estimated routing table. Finally, we present simulation results regarding networks that are characterized by different topologies and routing strategies. In particular, we discuss some interesting facts we observed while comparing HR-based to traditional routing in terms of complexity measures.

Keywords: Distributed systems · Recursive networking · Complexity measures

1 Introduction

Recursive networking refers to multi-layer virtual networks embedding networks as nodes inside other networks. It is based on *hierarchy*, *i.e.*, the categorization of a set of nodes according to their capability or status, and *recursion*, which is the repeated use of a single functional unit over different scopes of a distributed system. In the last decade, recursive networking has evolved to become a possible architecture for the future Internet [2]. In particular, it is a prominent approach to designing quantum networks [3]. In a recent work [1], we proposed to apply hierarchy and recursion (HR) to build self-aware and self-expressive distributed systems. In particular, we presented HR-based network exploration and routing algorithms.

In this paper, we continue the characterization of HR-based routing by means of a simple albeit powerful and general information-theoretic framework providing complexity measures, recently proposed by Gershenson and Fernandez [4]. Firstly, we introduce a novel and general (*i.e.*, not HR-specific) approach for computing the information associated to a known or estimated routing table. Then we present simulation results regarding networks that are characterized by different topologies and routing strategies. In particular, we discuss some interesting facts we observed, while comparing HR-based to traditional routing in terms of complexity measures.

The paper is organized as follows. In Sect. 2, we summarize the basic concepts of Gershenson and Fernandez's information-theoretic framework [4]. In Sect. 3, we illustrate our approach for computing the information associated to a routing table. In Sect. 4, we recall the working principles of HR-based routing. In Sect. 5, we present simulation results. Finally, in Sect. 6, we outline future research directions.

2   Complexity and Information

It is difficult to provide an exhaustive list of the ways of defining and measuring system complexity that have been proposed by the research community. Among others, the Gershenson-Fernandez information-theoretic framework provides abstract and concise measures of emergence, self-organization, complexity and homeostasis [4]. According to their framework, emergence is the opposite of self-organization, while complexity represents their balance. Homeostasis can be seen as a measure of the stability of the system.

In detail, a system can be described by a string X, composed by a sequence of variables with values $x \in \{1, .., n\}$ which follow a probability distribution $P(x)$. The information associated to that system is the normalized entropy

$$I = -\frac{\sum_x P(x) \log P(x)}{I_{max}} \tag{1}$$

where $I \in [0, 1]$ and $I_{max} = -\log(1/n)$, since the maximum information value is achieved when all values $1, .., n$ have the same probability.

Considering the dynamics of the system as a process, emergence can be defined as the novel information generated by that process:

$$E = \frac{I}{I_{init}} \tag{2}$$

where I and I_{init} are the current and initial information associated to the system, respectively. The initial information can be referred to the initial state or condition of the system. If the initial state is random, then $I_{init} = 1$.

Self-organization is seen as the opposite of emergence, since high organization (order) is characterized by low information. Vice versa, low organization is characterized by high information. Thus

$$S = I_{init} - I \tag{3}$$

Thus, self-organization occurs ($S > 0$) if the dynamics of the system reduce information.

Since E represents how much variety there is in a system, and S represents how much order, complexity is defined as their product:

$$C = a \cdot E \cdot S \tag{4}$$

where a is a normalization factor, due to the fact that E may be $> 1/S$.

Last but not least, homeostasis is defined as

$$H = 1 - d \tag{5}$$

where d is the normalized *Hamming distance* between the current and initial state of the system, measuring how much change has taken place. Being defined as its complement, homeostasis is a measure of the stability of the system. A high H implies that there is no change, that is, information is maintained.

This framework has been used to study different kinds of complex systems, ranging from self-organizing traffic lights [5] to adaptive peer-to-peer systems [6].

3   Information Associated to a Routing Table

Traditionally, routing strategies are compared in terms of effectiveness, efficiency and scalability [7,8]. To this purpose, selected independent variables should explain performance under a wide range of scenarios [9]. In particular, estimating routing tables is an important and challenging task, as details of how a route is chosen are diverse, and generally not publicly disclosed. An interesting strategy has been recently proposed by Rotenberg *et al.* [10].

In this context, we propose a novel and general approach for characterizing the whole network, namely, by averaging the emergence, self-organization, complexity and homeostasis values of its routers.

From now on, for simplicity, we assume that every node of the network is provided with a routing table, allowing to forward packets to neighbor nodes (routes), according to their destinations. A routing table can be modeled as a set of (destination, route) pairs.

Consider a node with k neighbors. Then, its routing table takes into account k possible routes. In terms of the framework illustrated in Sect. 2, this means that $x \in \{1, .., k\}$. By inspecting the routing table, it is possible to determine the relative frequency of each route. Thus, we define

$$P(x) = \frac{n_x}{n} \tag{6}$$

where n is the size of the routing table and n_x is the number of destinations whose route is x.

When a new node joins the network, its routing table is empty and every route has the same probability. Thus, $I_{init} = 1$. As a consequence, Eqs. 2–5 become:

- $E = I/I_{init} = I$
- $S = I_{init} - I = 1 - I$
- $C = aES = 4I(1 - I)$
- $H = 1 - d$

where $a = 4$ comes from: $\max\{ES\} = 0.5(1 - 0.5) = 1/4$; d is the normalized Hamming distance between the initial and current configurations of the routing

table. In general, the Hamming distance between any two consecutive configurations of the routing table is computed per-node according to the following equations:

$$\forall \text{ neighbor } i :$$
$$D_{i+1} = D_i + f(r)$$
$$f(i) = \begin{cases} 1 & \text{if } r \text{ is route in new routing table only} \\ 1 & \text{if } r \text{ is route in old routing table only} \\ 1 & \text{if } r \text{ is route in both routing tables, but } n_{old} \neq n_{new} \\ 0 & \text{else} \end{cases}$$

where n is the number of destinations associated to the selected route. Once normalized, D_i becomes d_i.

4   HR-Based Routing

We recall and explain HR-based routing by means of an example. Let us consider the network shown in Fig. 1. The routing table at node 4.2 contains information on how to reach any other node in the network. The table has more precise information about nearby destinations (node 4.4 and node 4.7), and vague information about more remote destinations (NET9).

Suppose that node 4.2 has to send a message to node 9.6. If routing tables were filled only with local information (*i.e.*, node 4.2's direct neighbors), routing would be quite inefficient. Instead, hierarchy and recursion make it possible to find the route more quickly. Node 4.2 knows that NET9 is reachable through NET6, whose node 6.1 is directly reachable. Thus, node 4.2 sends the message to node 6.1. The complete HR-based routing algorithm is described by the flowchart in Fig. 2.

HR-based routing is suitable for both intra-domain and inter-domain scenarios. Compared to the two main classes of intra-domain routing, namely Link-State and Distance-Vector [7], HR-based routing has the following advantages:

1. nodes are not required to know the whole network topology (unlike Link-State routing);
2. nodes build collective awareness by exchanging recursive and hierarchical information not only with direct neighbors, but also with neighbors of neighbors, etc. (unlike Distance-Vector routing).

For further details about HR-based versus Link-State and Distance-Vector, the reader may refer to our previous work [1]. Thanks to collective awareness, messages can be routed within the same subnetwork or from one subnetwork to another; doing so they enable, for example, the Unified Architecture for inter-domain routing proposed in RFC 1322.[1]
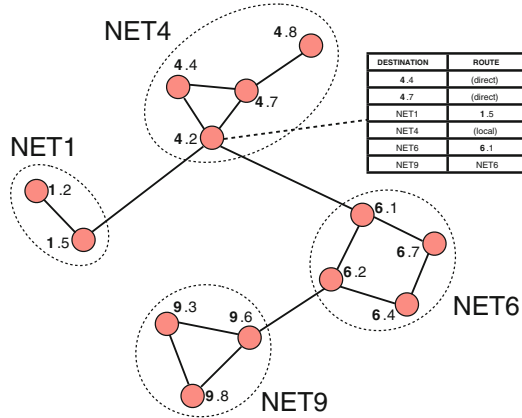
---

[1] http://www.rfc-editor.org/rfc/rfc1322.txt.

Fig. 1. Hierarchy and recursion: the routing table at node 4.2 contains information on how to reach any other node in the network.
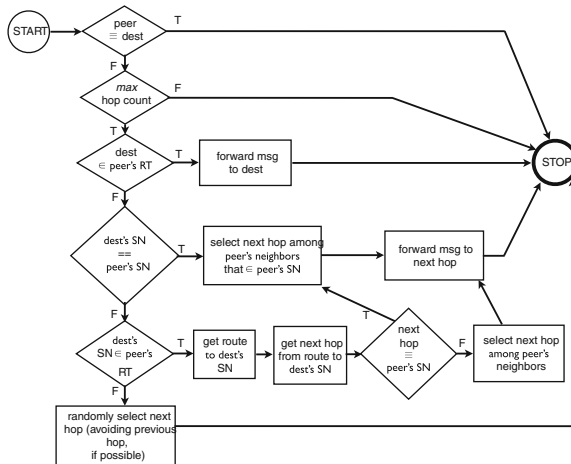

Fig. 2. HR-based routing algorithm. RT stands for routing table; SN for subnetwork.

5   Simulation Results

To evaluate the proposed approach, we used the general-purpose discrete event simulation environment DEUS [11]. The purpose of DEUS is to facilitate the simulation of highly dynamic overlay networks with several hundred thousands nodes, without needing to simulate also lower network layers.

Without loss of generality, we considered the (sub-optimal) scenario in which every node knows which subnetworks can be reached through its direct neighbors. In HR-based routing, no further knowledge — provided by neighbors of neighbors (of neighbors etc.) — is necessary, when the number of subnetworks M is of the

same order of magnitude as the mean node degree $\langle k \rangle$ of the network. Instead, for large networks, with $M \gg \langle k \rangle$, further knowledge is necessary to build effective routing tables.

We took into account two network topologies, characterized by different statistics for the *node degree*, which is the number of links starting from a node. The first network topology we considered is *scale-free*, meaning that its PMF decays according to a power law $P(k) = ck^{-\tau}$, with $\tau > 1$ (to be normalizable) and c normalization factor. Such a distribution exhibits the property of scale invariance (*i.e.*, $P(bk) = b^a P(k)$, $\forall a, b \in \mathbb{R}$). The second network topology we considered is a purely-random one, described by the well-known model defined by Erdös and Rényi (ER model). Networks based on the ER model have N vertices, each connected to an average of $\langle k \rangle = \alpha$ nodes. Scale-free and purely-random are the extremes of the range of meaningful network topologies, as they represent the presence of strong hubs and the total lack of hubs, respectively.

We evaluated the HR-based routing strategy in terms of success rate (*i.e.*, fraction of messages arrived to destination) and average route length, using different networks characterized by $N = 1000$ nodes, with $M = 20$ subnetworks. With the BA topology, when $m = 5$ and $m = 20$, the mean node degree is $\langle k \rangle = 10$ and $\langle k \rangle = 40$, respectively. To have the same $\langle k \rangle$ values for the ER topology, we set $\alpha = 10$ and $\alpha = 40$. Reported results are average values coming from 25 simulation runs.

As a basis for comparison, we also simulated a routing strategy where the nodes do not populate routing tables with information about subnetworks. Instead, they only keep trace of direct neighbors and neighbors of neighbors. Such a strategy (denoted as No-HR) has some common properties with Distance-Vector routing, although it does not manipulate vectors of distances to other nodes in the network. Mean values and standard deviations of success rate r_s and average route length n_h, reported in Table 1, show that the HR-based strategy outperforms the other one, provided that the average node degree $\langle k \rangle$ is suitably high. Interestingly, with low $\langle k \rangle$ values, the HR-based routing strategy has worse performance when the topology is ER. However, a small increase of $\langle k \rangle$ corresponds to a high performance increase of the HR-based routing strategy.

Then, we computed E, S, C and H at each node, from the initial configuration corresponding to $I_{init} = 1$, to the steady-state configuration corresponding to the filled routing table. We averaged the resulting values, considering the whole network. Their evolution is illustrated in Fig. 3.

Four main facts can be observed:

1. As m and α grow, E tends to 1, S tends to 0.
2. When m and α are low, HR-based and NoHR routing exhibit very different H values.
3. When m and α are high, the values of H in HR-based and NoHR routing are more similar.
4. Even if the mean node degree $\langle k \rangle$ is the same, BA and ER topologies result in very different E, S and C values.

Table 1. HR vs NoHR: success rate r_s and average route length n_h

Strategy	Topology	S	μ_{r_s}	σ_{r_s}	μ_{n_h}	σ_{n_h}
HR	BA, $m = 5$	20	0.88	2E-2	17.6	2.06
NoHR	BA, $m = 5$	20	0.74	2.9E-1	19.7	9.8
HR	BA, $m = 20$	20	0.99	9.3E-4	3.8	8E-2
NoHR	BA, $m = 20$	20	0.99	9E-3	9.85	1.2
HR	ER, $\alpha = 10$	20	0.64	3E-2	43.7	2.72
NoHR	ER, $\alpha = 10$	20	0.55	3.3E-1	21.64	17.91
HR	ER, $\alpha = 40$	20	0.99	1E-3	4.0	1.2E-1
NoHR	ER, $\alpha = 40$	20	0.93	1.9E-1	15.33	4.75
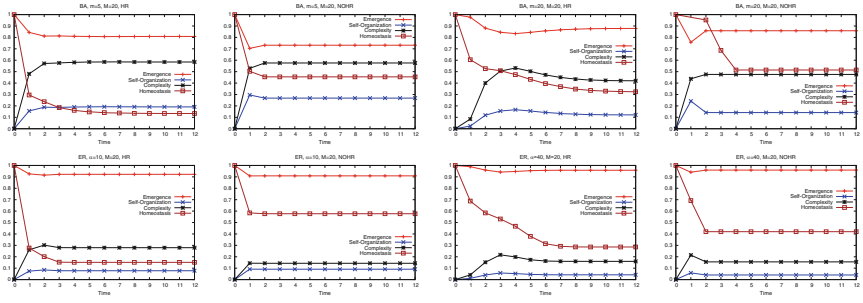
Fig. 3. Complexity measures of HR-based and NoHR routing with different topologies.

The reason for the first fact is that a higher number of connections, due to higher m and α, makes the routing table more varied in terms of available routes. The probability distribution $P(x)$ has fewer spikes, thus I is higher. As a consequence, E increases and S decreases. The second fact can be stated more precisely by means of the following inequality: $H_{HR} \ll H_{NoHR}$, when m and α are small. Our interpretation is that a reduced number of connections enhances the differences between routing tables, in HR-based and NoHR routing, $i.e.$, with respect to the initial state, the final state of the routing table is much more different in HR-based routing rather than NoHR routing. The impact on performance is evident: HR routing table are better than NoHR ones, thus producing a higher success rate. It is not possible, however, to generalize associating higher H values to higher performance. Conversely, a higher number of connections reduces the differences between routing tables, explaining the third fact. The fourth fact is further detailed by the following inequalities: $E_{BA} < E_{ER}$, $S_{BA} > S_{ER}$ and $C_{BA} \gg C_{ER}$, when m and α are such that the mean node degree $\langle k \rangle$ is the same in the BA and ER topologies. It is difficult to explain the relationship between C and performance, in general. It makes more sense to consider E and S separately. Regarding E, our interpretation is that scale-free properties

(characterizing the BA topology) make some routes intrinsically more probable than others. Indeed, only a few nodes have a high number of connections (such nodes are denoted as *hubs*). Thus, with respect to the ER topology, in scale-free networks the probability distribution $P(x)$ has more spikes, making I smaller. Consequently, E is lower and S is higher. Indeed, the presence of hubs makes routing more robust (S is higher), thus improving performance.

6   Conclusion

In this paper we have illustrated a novel approach to quantifying the information associated to a known or estimated routing table, allowing to characterize the whole network by averaging the emergence, self-organization, complexity and homeostasis values of its nodes. Our simulation study shows that these measures may represent an important complement to traditional performance indicators for routing protocols.

Regarding future work, we plan to improve the information-theoretical investigation of HR-based routing strategies, considering larger networks with multi-layered trees of subnetworks.

References

1. Amoretti, M., Cagnoni, S.: Toward collective self-awareness and self-expression in distributed systems. IEEE Comput. **48**(7), 29–36 (2015)
2. Touch, J., Baldine, I., Dutta, R., Ford, B., Finn, G., Jordan, S., Massey, D., Matta, A., Papadopoulos, C., Reiher, P., Rouskas, G.: A dynamic recursive unified internet design (DRUID). Comput. Netw. **55**(4), 919–935 (2011)
3. Van Meter, R.: Quantum networking and internetworking. IEEE Netw. **26**, 59 (2012)
4. Gershenson, C., Fernandez, N.: Complexity and information: measuring emergence. Self Organ. Homeost. Multiple Scales Complex. **18**(2), 29–44 (2012)
5. Zubillaga, D., Cruz, G., Aguilar, L.D., Zapotecatl, J., Fernandez, N., Aguilar, J., Rosenblueth, D.A., Gershenson, C.: Measuring the complexity of self-organizing traffic lights. Entropy **16**(5), 2384–2407 (2014)
6. Amoretti, M., Gershenson, C.: Measuring the complexity of adaptive peer-to-peer systems. Peer Peer Network. Appl. **9**, 1031 (2016)

7. Kurose, J.F., Ross, K.W.: Computer Networking: A Top-Down Approach. Addison-Wesley, Boston (2012)
8. Akkaya, K., Younis, M.: A survey on routing protocols for wireless sensor networks. Ad Hoc Netw. **3**(3), 325–349 (2005)
9. Stojmenovic, I.: Simulations in wireless sensor and ad hoc networks: matching and advancing models, metrics, and solutions. IEEE Commun. Mag. **46**(12), 102–107 (2008)
10. Rotenberg, É., Crespelle, C., Latapy, M.: Measuring routing tables in the internet. In: 6th IEEE International Workshop on Network Science for Communication Networks, Toronto, Canada (2014)
11. Amoretti, M., Picone, M., Zanichelli, F., Ferrari, G.: Simulating mobile and distributed systems with DEUS and ns-3. In: HPCS, Helsinki, Finland (2013)