Brief Announcement: Performance Analysis of Cyclon, an Inexpensive Membership Management for Unstructured P2P Overlays

François Bonnet[1], Frédéric Tronel[2], and Spyros Voulgaris[3]

[1] École Normale Supérieure de Cachan/IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France
[2] IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France
[3] Department of Computer Science, ETH Zurich, 8092 Zurich, Switzerland

Motivations. Unstructured overlays form an important class of peer-to-peer networks, notably for content-based searching algorithms. Being able to build overlays with low diameter, that are resilient to unpredictable joins and leaves, in a totally distributed manner is a challenging task. Random graphs exhibit such properties, and have been extensively studied in literature. Cyclon algorithm is an inexpensive gossip-based membership management protocol described in detail in [1] that meets these requirements.

An Overview of Cyclon. For a detailed description of Cyclon algorithm, the reader should refer to [1]. Briefly, Cyclon supports two different modes of operation : a basic shuffling mode, and an enhanced one. The basic mode, the only one to be studied in this article is a purely random mode, while the second mode uses a timestamp mechanism to improve performance with respect to node failures behavior. Each node maintains a cache of neighbor nodes of size c, hence each node knows exactly c nodes in the overlay. To correctly initialize nodes caches, we assume the existence of a predefined set of well-known supernodes. During the execution of the protocol, each node performs periodically a shuffle step. For a given node p, a shuffle step consists in contacting one node q among its neighbors. Then p and q exchange $\ell \leq c$ nodes from their respective caches. ℓ is a parameter of the algorithm. Counterintuitively, simulations in [1], have shown that the i nfluence of parameter ℓ is negligible (except for limit cases, when ℓ is close to 1 or c). One of the most fundamental operation performed by the shuffling step is that p sends its own identity to q, and remove q from its own cache. Consequently, the edge from p to q is reversed by the shuffling step. This guarantees the connectivity of the underlying overlay. In this paper, we propose two models to analyse Cyclon performances in term of convergence speed, and quality of the obtained overlay. In our work we evaluate this quality from the distribution of the in-degree[4] of nodes. We are interested by this distribution since it is highly related to the robustness of the overlay in the presence of failures. This gives also an indication of the distribution of resource usage (processing, bandwidth) across node s. We are looking for a distribution as uniform as possible.

[4] The number of nodes that have an edge directed to the considered node. It is an integer in $[0, n-1]$ since we do not authorized loop edges.

Model #1. We assume that there are n nodes in the system, gossip exchanges are atomic, and are triggered by a global scheduler which picks at random the next process to perform a shuffle operation with uniform probability. This is of course, a rather coarse model of reality, where each process would certainly locally triggers its shuffling operation, using timeout expiration. In fact this model corresponds to a complete asynchronous system; even though this model is questionable (see [2]) it has been introduced by [3] and [4]. With this random scheduling a given process has probability $1 - e^{-1}$ of taking at least one computation step, when exactly n steps are triggered. This must be compared to a real system based on local uniform time triggers, where the same process would have probability 1 to perform a step, every n steps counted globally. We are interested in evaluating the in-degree evolution of a given node since it is a good measure of the quality of the obtained overlay. We model Cyclon algorithm by a discrete time Markov chain (DTMC) whose states space S is the possible in-degrees for a given process. Note that we can focus on a particular process because they are all equivalent with respect to the scheduler. The evolution of the in-degree of a node after a shuffle step depends only on its value before the step; the impact of the detailed structure of the network is negligible. Moreover during a step, the in-degree can only change by one. Thus, the n-square matrix M_1 associated to this DTMC is a tridiagonal one. Its upper diagonal is $m^1_{i,i+1} = -\frac{i}{n(n-1)} + \frac{1}{n}$, while its lower diagonal is $m^1_{i,i-1} = \frac{i}{n(n-1)} \frac{n-1-c}{c}$. Being a stochastic tridiagonal matrix, diagonal elements $m_{i,i}$ of M_1 are equal to $1 - m^1_{i,i+1} - m^1_{i,i-1}$. We show that the generating function $G_\lambda(z) = \sum_{i=0}^{n-1} v_\lambda[i] z^i$ associated to the eigenvector v_λ[5] satisfies a first-order differential equation whose solutions are equal to $\left(\frac{z-1}{cz+n-c-1}\right)^{cn(\lambda-1)} \left(\frac{cz-n-c-1}{n-1}\right)^{n-1}$. Since by definition, this function must be a polynomial of degree $n-1$, we can conclude that eigenvalues are $1 - \frac{k}{nc}$ with $k \in [0, n-1]$. In particular we obtain for $k = 0$, a closed form for the generating function $\pi(z)$ associated with the stationary distribution $\pi = v_1$, namely $\pi(z) = \left(\frac{cz+n-c-1}{n-1}\right)^{n-1}$. Thus $\pi[i] = \binom{n-1}{i} \left(\frac{n-c-1}{n-1}\right)^{n-1-i} \left(\frac{c}{n-1}\right)^i$. This corresponds exactly to the in-degree distribution of a purely random directed graph where each vertex has exactly c outgoing edges, a highly desirable property for unstructured overlays. Using well-known properties on generating functions we can establish that the mean value $\overline{\pi}$ of stationary distribution is equal to c, which naturally satisfies the balance equation in a directed graph[6]. Similarly we can establish that standard deviation of π is equal to $c + O(1/n)$. Since we have access to the eigenvalues, and in particular the second largest one, namely $1 - \frac{1}{nc}$, we can establish an upper bound on the convergence speed of the DTMC. Using classical linear algebra, we can show that $\max_{X_0} \frac{1}{2} \|X_0 M_1^t - \pi\|_1 \leq \left(1 - \frac{1}{nc}\right)^t$ where X_0 denotes the initial distribution. Mixing time $\tau_1(\epsilon)$ as defined in [5] is

[5] v_λ is such that $v_\lambda M_1 = \lambda v_\lambda$ for a given λ called an eigenvalue of M_1.

[6] The number of outgoing edges, nc in this particular case, is equal to the number of ingoing edges, which is equal to $n\overline{\pi}$

thus bounded by $nc \log \epsilon^{-1} + O(1)$, which shows that Cyclon is a fast mixing algorithm [7].

Model #2. We consider a more refined model, where processes are fairly scheduled. We consider now that a step in our model corresponds to a whole cycle of the protocol, i.e. in a step every node performs one and exactly one shuffle. Contrary to model #1, this model corresponds to a synchronous system: all nodes execute the same number of exchanges and at the same time. This refinement comes at the price of a more complex stochastic matrix M_2. The evolution of the in-degree of a node after a (model) step still depends only on its value before the step, but its variation may now be larger than one. In particular, M_2 is an lower hesselberg[8] matrix whose general term is $m^2_{0\leq j-1\leq i\leq n-1} = \binom{i}{j-1}\left(\frac{1}{c}\right)^{i+1-j}\left(\frac{c-1}{c}\right)^{j-1}$. We show that the generating function $G_\lambda(z)$ is solution of the functional equation : $\lambda G_\lambda(\frac{cz-1}{c-1}) = fraccz - 1c - 1G_\lambda(z) + \frac{c(1-z)}{c-1}\left(\frac{cz-1}{c}\right)^{n-1} v_\lambda[n-1]$. We can extract and solve a recurrence equation giving $G_\lambda(z)$ and all of its successive derivatives $G_\lambda^{(k\geq 0)}(z)$ at point $z = 1$. For $\lambda = 1$, by using a Taylor series expansion, and by the fact that $G_1(z)$ is a polynomial of degree $n-1$, we have access to a closed formula for $\pi(z)$ the generating function associated to the stationary distribution of the considered DTMC, namely $\pi(z) = \sum_{k=0}^{n-1} \frac{G_1^{(k)}(1)}{k!}(z-1)^k$. For $\lambda \neq 1$, the fact that $G_\lambda(z)$ is a polynomial of degree $n-1$ can be expressed as a set of constraints on the successive derivatives at point $z = 1$, namely $G_\lambda^{(k\geq n)}(1) = 0$. These constraints reduce to a polynomial of degree $n-1$ whose roots are exactly the $n-1$ eigenvalues $\lambda < 1$. We show that the second largest eigenvalue is smaller than $1 - \frac{1}{c}$. Hence mixing time $\tau_2(\epsilon)$ is bounded by $c \log \epsilon^{-1} + O(1)$. Note that this is compatible with model #1. Indeed in model #2, each step of the Markov process corresponds to n steps of the previous Markov process. This explains why $\frac{\tau_1(\epsilon)}{\tau_2(\epsilon)} = n$. The reader is invited to refer to [6] for a detailed version of the results.

References

1. Voulgaris, S., Gavidia, D., van Steen, M.: CYCLON: Inexpensive membership management for unstructured P2P overlays. J. Network Syst. Manage **13**(2) (2005)
2. Aspnes, J.: Randomized protocols for asynchronous consensus. DISTCOMP: Distributed Computing **16** (2003)
3. Bracha, G., Toueg, S.: Asynchronous consensus and broadcast protocols. JACM: Journal of the ACM **32** (1985)
4. Aspnes, J.: Fast deterministic consensus in a noisy environment. In: PODC. (2000) 299–308
5. Randall, D.: Rapidly mixing Markov chains with applications in computer science and physics. Computing in Science and Engineering **8**(2) (2006) 30–41
6. Bonnet, F., Tronel, F., Voulgaris, S.: Performance analysis of cyclon, an inexpensive membership management for unstructured p2p overlays. Technical Report 1807, IRISA (2006)

[7] $\tau_1(\epsilon)$ is bounded by a polynomial of $\log \epsilon^{-1}$ and n. See [5] for a precise justification.

[8] A matrix whose terms above the upper diagonal are zero.