Decentralized Aggregation Protocols in Peer-to-Peer Networks: A Survey

Rafik Makhloufi, Grégory Bonnet, Guillaume Doyen, and Dominique Gaïti

ICD/ERA, FRE CNRS 2848, Université de Technologie de Troyes, 12 rue Marie Curie, 10010 Troyes Cedex, France
{rafik.makhloufi,gregory.bonnet,guillaume.doyen,dominique.gaiti}@utt.fr

Abstract. In large scale decentralized and dynamic networks such as Peer-to-Peer ones, being able to deal with quality of service requires the establishment of a decentralized, autonomous and efficient management strategy. In this context, there is a need to know the global state of a network by collecting a variety of distributed statistical summary information through the use of aggregation protocols. In this paper, we carried out a study on a set of aggregation protocols that can be used for autonomous network monitoring purposes in P2P networks, and we propose a classification and a comparison of them.

Keywords: aggregation protocols, Peer-to-Peer decentralized monitoring, gossip-based protocols, tree-based protocols.

1 Introduction

Size, complexity and user requirements in existing networks are constantly increasing, making current static and centralized management frameworks unadapted. Peer-to-Peer (P2P) networks are a good example of such networks since they are known as large scale and dynamic ones, where resources are distributed over the peers. This context involves the use of new decentralized and autonomic management approaches in order to ensure some level of performance and QoS. To reach this goal, a variety of statistical summary data about the network must be collected in order to infer global information about the system. These distributed numeric values are collected by the aggregation protocols. Aggregation is intended as a summarizing mechanism of the overall state within the network [25]. It refers to a set of functions that produce an indicator to evaluate a global property of a system [20]. The summary data are obtained through the use of a set of functions named aggregate functions. According to [26], there are basic aggregate functions: counts, sums, averages, minima and maxima, and over these simple aggregates, more advanced aggregates can be computed such as: histograms [13], parameter estimations [24,30], spectral analysis [16] or random linear projections [23]. Another classification of the aggregate functions is done in [18] according to other properties (e.g. duplicate sensitivity and monotony).

In this paper, we carry out a study of some of the aggregation protocols. We propose a new classification followed by the description of these protocols in Section 2 and we compare previously investigated protocols in Section 3 .

J.C. Strassner and Y.M. Ghamri-Doudane (Eds.): MACE 2009, LNCS 5844, pp. 111–116, 2009.

2 Taxonomy of Aggregation Protocols

Several aggregation protocols are proposed in large scale and dynamic networks such as grids, P2P, MANETs or sensor networks. In this paper, we focus our study on those that can be applied in a P2P context. Aggregation protocols are often classified in two prevailing categories: gossip-based protocols and tree-based protocols. Nonetheless, considering only these categories hides a lot of features and differences. Thus, we propose to refine this basic classification by introducing the following criteria, as shown in Figure 1.

1. **Network structure:** according to the degree of network structure, aggregation protocols can be classified into three categories: tree-based, gossip-based and hybrid protocols. Unlike tree-based techniques where nodes are organized into a tree, gossip-based protocols do not require a particular structure [5,10,14]. Computation of aggregates in tree-based techniques is often done hierarchically in a bottom-up fashion [3,9,17]. Finally, hybrid protocols combine a gossip dissemination mechanism with a tree structure.
2. **Propagation technique:** two way exist in which a node can exchange information with its neighbors: reactive and proactive[1] propagation techniques. Nodes use a reactive approach to reply by processing the query when a sender explicitly requests to compute an aggregate. A node uses the proactive approach to compute aggregates without explicitly asking for them (e.g. at each time interval or when changes occur) [12,19].
3. **Network view:** aggregates contain information about the global state of a network, but they may be computed across nodes in a neighborhood, a network domain or the entire network. Then, a node may have a situated view, limited to the neighborhood or a part of the network, or may have a global view about the entire network [8]. In our classification, a protocol has a situated view when it uses an explicit parameter that can be adjusted in order to have a view about some nodes or all the network nodes.
4. **Neighborhood information:** communication between nodes can be blind or informed. In blind communication, nodes do not hold information about other nodes. Thus a node selects neighbors to exchange information uniformly at random. By contrast, informed communication methods use heuristics for node selection (i.e. a non-uniform probabilistic distribution) [19].

2.1 Gossip-Based Aggregation Schemes

In each round of the basic gossip algorithms [7], a random pair of neighboring nodes is chosen to exchange their information. According to [26], gossip-based protocols can be divided regarding node selection into uniform gossip and standard gossip protocols. In uniform gossip, each node chooses to exchange information with a uniformly chosen node [14]. In standard gossip, neighbors are chosen according to a non-uniform probabilistic distribution [15,24]. Kempe et al. [14]

[1] Reactive and proactive methods are also called, respectively, pull and push methods.

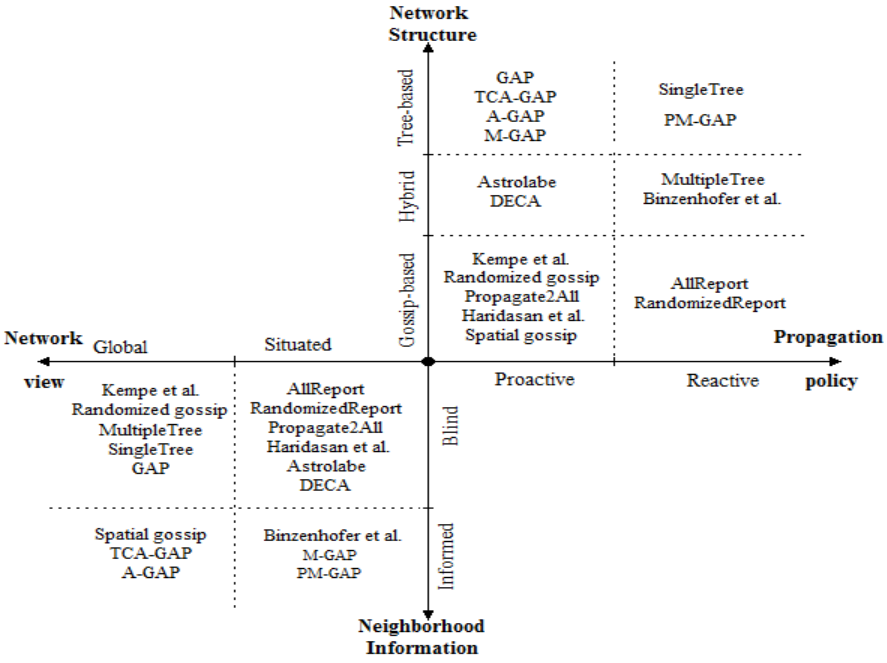

Fig. 1. Taxonomy of aggregation protocols in a P2P network

propose some uniform aggregation algorithms (e.g. the proactive blind `Push-sum` algorithm with a global view). Moreover, Bawa et al. [3] propose `Propagate2All` that enables a situated view by using a diameter D that denotes the upper bound to which the network is known, and they also present some reactive protocols: `AllReport` and `RandomizedReport` where a node replies with probability p. In the same category as `Propagate2All`, Haridasan et al. [11] propose an estimation-based protocol using data synopsis techniques [1]. An informed aggregation protocol named `Spatial gossip` is proposed in [15,24] where node selection is done according to the distance between the nodes.

2.2 Tree-Based Aggregation Schemes

Tree-based protocols use a tree for computing aggregates in a bottom-up fashion. The most basic protocol in this category is the blind reactive global view algorithm `SingleTree` [3], where a node q broadcasts a query to construct a spanning tree on the network. Dam and Stadler [9] propose a proactive protocol named GAP (Generic Aggregation Protocol) that builds and maintains a BFS spanning tree on the overlay network and uses it to incrementally and continuously compute and propagate aggregates. Improvements made on GAP gave form to several other protocols: a situated view and informed M-GAP where

aggregates are available at all nodes, a reactive version of M-GAP named PM-GAP [29], TCA-GAP [28] for the decentralized detection of threshold crossings, A-GAP [21,22] an adaptive extension of GAP.

2.3 Hybrid Aggregation Schemes

In order to combine the benefits of both gossip and tree, some protocols propose an hybrid approach that uses a gossip dissemination over a tree structure. Bawa et al. [3] propose the reactive blind situated view algorithm `MultipleTree`, an enhancement of `SingleTree` that creates k independent spanning trees rooted at the querying node. Moreover, others propose to organize nodes into a hierarchy of a maximal height h) with a situated view. For exemple, `Astrolabe` [6,25] is a distributed information management system that uses gossip to construct an overlay tree for computing aggregates. For structured networks, Artigas et al. propose the use of `DECA` [2], where nodes are organized into clusters and super-clusters. In this same context, Binzenhofer et al. [4] propose an informed and reactive approach that uses a snapshot algorithm at an arbitrary peer of a Chord DHT [27] to divide recursively the overlay into contiguous subparts of size c.

3 Comparison of Aggregation Protocols

We carry out a comparison of some aggregation protocols seen above according to the following evaluation criteria: (1) the computation cost for a protocol is the maximum computation cost among all the nodes in the network, and for a single node the computation cost is the number of steps taken by the process that is executed on the node; (2) the communication cost is the sum of sizes of messages sent between any node pairs during aggregation; (3) the robustness that defines the capacity of a system to operate correctly and to ensure accuracy despite of external factors such as node or link failures; (4) the convergence time that is the necessary time between the initialization of the aggregation and the time when all nodes (or querying node) hold the aggregation results (i.e. the elapsed time for both communication and computation).

The collected comparison results presented in Table 1 obtained from the literature show that gossip-based protocols ensure fault-tolerance due to their decentralization, and also provide a better convergence time due to the gossip diffusion speed. However, the large number of exchanged messages causes more overhead and a high communication and computation cost. Tree-based protocols execute themselves in a better convergence time and a lower communication and computation cost due to their optimization of the number of exchanged messages on the tree. However, the hierarchical structure and the unique path between each node and the root let tree-based protocols be more sensitive to faults than the decentralized gossip-based protocols. Hybrid protocols combine the benefits of the two approaches. So, they are more resilient than tree-based protocols and they have less overhead than gossip-based protocols.

Table 1. Comparison of aggregation protocols

		Communication cost	Computation cost	Convergence time	Robustness
Gossip	**Kempe et al.**	$O(n \log n)$	$O(\log n)$	$O(\log n)$	$\surd$
	Propagate2All	$O(mnr)$	$O(n)$	$O(\hat{D})$	$\surd$
	AllReport	$O(m+n)$	$O(n)$	NA	$\surd$
	Randomized-Report	$O(m+np)$	$O(n)$	NA	$\surd$
	Spatial gossip	$O(\sqrt{n})$	$O(n \log n)$	$O(n^{3/2} \log n)$	$\surd$
Tree	**SingleTree**	$O(m+nr)$	$O(b)$	$O(D)$	$\times$
	GAP	$1.7msg/sec/node*$	NA	$1.5-3sec*$	$\times$
	A-GAP	NA	$< 0.5sec/node*$	$< 4ms*$	$\surd$
	TCA-GAP	$\simeq GAP*$	NA	$0.5-1.75sec*$	$\times$
Hybrid	**Multiple-Tree**	$O(m+knr)$	$O(k+b)$	$O(D)$	$\surd$
	Astrolabe	$\simeq 1msg/round/node*$	$O(\log n)$	NA	$\surd$
	DECA	$O(logn)$	$O(rn \log n + rn)$	$O(\log c + h)$	$\surd$

$\surd$: supported; $\times$: not supported; NA: not available; $*$: depends on experiments.

4 Conclusion and Future Work

In this paper, we handle the problem of decentralized monitoring and computing aggregates in P2P networks. We examine and classify a set of aggregation protocols for P2P networks. We also address a comparison of these protocols. The aggregation protocol has to be simple, scalable and ensures robustness, convergence time and low communication and computation cost. Despite the amount of work carried out towards the development of efficient aggregation protocols, there is no protocol that guarantees all the desired performance criteria at the same time. Gossip-based protocols are generally simple, scalable and more resilient to faults, but they cause a high communication and computation cost while tree-based protocols cause a lower communication cost with sensitivity to faults. We plan to extend our work by consolidating the comparison of aggregation protocols with more performance criteria and quantitative results.

References

1. Aggarwal, C.C., Yu, P.S.: A survey of synopsis construction in data streams. In: Data streams: models and algorithms. Springer, Heidelberg (2006)
2. Artigas, M.S., López, P.G., Gómez-Skarmeta, A.F.: DECA: a hierarchical framework for decentralized aggregation in DHTs. In: State, R., van der Meer, S., O'Sullivan, D., Pfeifer, T. (eds.) DSOM 2006. LNCS, vol. 4269, pp. 246–257. Springer, Heidelberg (2006)
3. Bawa, M., Garcia-Molina, H., Gionis, A., Motwani, R.: Estimating aggregates on a peer-to-peer network. Tech. rep., Stanford InfoLab (2003)
4. Binzenhofer, A., Kunzmann, G., Henjes, R.: A scalable algorithm to monitor chord-based p2p systems at runtime. In: Proc. IPDPS (2006)
5. Birman, K.: The promise, and limitations, of gossip protocols. SIGOPS Oper. Syst. Rev. 41(5), 8–13 (2007)
6. Birman, K., van Renesse, R., Vogels, W.: Scalable data fusion using astrolabe. In: Proc. FUSION (2002)

7. Boyd, S., Ghosh, A., Prabhakar, B., Shah, D.: Randomized gossip algorithms. IEEE/ACM Trans. Netw. 14, SI, 2508–2530 (2006)
8. Bullot, T., Khatoun, R., Hugues, L., Gaïti, D., Merghem-Boulahia, L.: A situatedness-based knowledge plane for autonomic networking. Int. J. Netw. Manag. 18(2), 171–193 (2008)
9. Dam, M., Stadler, R.: A generic protocol for network state aggregation. In: Proc. RVK (2005)
10. Dietzfelbinger, M.: Gossiping and broadcasting versus computing functions in networks. Discrete Appl. Math. 137(2), 127–153 (2004)
11. Haridasan, M., van Renesse, R.: Gossip-based distribution estimation in peer-to-peer networks. In: Proc. IPTPS (2008)
12. Jelasity, M., Montresor, A., Babaoglu, O.: Gossip-based aggregation in large dynamic networks. ACM Trans. Comput. Syst. 23(3), 219–252 (2005)
13. Jurca, D., Stadler, R.: Computing histograms of local variables for real-time monitoring using aggregation trees. In: Proc. IM (2009)
14. Kempe, D., Dobra, A., Gehrke, J.: Gossip-based computation of aggregate information. In: Proc. FOCS (2003)
15. Kempe, D., Kleinberg, J., Demers, A.: Spatial gossip and resource location protocols. In: Proc. STOC (2001)
16. Kempe, D., McSherry, F.: A decentralized algorithm for spectral analysis. In: Proc. STOC (2004)
17. Li, J., yoh Lim, D.: A robust aggregation tree on distributed hash tables. In: Proc. MIT SOW (2004)
18. Madden, S., Franklin, M.J., Hellerstein, J.M., Hong, W.: TAG: a tiny aggregation service for ad-hoc sensor networks. SIGOPS Oper. Syst. Rev. 36, SI, 131–146 (2002)
19. Meshkovaa, E., Riihijärvia, J., Petrovaa, M., Mähönen, P.: A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks. Computer Networks 52(11), 2097–2128 (2008)
20. Montresor, A., Jelasity, M., Babaoglu, O.: Robust aggregation protocols for large-scale overlay networks. In: Proc. DSN (2004)
21. Prieto, A.G., Stadler, R.: Adaptive distributed monitoring with accuracy objectives. In: Proc. INM (2006)
22. Prieto, A.G., Stadler, R.: A-GAP: an adaptive protocol for continuous network monitoring with accuracy objectives. IEEE TNSM 4(1), 2–12 (2007)
23. Rabbat, M., Haupt, J., Singh, A., Nowak, R.: Decentralized compression and predistribution via randomized gossiping. In: Proc. IPSN (2006)
24. Rabbat, M.G.: On spatial gossip algorithms for average consensus. In: Proc. SSP (2007)
25. Renesse, R.V., Birman, K.P., Vogels, W.: Astrolabe: a robust and scalable technology for distributed system monitoring, management, and data mining. ACM Trans. Comput. Syst. 21(2), 164–206 (2003)
26. Sarkar, R., Zhu, X., Gao, J.: Hierarchical spatial gossip for multi-resolution representations in sensor networks. In: Proc. IPSN (2007)
27. Stoica, I., Morris, R., Karger, D., Kaashoek, M.F., Balakrishnan, H.: Chord: a scalable peer-to-peer lookup service for internet applications. In: Proc. SIGCOMM (2001)
28. Wuhib, F., Dam, M., Stadler, R.: Decentralized detection of global threshold crossings using aggregation trees. Computer Networks 52(9), 1745–1761 (2008)
29. Wuhib, F., Stadler, R.: M-GAP: a new pattern for cfengine and other distributed software. Tech. rep., Royal Institute of Technology, KTH (2008)
30. Xiao, L., Boyd, S., Lall, S.: A scheme for robust distributed sensor fusion based on average consensus. In: Proc. IPSN (2005)