Optimal Broadcast Strategy in Homogeneous Point-to-Point Networks

Franco Robledo, Pablo Rodríguez-Bocca, and Pablo Romero

Instituto de Computación, INCO
Facultad de Ingeniería - Universidad de la República
Montevideo Uruguay
(frobledo,prbocca,promero)@fing.edu.uy,

Abstract. In this paper we address a fundamental combinatorial optimization problem in communication systems. A fully-connected system is modeled by a complete graph, where all nodes have identical capacities. A message is owned by a singleton. If he/she decides to forward the message simultaneously to several nodes, he/she will take longer, with respect to a one-to-one forwarding scheme. The only rule in this communication system is that a message can be forwarded by a node that owns the message. The *makespan* is the time when the message is broadcasted to all the nodes. The problem under study is to select the communication strategy that minimizes both the makespan and the average waiting time among all the nodes. A previous study claims that a sequential or *one-to-one* forwarding scheme minimizes the average waiting time, but they do not offer a proof. Here, a formal proof is included. Furthermore, we show that the sequential strategy minimizes the makespan as well. A discussion of potential applications is also included.

Keywords: Communication System, Forwarding, Waiting Time, Makespan.

1 Motivation

The Internet is supported by the client-server architecture, where users connect with a specific server to download data. This architecture has some benefits. The service is both simple and highly predictable. However, the server infrastructure is not scalable when demand is increased. A natural idea to overcome this scalability issue is to consider content popularity, where the most popular contents can be shared by the users. The server invites users to communicate and offer those files which are normally replicated in the network. An abstraction of this concept is accomplished with peer-to-peer systems (P2P for short). They are self-organized virtual communities developed on the Internet Infrastructure, where users, called peers, share resources (content, bandwidth, CPU-time, memory) to others, basically because they have common interests. From a game-theoretic point of view, cooperation is better than competition. From an engineering point of view, we understand that the power of user-cooperation in P2P systems is maximized, but the real-life design is jeopardized by other factors. Indeed, broadband resources are better exploited with cooperation. The altruistic behaviour in P2P networks is achieved with incentives, using a *give-to-get* concept [14, 2]. Nevertheless, the design of a resilient P2P network has several challenges. Indeed, the Internet access infrastructure is usually asymmetric, hindering peer exchange; peers arrive and depart the system when they wish [15]; free-riders exploit network resources but do not contribute with the system; a failure in the underlying network usually damage the P2P service; there is an explicit trade-off between the full knowledge of the network (topology, peers resources) and payload, which directly impacts in the throughput and network performance.

The main purpose of this paper is to understand the best forwarding schemes in an ideal abstract setting, and how different forwarding schemes are used in real-life systems. Even though we motivate this paper by P2P systems, our main result apply to several communication systems, such as scheduling in parallel unrelated machines, social networks and content delivery networks. Our work is inspired by a fundamental problem posed for the first time by Qiu and Srikant, where they state that *it should be clear that a good strategy* is the one-to-one forwarding scheme [13]. Even though the authors study the service capacity of a file sharing peer-to-peer system, its formulation is general enough. For practical purposes they find a closed formula for the average waiting time following a one-to-one forwarding scheme when the population N is a power of two. In [9], a formal proof that the one-to-one forwarding scheme achieves the minimum waiting time is included, when the population is a power of two. Here, we formally prove that it is not only good, but also optimal, for both makespan and waiting time measures. The result holds for an arbitrary population size. We remark that the optimum forwarding scheme rarely appears in real-life systems. We discuss this phenomenon showing the gap between this theoretical result and real-life implementations of communication systems. The main contributions of this paper are two-fold:

1. The best forwarding scheme in complete homogeneous communication networks is found.
2. The gap between the best theoretical forwarding scheme and real-life implementations is discussed.

This paper is organized as follows. The problem under study is presented in Section 2. The mathematical analysis provides a full solution of the problem, which is derived in Section 3. The gap between theory and real-life applications is considered in Section 4. Section 5 contains the main conclusions and trends for future work.

2 Problem

We are given a full network composed by N peers with identical capacity b (in bits per second), and a message with size M (measured in bits). A single node that belongs to the network owns the message, and at time $t_1 = 0$ he/she forwards the message to one or several peers belonging to the network.

Let us denote $\tau = M/b$ the time-slot following one-to-one forwarding time. If some peer sends the message to c other peers, it will take $c\tau$ seconds to perform the forwarding task. Let us denote $0 = t_1 \leq t_2 \leq \ldots \leq t_N$ the corresponding completion times of the N peers in this cooperative system. The *makespan* is t_N, while the *average waiting time*, $\bar{t}$, is the average over the set $\{t_1, \ldots, t_N\}$. Clearly, $\bar{t} \leq t_N$.

In a *one-to-many* forwarding scheme, every peer selects a fixed number c of peers to forward the message. In general, in a simultaneous forwarding scheme there is some peer i that, at time t_i, simultaneously forwards the message to more than one peer. In contrast, the only remaining strategy is a sequential or *one-to-one* forwarding strategy.
The goal in the Minimum Point-to-Point Makespan (MPTPM) is to minimize t_N:

$$\min_{s \in \mathcal{S}} \max_{1 \leq i \leq N} t_i, \tag{1}$$

being s a member belonging to the family of forwarding strategies $\mathcal{S}$, where each node decides to send the message either to one or to multiple peers, and may decide a delay to start the transmission as well.

An analogous problem is the Minimum Point-to-Point Waiting Time (MPTPWT), where the goal is to minimize $\bar{t}$ among all possible forwarding strategies:

$$\min_{s \in \mathcal{S}} \frac{1}{N} \sum_{i=1}^{N} t_i, \tag{2}$$

In a *one-to-many* forwarding scheme, every peer selects a fixed number c of peers to forward the message. In general, in a simultaneous forwarding scheme there is some peer i that, at time t_i, simultaneously forwards the message to more than one peer. In contrast, the only remaining strategy is a sequential or *one-to-one* forwarding strategy. A peer can decide to delay the transmission as well.

Here we formally prove that the one-to-one forwarding strategy is optimal for both the MPTPM and MPTPWT. For short, we will use $n = \lceil log_2(N) \rceil$ and $n_c = \lceil log_c(N) \rceil$.

3 Solution

A straight calculation provides the makespan and average waiting time in the one-to-one forwarding scheme:

Lemma 1. *The makespan in the one-to-one forwarding scheme is $n\tau$.*

Proof. The message is fully owned by 2^i peers at time $i\tau$, for $i = 1, \ldots, n-1$. The remaining $N - 2^{n-1}$ peers receive the message at time $n\tau$. □

Lemma 2. *The average waiting time in the one-to-one forwarding scheme is $\bar{t} = \frac{\tau}{N}(nN - 2^n + 1)$.*

Proof.

$$\begin{aligned}
\bar{t} &= \frac{1}{N}[\sum_{i=1}^{n-1} 2^{i-1} i\tau + (N - 2^{n-1})\tau] \\
&= \frac{\tau}{N}[(n2^{n-1} - 2^n + 1) + (N - 2^{n-1})n] \\
&= \frac{\tau}{N}(nN - 2^n + 1).
\end{aligned}$$

□

Since $n = \lceil log_2(N) \rceil$, we conclude that both makespan and average waiting time grow logarithmically with the population of the system and linearly with respect to the time-slot τ when the one-to-one strategy is considered. Let us contrast the result with a one-to-many strategy in what follows, where each peer forwards the message to $c - 1$ different peers, for some $c > 2$.

Lemma 3. *The makespan in the one-to-many forwarding scheme of type $c - 1$ is $n_c(c-1)\tau$.*

Proof. By the definition of one-to-many forwarding schemes of type $c - 1$, there are $N_i = c^{i-1}(c-1)$ peers whose completion time is $T_i = i(c-1)\tau$. Therefore, the message is fully owned by c^i peers at time T_i, for $i = 1, \ldots, n_c - 1$. The remaining $N - c^{n_c - 1}$ peers receive the message at time $T_{n_c} = n_c(c-1)\tau$. □

Lemma 4. *The average waiting time in a one-to-many forwarding scheme of type $c - 1$ is $\bar{t}_{c-1} = \frac{\tau}{N}[n_c(c-1)N - c^{n_c} + 1]$.*

Proof.

$$\begin{aligned}
\bar{t}_{c-1} &= \frac{1}{N}[\sum_{i=1}^{n_c-1} N_i T_i + (N - c^{n_c-1})T_{n_c} \\
&= \frac{1}{N}[\sum_{i=1}^{n_c-1} c^{i-1}(c-1)i(c-1)\tau + (N - c^{n_c-1})T_{n_c}] \\
&= \frac{\tau}{N}[n_c c^{n_c-1}(c-1) - (c^{n_c} - 1) + (N - c^{n_c-1})T_{n_c}] \\
&= \frac{\tau}{N}[n_c(c-1)N - c^{n_c} + 1].
\end{aligned}$$

□

On one hand, we can check that $\bar{t}_{c-1}$ equals $\bar{t}$ when $c = 2$, as expected. In fact, the one-to-one strategy is the one-to-many if $c - 1 = 1$. On the other, $\bar{t}_{c-1} > \bar{t}$ for every $c > 2$. The makespan is studied first:

Lemma 5. *The makespan in the one-to-one strategy is never greater than in the one-to-many strategy.*

Proof. If $c = 2$ we see that $n_2 = n$, so $(c-1)n_c = n$. It suffices to prove that $(c-1)n_c \geq n$ for any $c \geq 3$, being $n = \lceil log_2(N) \rceil$ and $n_c = \lceil log_c(N) \rceil$:

$$\begin{aligned}(c-1)n_c &\geq log_c(N^{c-1}) = (c-1)log_c(2)log_2(N) \\ &= log_c(2^{c-1})log_2(N) > log_2(N);\end{aligned}$$

where the last inequality follows from the fact that $2^{c-1} > c$ whenever $c \geq 3$. Since $(c-1)n_c$ is an integer, we obtain that $(c-1)n_c \geq \lceil log_2(N) \rceil$, and the result follows. □

A technical lemma will be used in the main result:

Lemma 6. *Given two partitions of* $N = \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i$ *such that* $x_i \geq y_i \geq 0, \forall i = 1, \ldots, m-1$ *and* $0 \leq x_m < y_m$. *Consider an arrange of times* $0 \leq t_1 \leq t_2 \leq \ldots \leq t_m$, *and a partition for each* t_i, $t_i = \sum_{j=1}^{m_i} t_{ij}$, *where* $0 \leq t_{ij} \leq t_i$. *Given any related partition of* x_i, $x_i = \sum_{j=1}^{m_i} x_{ij}$, *then* $\overline{W}_x = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{m_i} x_{ij}t_{ij}$ *is strictly lower than* $\overline{W}_y = \frac{1}{N}\sum_{i=1}^{m} y_i t_i$.

Proof.

$$\begin{aligned}\overline{W}_x &= \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{m_i} x_{ij}t_{ij} < \frac{1}{N}\sum_{i=1}^{m} x_i t_i \\ &= \frac{1}{N}(\sum_{i=1}^{m-1} x_i t_i + x_m t_m) \\ &= \frac{1}{N}(\sum_{i=1}^{m-1}(x_i - y_i)t_i + \sum_{i=1}^{m-1} y_i t_i - y_m t_m + x_m t_m) \\ &= \overline{W}_y + \sum_{i=1}^{m-1}(x_i - y_i)t_i - (y_m - x_m)t_m \\ &= \overline{W}_y + \sum_{i=1}^{m-1}(x_i - y_i)t_i - (\sum_{i=1}^{m}(x_i - y_i))t_m < \overline{W}_y.\end{aligned}$$

□

In words, if more peers own the message at any time t_i using strategy x instead of y ($x_i \geq y_i, i = 1, \ldots, m-1$) and the population is constant (N is constant, so $x_m < y_m$), then x outperforms y in terms of average waiting time.

Lemma 7. *The average waiting time in the one-to-one strategy is never greater than in the one-to-many strategy.*

Proof. When $t_i = i(c-1)\tau$ we know that c^i peers own the message following the one-to-many forwarding scheme of type $c-1$ versus $2^{i(c-1)}$ following the one-to-one forwarding scheme. By Lemma 6, it suffices to prove that $2^{i(c-1)} \geq c^i$ whenever $c \geq 3$. Taking logarithms on both sides yields $c - 1 \geq log_2(c)$, which holds for all $c \geq 2$. □

In order to illustrate the previous results, Figures 1 and 2 present the makespan and average waiting times respectively for one-to-one and one-to-many strategies in representative cases.

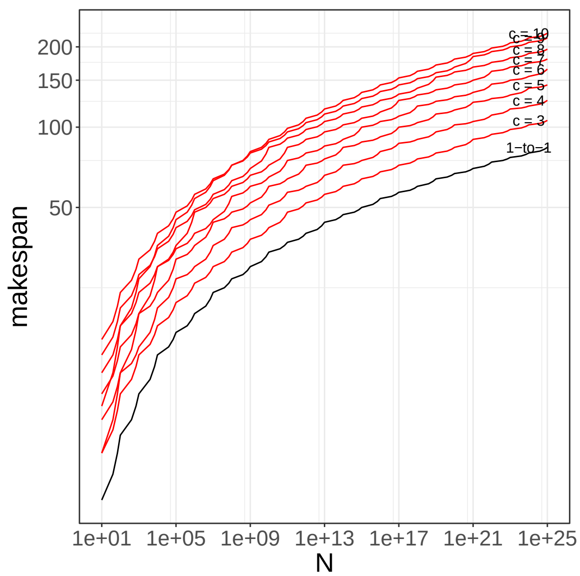

Fig. 1. Makespan as a function of the size of the system N, for $c \in \{1, \ldots, 10\}$).

Lemma 8 (Local Replacement). *If we are given a strategy where some peer x forwards the message to k new peers in a given time-slot $[t, t+T]$, and there exists an alternative strategy where x forwards the message to $k' > k$ peers in the same time-slot, then the local replacement for the alternative strategy in x reduces both the makespan and average waiting time if all the k' nodes behave as in the original strategy.*

Proof. During the specific time-slot $[t, t+T]$, the message is fully owned by more peers. By Lemma 6, the local replacement has lower average waiting time.

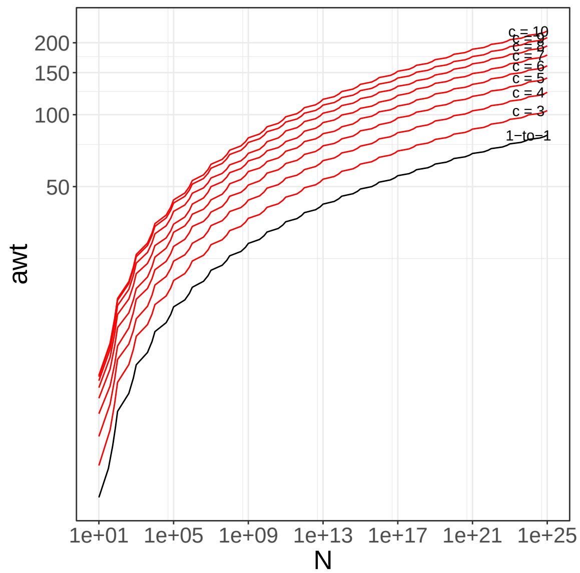

Fig. 2. Average waiting time as a function of the size of the system N, for $c \in \{1, \dots, 10\}$).

Analogously, there are more successors of x, so they feed more peers and the makespan is lower as well. □

Theorem 1 (Main Result). *The one-to-one forwarding scheme is optimal for both the MPTPM and the MPTPWT.*

Proof. If some peer deliberately produces a positive delay in the forwarding, there is a corresponding shift in both makespan and average waiting time. Therefore, delays are not included in an optimal strategy. If some peer x forwards the message to $c - 1 > 1$ nodes, we can consider a local replacement into the one-to-one strategy for x. By Lemmas 5 and 7, the one-to-one forwarding scheme offers lower makespan and average waiting times. By Lemma 8, a local replacement improves both measures. A local replacement is conducted in every node that forwards the message to many nodes. The result is a one-to-one forwarding scheme. □

It is worth to mention that a historical problem from telephonic services is the Minimum Broadcast Time or MBT for short [5]. In the MBT, we are given a simple graph, and a target node which owns the message. The goal is to select a forwarding one-to-one strategy, in order to minimize the broadcast time (i.e., the makespan). Observe that we studied complete networks. However, the makespan in an arbitrary simple (non-complete) graph is extremely challenging. In fact, the problem is formally known as the Minimum Broadcast

Time (MBT), and it belongs to the class of $\mathcal{NP}$-Complete problems [6]. The MBT is equivalent to find a spanning tree rooted at the holder of the message with the minimum makespan. The hardness of the MBT promotes the development of metaheuristics. In particular, a Greedy randomized heuristic is already available in the literature, together with an efficient Integer Linear Programming (ILP) formulation for the MBT [4]. In the following result, we consider the MPTPM for general, for non-complete networks, here called MPTPMNC:

Theorem 2. *Finding the optimal strategy for the MPTPMNC belongs to the class of $\mathcal{NP}$-Hard problems.*

Proof. Consider the MPTPMNC in general, for an arbitrary graph G and target node v. By an iterative application of Lemma 8, a one-to-many strategy can by replaced by a one-to-one strategy, minimizing the makespan. Then, the globally optimum solution for the MPTPMNC is precisely the globally optimum solution for the MBT. Then, the MPTPMNC is at least as hard as the MBT. Since the latter belongs to the class of $\mathcal{NP}$-Hard problems [6], the result follows. □

4 Discussion

As far as we know, the MPTPWT was posed for the first time by Yang and de Veciana [13]. The authors study the service capacity of a file sharing peer-to-peer system, and the problem under study serves as a fluid model for replication. They literally state that *it should be clear that a good strategy* is the one-to-one forwarding scheme. For practical purposes they find a closed formula for the average waiting time following a one-to-one forwarding scheme when N is a power of two. In [9], a formal proof of Lemma 7 is provided when the population is a power of two. Here, we formally prove that it is not only good, but also optimal, for both makespan and waiting time measures (this is, for the MPTPM as well). The result holds for an arbitrary population size.

Theorem 1 is counterintuitive, and could be used in several fields of knowledge. For instance, the earliest-finish-time in the context of parallel computing systems is precisely our makespan, and forwarding strategies are identified with a formal scheduling on this machines [1, 11]. The main goal in a Content Delivery Network is to minimize the delivery time, which is strictly related with makespan and average waiting time [8, 7, 12]. The time needed to distribute information in a social network, or a virus by an epidemic, are one of the main factors studied in these disciplines [10, 3]. Several real networks use one-to-many forwarding schemes. This fact suggests that in practice at least one assumption does not hold. First, we remark that full connectivity holds in overlay networks, but does not hold in most real-life scenarios, such as social networks. Second, there is no matching between modelling and reality when identical capacity is assumed. Last but not least, in an information-centric network the behaviour of nodes could be affected with information.

5 Conclusions and Trends for Future Work

In this paper we show that a one-to-one forwarding scheme provides both the lowest makespan and average waiting time, under complete homogeneous networks. The merit of this strategy was suggested by previous authors in the context of peer-to-peer systems for average waiting times. Forwarding schemes have a direct implication in many different contexts, such as scheduling in parallel unrelated machines, social networks and content delivery networks. As a future work, we would like to extend our analysis to incomplete graphs with heterogeneous and dynamic nodes. Observe that the optimum forwarding scheme is completely deterministic. Furthermore, we would like to better understand the gap between the theoretical predictions from this paper and real-life applications such as social networks and cellular systems.

6 Acknowledgment

This work is partially supported by MATHAMSUD 19-MATH-03 Raredep, *Rare events analysis in multi-component systems with dependent components*, STICAMSUD 19-STIC-01 ACCON, *Algorithms for the capacity crunch problem in optical networks* and Fondo Clemente Estable *Teoría y Construcción de Redes de Máxima Confiabilidad*.

References

1. T. L. Casavant and J. G. Kuhl. A taxonomy of scheduling in general-purpose distributed computing systems. *IEEE Transactions on Software Engineering*, 14(2):141–154, Feb 1988.
2. Bram Cohen. Incentives Build Robustness in BitTorrent. *www.bramcohen.com*, 1:1–5, May 2003.
3. Daryl J Daley, Joe Gani, and Joseph Mark Gani. *Epidemic modelling: an introduction*, volume 15. Cambridge University Press, 2001.
4. Amaro de Sousa, Gabriela Gallo, Santiago Gutierrez, Franco Robledo, Pablo Rodrguez-Bocca, and Pablo Romero. Heuristics for the minimum broadcast time. *Electronic Notes in Discrete Mathematics*, 69:165 – 172, 2018. Joint EURO/ALIO International Conference 2018 on Applied Combinatorial Optimization (EURO/ALIO 2018).
5. Arthur M. Farley. Broadcast time in communication networks. *SIAM Journal on Applied Mathematics*, 39(2):385–390, 1980.
6. Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman & Company, New York, NY, USA, 1979.
7. Yong Liu, Yang Guo, and Chao Liang. A survey on peer-to-peer video streaming systems. *Peer-to-Peer Networking and Applications*, 1(1):18–28, 2008.
8. Dongyu Qiu and R. Srikant. Modeling and performance analysis of bittorrent-like peer-to-peer networks. In *Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM '04, pages 367–378, New York, NY, USA, 2004. ACM.

9. Pablo Romero. *Mathematical Analysis of Scheduling Policies in Peer-to-Peer video streaming networks*. PhD thesis, Universidad de la República, Montevideo, Uruguay, November 2012.
10. Kazumi Saito, Ryohei Nakano, and Masahiro Kimura. Prediction of information diffusion probabilities for independent cascade model. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 67–75. Springer, 2008.
11. H. Topcuoglu, S. Hariri, and Min-You Wu. Performance-effective and low-complexity task scheduling for heterogeneous computing. *IEEE Transactions on Parallel and Distributed Systems*, 13(3):260–274, Mar 2002.
12. A. Vakali and G. Pallis. Content delivery networks: status and trends. *IEEE Internet Computing*, 7(6):68–74, Nov 2003.
13. Xiangying Yang and Gustavo de Veciana. Service Capacity of Peer to Peer Networks. In *Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies (INFOCOM'04)*, volume 4, pages 2242–2252, 2004.
14. Bin Yu and Munindar P. Singh. *Incentive Mechanisms for Peer-to-Peer Systems*, pages 77–88. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
15. Y. Zhou, L. Chen, C. Yang, and D. M. Chiu. Video popularity dynamics and its implication for replication. *IEEE Transactions on Multimedia*, 17(8):1273–1285, Aug 2015.