Proofs for "Generosity Pays Off: A Game-Theoretic Study of Cooperation in Decentralized Learning"

G. Di Giacomo[†], F. Malandrino[‡,*], C. F. Chiasserini[†,‡,*]
[†]Politecnico di Torino, Italy; [‡]CNR-IEIIT, Italy; [*]CNIT, Italy

Appendix

We provide the proofs of the theorems of "Generosity Pays Off: A Game-Theoretic Study of Cooperation in Decentralized Learning", which are adapted from the proofs of [1].

A. Proof of Theorem 1

Theorem 1. *Consider a system of N nodes, all belonging to class 1, with mean and maximum computing capability constraint (resp.) $\bar{\chi}_1$ and $\chi_1^{\max}$. Models have complexity m_1, which implies that d_1 samples and e_1 epochs are required to reach the target learning quality. Also, $H{=}h{=}1$, and if the node receiving the training request accepts, the time required for the training is T_{11}, otherwise a time T_w is waited. Then:*

1) *if all nodes, except for node n, employ GENIAL, then $\limsup_{t\to\infty}\mu_n^{11}(t){\le}\pi_{11}{=}\frac{\bar{\chi}_1 N}{2}\frac{T_w}{m_1d_1e_1-\frac{\bar{\chi}_1 N}{2}(T_{11}-T_w)}$;*
2) *if all nodes apply GENIAL, then $\lim_{t\to\infty}\mu_n^{11}(t){=}\pi_{11}{=}\frac{\bar{\chi}_1 N}{2}\frac{T_w}{m_1d_1e_1-\frac{\bar{\chi}_1 N}{2}(T_{11}-T_w)}$, $\forall n{\in}\mathcal{N}$.*

Proof: First, since $H{=}h{=}1$, from (1) one can derive that $\Pi_{11}{=}\pi_{11}$, whose value is computed as follows. Using (3) and denoting the time required for training as T_{11}, we can derive the mean time required for each session:

$$T_s = T_{11}\Pi_{11} + T_w(1-\Pi_{11}) = \frac{m_1d_1e_1}{\chi_1^{max}}\pi_{11} + T_w(1-\pi_{11})\,. \tag{A.1}$$

As the probability of being selected by the orchestrator as requester is $\frac{1}{N}$, the mean computing cost for a node n when it acts as requester is:

$$c_n^r = \frac{1}{N}m_1d_1e_1\frac{1}{T_s}\pi_{11}\,. \tag{A.2}$$

The mean computing expenditure when the node is a learner c^l is equal to c^r. Thus, imposing the computing constraint it follows that:

$$c_n^{tot} = c_n^r + c_n^l = \frac{2}{N}m_1d_1e_1\frac{1}{T_s}\pi_{11} \le \bar{\chi}_1. \tag{A.3}$$

Solving equation (A.3) by using (A.1) we obtain the Pareto-optimal value:

$$\pi_{11} = \frac{\bar{\chi}_1 N}{2}T_w\frac{1}{m_1d_1e_1 - \frac{\bar{\chi}_1 N}{2}(T_{11}-T_w)}. \tag{A.4}$$

We prove now the first part of the theorem. Consider the scenario where $N{-}1$ nodes employ GENIAL, while node u does not; a node n that uses GENIAL rejects a training request when $\nu_n(t){>}\pi_{11}$; therefore, $\limsup_{t\to\infty}\nu_n(t){\le}\pi_{11}, n{\ne}u$. As GENIAL does not depend on the requester identity, all nodes receive the same amount of service. Thus, $\limsup_{t\to\infty}\mu_n(t){\le}\pi_{11}, n{=}1,\dots,N$. This proves the first part of the theorem.

To prove the second part of the theorem, we define for the generic node n:

$$\begin{aligned}\phi_n(S) &= \frac{\text{No. of accepted requests made by } n \text{ till } S}{S}\\ \psi_n(S) &= \frac{\text{No. of received requests accepted by } n \text{ till } S}{S}\end{aligned} \tag{A.5}$$

with S denoting the number of total training sessions, both accepted and rejected. Denoting $\phi_n{=}\lim_{S\to\infty}\phi_n(S)$, it is possible to write:

$$\phi_n = \lim_{S\to\infty}\frac{\text{No. of accepted requests made by } n \text{ till } S}{S}\cdot \frac{\text{No. of received requests accepted by } n \text{ till } S}{\text{No. of received requests accepted by } n \text{ till } S} = \frac{NRS}{N(N-1)} \tag{A.6}$$

From (A.6) we can see that the NRS converges if and only if ϕ_n converges. Notably, since $H{=}h{=}1$, the probability for a requester that its training request is accepted, hence performed, is precisely equal to the probability that the only node receiving the request accepts, i.e., π_{11}. For this reason, for all nodes, the number of made requests that are accepted must be equal to the total number of accepted received requests. Then, we define the following Lemma.

Lemma 1: $\sum_{n=1}^N(\phi_n(S){-}\psi_n(S)){=}0$

Proof: First, we consider the node n at training session S and we prove that $\phi_n(S){-}\psi_n(S)$ converges to 0.

For $\phi_n(S)$ and $\psi_n(S)$ it holds that:

$$\begin{aligned}\phi_n(S+1) &= \frac{S\phi_n(S)+1_R}{S+1}\\ \psi_n(S+1) &= \frac{S\psi_n(S)+1_L}{S+1}\end{aligned}$$

where

$$1_R = \begin{cases} 1, \text{if request made by node } n\text{' is accepted}\\ 0, \text{else}\end{cases}$$

$$1_L = \begin{cases} 1, \text{if node } n \text{ accepts a received training request}\\ 0, \text{else.}\end{cases}$$

Then, we define

$$\rho(S) = [\phi_1(S) - \psi_1(S), \ldots, \phi_N(S) - \psi_N(S)]^T \tag{A.7}$$

and we write the following recursion on $\rho(S)$ as:

$$\rho(S+1) = \rho(S) + \frac{1}{S+1}(-\rho(S) + \omega(S)) \tag{A.8}$$

where

$$\omega(S) = \begin{cases} -1, \text{if node } n \text{ accepts a received training request;} \\ 1, \text{if request made by node } n \text{ is accepted;} \\ 0, \text{else.} \end{cases}$$

The objective is to prove that the sequence $\{\rho_S\}$ converges to $\rho^*{=}[0, \ldots, 0]^T$ when using GENIAL, as this leads to the proof that $\phi_n(S){-}\psi_n(S)$, and hence $\nu_n(S){-}\mu_n(S)$, converges to 0. To prove this, we utilize the following corollary [2]. ∎

Corollary 1: Consider a sequence $\{\theta(S)\}$ such that:

$$\begin{aligned} &\theta(S+1) = \theta(S) + \lambda(S)\sigma(\omega(S), \theta(S)) \\ &\sum_{S=1}^{\infty} \lambda(S) = \infty \\ &\sum_{S=1}^{\infty} \lambda^2(S) < \infty \end{aligned} \tag{A.9}$$

Define $\bar{\sigma}(\theta){=}E[\sigma(\theta,\omega)]$. Then, if:

a) $(\theta^*{-}\theta)^T \bar{\sigma}(\theta) {\geq} C_1 \left\|\theta^\star{-}\theta\right\|^2$ for some $C_1{>}0$;

b) $E\left[\|\sigma(\theta,\omega)\|^2\right] {\leq} C_2 \left[\left\|\theta^\star{-}\theta\right\|^2 {+}1\right]$ for some $C_2 > 0$;

we have $\lim_{S\to\infty} \theta(S){=}\theta^\star$ with probability 1.

We must prove that $\rho(S)$ converges to $\rho^\star$. With $\lambda(S){=}1/(S+1)$ and $\sigma(\theta,\omega){=}{-}\rho(S){+}\omega(S)$, we verify that (A.8) satisfies (A.9); also, condition b) of Corollary 1 is satisfied considering sufficiently large C_2. We need to prove that condition a) holds, i.e.,

$$(-\rho)^T \bar{\sigma}(\rho) \geq C_1 \| -\rho\|^2. \tag{A.10}$$

At training session S, if a set $\mathcal{A}$ of A out of the N nodes are accepting the training request, we have $\mu_n(S){-}\nu_n(S){>}\delta, n{=}1, \ldots, A$; on the other hand, for the remaining $N{-}A$ nodes rejecting the requests it holds that $\mu_n(S){-}\nu_n(S){<}\delta$, $n{=}A{+}1, \ldots, N$. Thus, for some $\epsilon{>}0$, it holds $\phi_n(S){-}\psi_n(S){>}\epsilon, n{=}1, \ldots, A$ and $\phi_n(S){-}\psi_n(S){<}\epsilon$, $n{=}A{+}1, \ldots, N$.

Considering that a node becomes a requester with probability equal to $1/N$, the probability that a node belonging to $\mathcal{A}$ makes a request and that its request is accepted is equal to $(A{-}1)/[N(N{-}1)]$. On the other hand, the node receives and accepts a training request with probability $(N{-}1)/[N(N{-}1)]$, while $A/[N(N{-}1)]$ defines the probability that a node in the set of the rejecting ones makes a training request and that the latter is accepted. Finally, the probability that it accepts a request is equal to 0. Hence, we can derive that:

$$\begin{aligned} &\bar{\sigma}_n(\rho(S), \omega(S)) \\ &\quad = \begin{cases} -\rho_n(S) + \frac{A-N}{N(N-1)}, & \text{if } n = 1, \ldots, A \\ -\rho_n(S) + \frac{A}{N(N-1)}, & \text{if } n = A+1, \ldots, N \end{cases} \end{aligned}$$

which allow us to write

$$\begin{aligned} (-\rho_S)^T \bar{\sigma}(\rho_S) =& \|\rho(S)\|^2 - \frac{1}{N(N-1)} \sum_{n=1}^{A} (A-N)\rho_n(S) \\ & - \frac{1}{N(N-1)} \sum_{n=A+1}^{N} A\rho_n(S) \\ =& \|\rho(S)\|^2 + \frac{1}{N-1} \sum_{n=1}^{A} \rho_n(S) \\ >& \|\rho(S)\|^2 + \frac{A}{N-1}\epsilon \\ >& \|\rho(S)\|^2. \end{aligned}$$

Therefore, a) is satisfied for $C_1{=}1$, allowing us to use Corollary 1. We have $\lim_{S\to\infty} \rho(S){=}\rho^*$ with probability 1, i.e., $\phi_n(S){-}\psi_n(S)$, and hence $\mu_n(S){-}\nu_n(S)$, converges to 0 for each n.

Considering a node n that employs the GENIAL algorithm, we can derive that $\limsup_{S\to\infty} \nu_n(S){\leq}\pi_{11}$. Also, since $\lim_{S\to\infty} \nu_n(S){-}\mu_n(S){=}0$, $\liminf_{S\to\infty} \nu_n(S){\geq}\pi_{11}$. This is because, if a node n uses the GENIAL algorithm and $\nu_n(S){\leq}\pi_{11}$, it always accepts a training request when $\nu_n(S){-}\mu_n(S){=}0$, thus increasing $\nu_n(S)$. It follows that $\liminf_{S\to\infty} \nu_n(S){\not<}\pi_{11}$, thus $\lim_{S\to\infty} \nu_n(S){=}\pi_{11}$. Since $\nu_n(S){-}\mu_n(S)$ converges to zero, it holds that $\lim_{S\to\infty} \mu_n(S){=}\pi_{11}$.

If both $\nu_n(S)$ and $\mu_n(S)$ converge to π_{11} considering the number of training sessions, they converge also in time, thus we can write $\lim_{t\to\infty} \nu_n(t){=}\lim_{t\to\infty} \mu_n(t){=}\pi_{11} = \frac{\bar{\chi}_1 N}{2} T_w \frac{1}{m_1 d_1 e_1 - \frac{\bar{\chi}_1 N}{2}(T_{11} - T_\omega)}$. ∎

B. Proof of Theorem 2

Theorem 2. *Consider a system of N nodes and K classes, $H{=}h{=}1$, and N_k nodes in class k. Also, the computing capabilities constraints are as follows: $\bar{\chi}_1{<}\bar{\chi}_2{<}\ldots < \bar{\chi}_K$ and $\chi_1^{\max}{<}\chi_2^{\max}{<}\ldots < \chi_K^{\max}$. Then:*

1) *if all nodes, except for node n, employ GENIAL, then $\limsup_{t\to\infty} \mu_n^{kj}(t){\leq}\pi_{kj}$;*
2) *if all nodes apply GENIAL, then $\lim_{t\to\infty} \mu_n^{kj}(t){=}\pi_{kj}$, $\forall n{\in}\mathcal{N}$ and $k, j{=}1, \ldots, K$.*

Proof: Considering the case where nodes are involved in sessions of type (kk), all nodes behave as if they were of the same class, i.e., as they had the same computing constraint $\bar{\chi}_k$ and $\chi_k^{\max}$. From Theorem 2, one can notice that when considering sessions of type (kk) $\nu_n^{kk}(t)$ and $\mu_n^{kk}(t)$ will converge. Therefore, using (8) and (9) to compute the ratio $\frac{\Pi_{kj}}{\Pi_{kk}}$ and, hence, $\frac{\pi_{kj}}{\pi_{kk}}$, we can see that $\nu_n^{kj}(t)$ and $\mu_n^{kj}(t)$ converge for all type session. ∎

C. Proof of Theorem 3

Theorem 3. *Consider a system with N nodes, K classes, $H{>}1$, $h{>}1$ and N_k nodes in class k, $k{=}1, \ldots, K$. As for the computing capabilities constraints, assume that $\bar{\chi}_1{<}\bar{\chi}_2{<}\ldots{<}\bar{\chi}_K$ and $\chi_1^{\max}{<}\chi_2^{\max}{<}\ldots{<}\chi_K^{\max}$. Then:*

1) *if all nodes, except for node* n*, use GENIAL,* $\limsup_{t\to\infty} \mu_n^{kj}(t) \leq \Pi_{kj}$*;*
2) *if all nodes apply GENIAL, then* $\lim_{t\to\infty} \mu_n^{kj}(t) = \Pi_{kj}$, $\forall n \in \mathcal{N}$ *and* $k, j = 1, \ldots, K$.

Proof: In the general case, we have that $H>1$ $h>1$; thus, fixing such quantities, we can prove that $\nu_n^{kk}(t)$ and $\mu_n^{kk}(t)$ converge, by following the same procedure of the proof of Theorem 3 and by scaling Lemma 1.

REFERENCES

[1] Srinivasan et al., *An analytical approach to the study of cooperation in wireless ad hoc networks*, in IEEE Transactions on Wireless Communications, 2005

[2] D. Bertsekas and J.N. Tsitsiklis, *Neuro-dynamic programming*, Athena Scientific, 1996

■