A Cooperative Game Based Allocation for Sharing Data Center Networks

Presenter: Shengkai Shi

March 15 , 2013

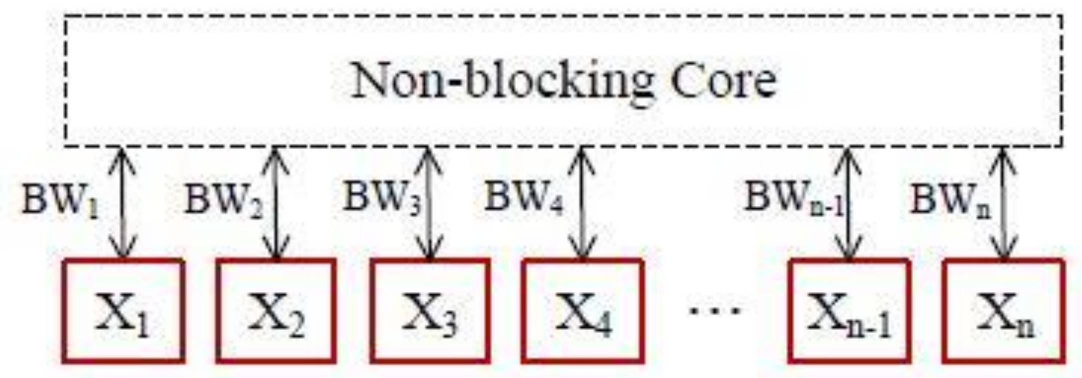
Background

- The network bandwidth is usually shared in a best-effort manner. No bandwidth guarantee.
- Minimum bandwidth guarantee and fair bandwidth guarantee.
- High utilization.
- A conflict of different objectives.

Goal

- Achieve Nash bargaining solution, ensuring minimum bandwidth for each VM-pair and fairness for all VM pairs.
- Design a distributed algorithm to achieve Nash bargaining solution and high utilization.

Data center network model



Data center network model(Cont).

- M servers:

  $M = \{p_1, p_2, ..., p_M\}$

- N VMs:

  $N = \{v_1, v_2, ..., v_N\}$

- Placement matrix $V = (v_{mi})_{M \times N}$:

$$v_{mi} = \begin{cases} 1, & v_i \text{ is on } p_m; \\ 0, & \text{otherwise.} \end{cases}$$

Data center network model(Cont).

- Bandwidth demand matrix $D_N(t) = [D_{ij}(t)]_{N \times N}$:
  $D_{ij}(t)$ denotes the traffic demand from VM v_i to VM v_j at time t.
- Bandwidth allocation strategy $r_N(t) = [r_{ij}(t)]_{N \times N}$:
  $r_{ij}(t)$ denotes the bandwidth allocated for VM v_i to VM v_j at time t.

Data center network model(Cont).

- Base bandwidth:

  - if the bandwidth demand of a VM is lower than its base bandwidth, allocate sufficient bandwidth to satisfy its demand.

  - otherwise, ensure the base bandwidth and set an upper-bounded bandwidth for each VM pair to maintain fairness among VMs.

Data center network model(Cont).

- For VM v_i : $(r_i^I(t), r_i^E(t)), D_i^I(t), D_i^E(t), B_i^I(t), B_i^E(t))$.

  - $r_i^I(t), r_i^E(t)$: total ingress and egress bandwidth allocated.

  $r_j^I(t) = \sum_{i=1}^{N} r_{i,j}(t)$ $r_j^E(t) = \sum_{i=1}^{N} r_{j,i}(t)$

  - $D_i^I(t), D_i^E(t)$: total ingress and egress bandwidth demand.

  $D_j^I(t) = \sum_{i=1}^{N} D_{i,j}(t)$ $D_j^E(t) = \sum_{i=1}^{N} D_{j,i}(t)$

Data center network model(Cont).

- Base bandwidth B_{ij} :

$$B_{i,j} = \min\{B_i^E \frac{B_j^I}{\sum_{D_{ik} \neq 0} B_k^I}, B_j^I \frac{B_i^E}{\sum_{D_{kj} \neq 0} B_k^E}\}$$

- a portion of the egress base bandwidth of VM v_i or a portion of the ingress base bandwidth of VM v_j.

Data center network model(Cont).

- Server p_m : $(C_m^I(t), C_m^E(t))$

  - $C_m^I(t), C_m^E(t)$: the total ingress and egress bandwidth capacity.

  - assumption: $C_m = C_m^I(t) = C_m^E(t)$.

Bandwidth allocation principles

- Minimum bandwidth guarantee

  - poor VM: $D_i < B_i$

  - rich VM : $D_i \geq B_i$

- Fairness

  Fairness in game theory.

- High utilization

Bargaining problem

- Bargaining problems represent situations in which:

  - there is a conflict of interest about agreements.

  - individuals have the possibility of concluding a mutually beneficial agreement.

  - no agreement may be imposed on any individual without his approval.

Bargaining problem

- The strategic or noncooperative model involves explicitly modeling the bargaining process (i.e., the game form).
- Axiomatic approach involves abstracting away the details of the process of bargaining and considers only the set of outcomes or agreements that satisfy ”reasonable” properties.

Nash bargaining framework

- Basic setting:

  - N VMs are players competing for the use of bandwidth.

  - each player has a performance function f_i and a desired initial performance u_i^0.

  - X represents the space of available bandwidth strategies for N VMs. X defined as a convex , nonempty, and compact subset of R^N .

Nash bargaining framework(Cont).

- The initial performance of each player is a minimum guarantee that network must provide the player.

  $u_0 = (u_1^0, \ldots., u_N^0)$

- (X, u_0) is called bargaining problem.
- $G = \{S(X, u_0) \mid X \subset R^N\}$ is the set of achievable performance with respect to the initial performance. $U = \{x \mid x \in X, x_i \geq u_i^0., \forall i\}$

Nash bargaining framework(Cont).

- Pareto efficiency

  A bargaining solution f is Pareto efficient if it does not exist $a \in X$ such that $a \geq f, a_i > f_i, \exists i$

- Symmetry

  X is symmetric with respect to a subset of indices.

  $J \subseteq \{1, ..., N\}, i \in J, j \in J$. If $u_i^0 = u_j^0$, $S(X, u^0)_i = S(X, u^0)_j$

Nash bargaining framework(Cont).

- Linearity

$\phi : R^N \to R^N, \phi(u) = u', u_j' = a_j u_j + b_j, a_j > 0, j = 1, \ldots.., N$

then, $S(\phi(X), \phi(u_0)) = \phi(S(X, u_0))$

- Irrelevant alternatives

$(X, u_0), (X', u_0)$ are two bargaining problems, and

$X' \subseteq X$. If $S(X, u_0) \in X'$, then $S(X, u_0) = S(X', u_0)$.

Nash bargaining framework(Cont).

Definition 1: A mapping $S:G \to R^N$ is said to be a Nash Bargaining Solution(NBS) if:

- $S(X,u_0) \in U, U = \{x \mid x \in X, x_i \geq u_i^0, \forall i\}$
- $S(X,u_0)$ is Pareto efficient.
- $S(X,u_0)$ satisfies the linearity axiom.
- $S(X,u_0)$ satisfies the irrelevant alternatives.
- $S(X,u_0)$ satisfies the symmetry axiom.

Nash bargaining framework(Cont).

Definition 2:

Let $J = \{ j \in \{1....N\} \mid \exists x \in X, x_j > u_j^0 \}$.

We say x is a NBS if the vector x solves the following optimal problem(P_J):

$$\max \prod_{j \in J} (x_j - u_j^0), x \in X$$

- A unique optimal function.

Nash bargaining framework(Cont).

Proposition 1：

The optimal solution of the optimal problem P_J is a unique bargaining solution that satisfies requirements of a NBS.

Proof :

The proof has two steps. First step is to prove that the optimal solution of P_J satisfies the requirements of NBS. Then we can prove any bargaining solution that satisfies requirements of a NBS is equal to the optimal solution of P_J

Nash bargaining framework(Cont).

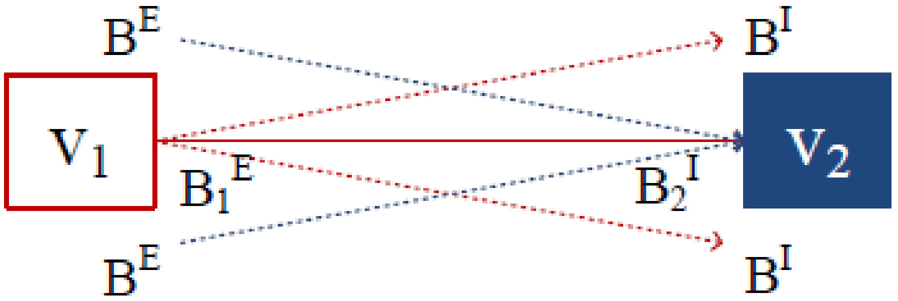
- In [2], proportional fairness is shown to be in fact an NBS.
- Taking the logarithm of the objective, we can derive an equivalent optimization problem:

$$\max \sum_{j \in J} \ln(x_j - u_j^0), x \in X$$

Optimization problem based on NBS

- Define an optimization problem in sharing data center network to achieve NBS.
- For VM v_i: $(r_i^I(t), r_i^E(t)), D_i^I(t), D_i^E(t), B_i^I(t), B_i^E(t))$

  r_{ij} is what we should obtain.



Optimization problem based on NBS(Cont).

- The minimum bandwidth to guarantee L_{ij}:

$$L_{ij} = \min(D_{ij}, B_{ij})$$

- The upper bound for bandwidth allocation from v_i to v_j

$$U_{ij} = \min(C_m, C_n), s.t. v_i \in p_m, v_j \in p_n$$

Optimization problem based on NBS(Cont).

- The joint optimization problem:

$$
\begin{aligned}
\max_{r} \quad & \sum_{j} \sum_{i} \ln(r_{i,j} - L_{i,j}) \\
\text{s.t.} \quad & r_{i,j} \leq U_{i,j}, \ \forall i, j \in \{1, \ldots, N\}, \\
& r_{i,j} \geq L_{i,j}, \ \forall i, j \in \{1, \ldots, N\}, \\
& \sum_{v_i \in V_{p_m}} r_i^I \leq C_m, \ \forall p_m \in \mathcal{M}, \\
& \sum_{v_i \in V_{p_m}} r_i^E \leq C_m, \ \forall p_m \in \mathcal{M},
\end{aligned}
$$

Centralized optimal solution

- Apply the method of Lagrange multipliers, we could get the optimal solution r_{ij}^* :

$$r_{i,j}^* = L_{i,j} + \frac{1}{\sum_{m=1}^{M} \gamma_m^E v_{mi} + \sum_{m=1}^{M} \gamma_m^I v_{mj}}$$

- This convex optimization has 2(M+N) constraints. Computational complexity may increase significantly as the number of VMs and servers scales up.

Dual-based decomposition

- Create a new optimization with the same optimal solution:

$$
\begin{aligned}
\min_{r} \quad & P(r) = -\sum_{j=1}^{N}\sum_{i=1}^{N} \ln(r_{i,j} - L_{i,j}) \\
\text{s.t.} \quad & r_{i,j} \le U_{i,j},\ \forall i, j \in \{1, \ldots, N\}, \\
& r_{i,j} \ge L_{i,j},\ \forall i, j \in \{1, \ldots, N\}, \\
& \sum_{v_i \in V_{p_m}} r_i^I \le C_m,\ \forall p_m \in \mathcal{M}, \\
& \sum_{v_i \in V_{p_m}} r_i^E \le C_m,\ \forall p_m \in \mathcal{M}.
\end{aligned}
$$

Dual-based decomposition(Cont).

- Discuss the general situation: $L_{i,j} < r_{i,j} < U_{i,j}$
- Lagrangian associated with this problem:

$$\mathcal{L}(r,\gamma^I,\gamma^E) = -\sum_{j=1}^{N}\sum_{i=1}^{N}\ln(r_{i,j}-L_{i,j})$$
$$+\sum_{m=1}^{M}\gamma_m^I((V\cdot r^I)_m - C_m) + \sum_{m=1}^{M}\gamma_m^E((V\cdot r^E)_m - C_m)$$

- The Lagrange dual function:

$$d(\gamma^I,\gamma^E) = \inf_{r\in\mathcal{R}^{N\times N}} \mathcal{L}(r,\gamma^I,\gamma^E)$$

Dual-based decomposition(Cont).

- Slater's condition holds. There is no duality gap. The dual problem corresponding to the optimization problem

$$\max_{\gamma^I, \gamma^E \in \mathcal{R}^M} d(\gamma^I, \gamma^E) = \mathcal{L}(r^*, \gamma^I, \gamma^E)$$

Gradient projection method

- Apply gradient projection method to converge to the optimal γ^E and γ^I.
- Define a recursion:

$$\gamma_m^{(k+1)} = \max(0, \gamma_m^{(k)} + \xi \frac{\partial d}{\partial \gamma_m}), \forall m \in \{1, 2, \ldots, M\}$$

Gradient projection method(Cont).

Theorem 3. *For the recursive sequence* $\{\gamma^{I(k)}\}$, *if* $\gamma^{I(0)} \in \mathcal{R}^{+M}$ *and* $\xi \in (0, \frac{2}{K}]$, *then* $\{\gamma^{I(k)}\}$ *converges, thus*

$$\lim_{k\to\infty} \gamma^{I(k)} = \gamma^{I*} \in \overline{\Gamma}, \tag{20}$$

where K *is the Lipschitz constant [15] of the dual function in Eq. (18), such that*

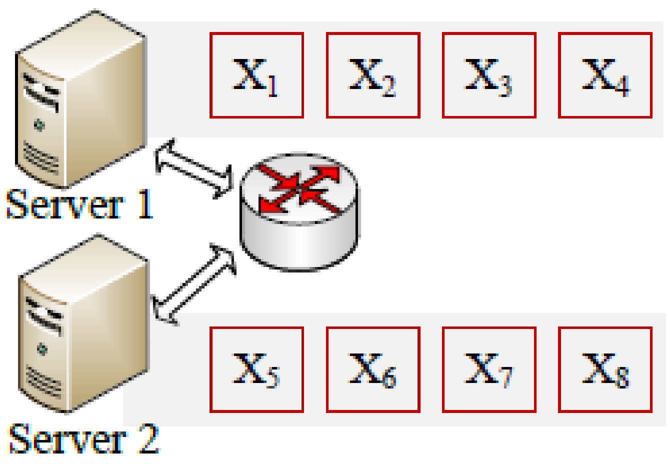
$$K = \sqrt{M}\sum_{j=1}^{N}\sum_{i=1}^{N}(U_{i,j} - L_{i,j})^2. \tag{21}$$

Distributed cooperative algorithm

for all $r_{i,j}$ **do**
  Initialize $L_{i,j}$ and $U_{i,j}$ by Eq. (2) and (9), respectively;
  $r_{i,j} = L_{i,j}$;
end for
while steps $<$ S **do**
  for all server p_m **do**
    Update $r^E_{p_m} = \sum_i v_{mi} r^E_i$, $r^I_{p_m} = \sum_i v_{mi} r^I_i$;
    $\gamma^E_m = \max(0, \gamma^E_m - \xi(C_m - r^E_{p_m}))$ (Eq. 19);
    $\gamma^I_m = \max(0, \gamma^I_m - \xi(C_m - r^I_{p_m}))$ (Eq. 19);
  end for
  for all $r_{i,j}$ **do**
    if $\frac{1}{\gamma^E + \gamma^I} \leq U_{i,,j} - L_{i,j}$ **then**
      $r_{i,j} = U_{i,j}$ (Eq. 10);
    else
      $r_{i,j} = L_{i,j} + \frac{1}{\gamma^E + \gamma^I}$
    end if
  end for
  steps++;
end while

Simulation

- Simulation scenario:



Simulation results

- Bandwidth allocation for VMs：

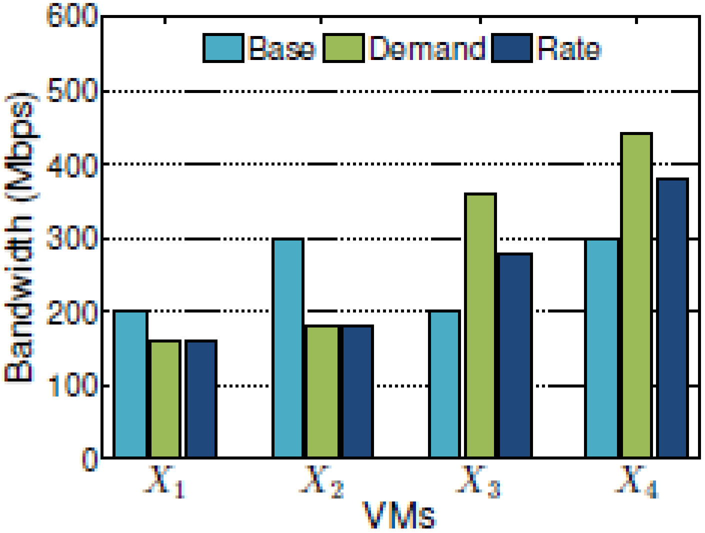


Fig. 4: Bandwidth allocation to VMs on server 1 with different demands and base bandwidths.

Simulation results(Cont).

- Rate of convergence:

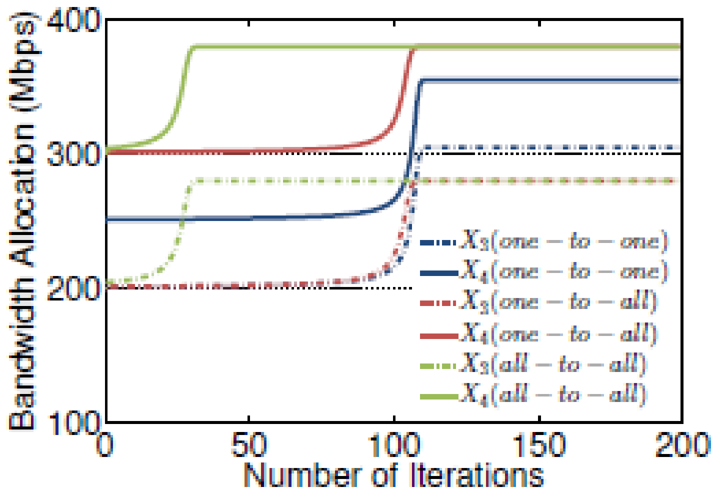


Fig. 6: Rates of VM on server 1 with increasing number of iterations.

Simulation results(Cont).

- Varying bandwidth demand of VMs:

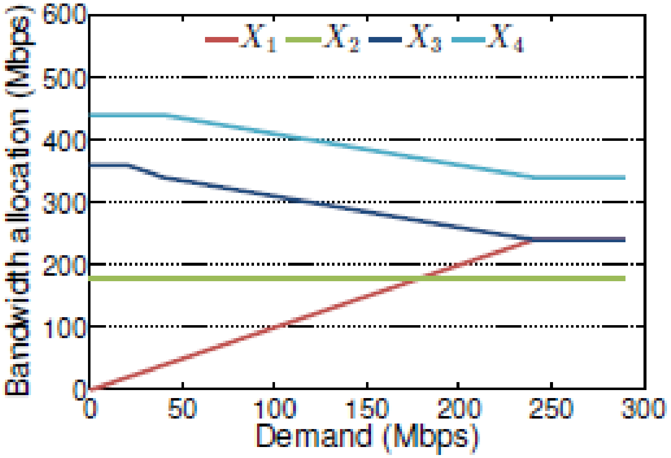


Fig. 5: Bandwidth allocation to VMs on server 1 with increasing demand of X_1 for all-to-all communication patterns.

References

[1] Jian Guo, Fangming Liu, Dan Zeng, John C.S. Lui, Hai Jin, "A Cooperative Game Based Allocation for Sharing Data Center Networks", in Proc. of IEEE INFOCOM , April, Italy, 2013.

[2] H. Yaıche, R. Mazumdar, and C. Rosenberg, “A game theoretic framework for bandwidth allocation and pricing in broadband networks,” IEEE/ACM Transactions on Networking (TON), vol. 8, no. 5, pp. 667–678, 2000.

Thanks!