Stochastic Games for Cooperative Network Routing and Epidemic Spread

Lorenzo Maggi
Eurecom
Mobile Communications Department
BP193, F-06560 Sophia Antipolis, France
Email: lorenzo.maggi@eurecom.fr

Konstantin Avrachenkov
INRIA
BP95, 06902 Sophia Antipolis, France
Email: k.avrachenkov@sophia.inria.fr

Laura Cottatellucci
Eurecom
Mobile Communications Department
BP193, F-06560 Sophia Antipolis, France
Email: laura.cottatellucci@eurecom.fr

***Abstract*—We consider a system where several providers share the same network and control the routing in disjoint sets of nodes. They provide connection toward a unique server (destination) to their customers. Our objective is to facilitate the design the available network links and their costs such that all the network providers are interested in cooperating and none of them withdraw from the coalition. More specifically, we establish the framework of a coalition game by providing an algorithm to compute the transferable coalition values. As by-product, we apply the proposed algorithm to two-player games both in networks subject to hacker attacks and in epidemic networks.**

I. Introduction

Sharing resources among competitive operators is a fundamental issue in 4G wireless systems. Cooperation enables a better exploitation of the resources and promises higher revenues to network providers. However, cooperation among competitive entities is complicated by the sensitive issue of conflicting interests. Thus, it becomes imperative to motivate and guarantee a fair cooperation among these entities. This can be achieved by a careful distribution of the costs or the incremental revenues obtained by cooperating. Coalition games offer a suitable theoretical framework to address this problem.

Several providers share a network to provide connection towards a unique common destination to their customers. We provide a framework of a coalition game to facilitate the design of the available network links and their costs such that there exists an optimum routing strategy and a cost sharing satisfying all the subsets of providers. More specifically, we provide algorithms to compute the coalition values, i.e. the minimum costs that each coalition can ensure for itself. The proposed algorithm is based on some results for two-player zero-sum stochastic games with perfect information in [1].

It is worth noticing that the analyzed problem differs substantially from the noncooperative routing games thoroughly studied in literature (for additional details see e.g. [7] and references therein). At the best of the authors' knowledge, this work is the first one applying coalition games to determine an optimum routing solution and cost allocation in a shared network.

II. Routing model

We consider a network consisting of a set of nodes $V = \{1, \dots, N\}$. M service providers share the network to offer their customers connection toward a single destination node N. The customers's traffic is injected in the network at $n \le N-1$ nodes, called sources, located in nodes $\mathcal{T} = \{i_1, \dots, i_n\} \subseteq V/\{N\}$. There is only one destination, in node N. We assume that all the sources transmit at the same rate the packets of a provider k, for all possible k. Let $c_k(i,j) > 0$ represent the cost per unit time that provider k has to sustain to convey its own packets, sent by any of the sources in $\mathcal{T}$, through the link $i \to j$.

The k-th service provider controls the routing, i.e. the activation of outgoing links, in the set of nodes V_k. We suppose that a node is controlled at most by one provider, i.e., $V_i \cap V_j = \emptyset,\ \forall\, i \neq j$ and $\bigcup_i V_i \subseteq V$. Each node i is assigned a subset $\alpha_i \subseteq V$, such that the *directed* link $i \to j$ can be activated if and only if $j \in \alpha_i$. In the generic node $i \in V_k$ controlled by provider k, provider k himself can assign a probability distribution $\mathbf{f}_k$ to the each node $j \in \alpha_i$ such that the probability that the network link (i,j) is utilized for routing is $\mathbf{f}_k(i,j)$ at *any* routing decision moment. The destination node is a "sink", and it does not route the incoming packets to any of the other nodes. We remark that all the nodes $\{1, \dots, N-1\}$, included the sources, serve as routing nodes.

Let $\mathbf{\Phi}_\beta^{(k)}$, with $\beta \in [0;1]$, be a N-by-1 vector whose i-th component is the expected β-discounted sum of costs:

$$E_{\mathbf{f}_1,\dots,\mathbf{f}_M}\Big[\sum_{t\ge 0} \beta^t c_k(i_t, i_{t+1})\Big], \text{ with } i_0 = i,$$

where i_t is the t-th node crossed by the packets. It is worth noticing that, for $\beta = 1$, $\mathbf{\Phi}_1^{(k)}$ is the undiscounted sum and its l-th component, with $l \in \mathcal{T}$, is the cost per unit time that provider k incurs for the stream of packets going from the l-th source to the destination.

III. Routing coalition game

Let $\mathcal{M} = \{1, \dots, M\}$ be the grand coalition of service providers. We assume that the providers belonging to a generic coalition $\mathcal{C} \subseteq \mathcal{M}$ can stipulate binding agreements among them to enforce the optimum strategy for the coalition and distribute the costs among themselves.

Let $\mathbf{F}_\mathcal{C}$ be the set of strategies available to coalition $\mathcal{C} \subseteq \mathcal{M}$. It is easy to show that $\mathbf{F}_\mathcal{C}$ is the Cartesian product of the strategies available to all the members of $\mathcal{C}$, i.e. $\mathbf{F}_\mathcal{C} = \times_{k\in\mathcal{C}} \mathbf{F}_k$, and the set of strategies $\mathbf{F}_\mathcal{C}$ is dubbed *not correlated*. Moreover, thanks to available results on stochastic games (see Appendix

A), we can focus only on *pure* (deterministic) strategies. Let $\mathbf{F}_{\mathcal{C}}$ be the set of pure strategies for $\mathcal{C}$, i.e.

$$\mathbf{F}_{\mathcal{C}} = \Big\{ \mathbf{f}_k : \ \{k\} \in \mathcal{C}; \ \forall\, i \in V_k, \ \exists\, j : \mathbf{f}_k(i,j) = 1 \Big\}.$$

Let us define, for any $\beta \in [0;1]$, the expected β-discounted sum

$$\mathbf{\Phi}_{\beta}^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) = \sum_{\{k\}\in\mathcal{C}} \mathbf{\Phi}_{\beta}^{(k)}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$$

Let $\mathbf{e}_{\mathcal{T}}$ be a N-by-1 vector containing 1's in correspondance of the sources and 0's otherwise. In this paper we are interested in the case $\beta = 1$, since the quantity $\mathbf{e}_{\mathcal{T}}^T \mathbf{\Phi}_{\beta=1}^{(\mathcal{C})}$ is the total cost per unit time that $\mathcal{C}$ incurs to sustain its $n|\mathcal{C}|$ information streams. The minimum cost $v(\mathcal{C})$ that coalition $\mathcal{C}$ can ensure for itself is

$$v(\mathcal{C}) = \min_{\mathbf{f}_{\mathcal{C}} \in \mathbf{F}_{\mathcal{C}}} \ \max_{\mathbf{f}_{\mathcal{M}/\mathcal{C}} \in \mathbf{F}_{\mathcal{M}/\mathcal{C}}} \mathbf{e}_{\mathcal{T}}^T \mathbf{\Phi}_1^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}). \qquad (1)$$

Under the transferable utility (TU) condition, we suppose that $v(\mathcal{C})$ can be partitioned among the providers of $\mathcal{C}$ in any manner, thanks to a binding agreement among its members. We can say that $v(\mathcal{C})$ is the minmax value of a zero-sum game between the coalition $\mathcal{C}$ and the rest of the providers $\mathcal{M}/\mathcal{C}$, who are willing to "punish" the coalition $\mathcal{C}$. The formulation of this conflict among coalitions as a two-player stochastic game with perfect information is available in Appendix B.
The overall optimum global routing strategy $\mathbf{F}^{\text{o}}$ satifies

$$v(\mathcal{M}) = \mathbf{e}_{\mathcal{T}}^T \mathbf{\Phi}_1^{(\mathcal{M})}(\mathbf{F}^{\text{o}}) = \min_{\mathbf{f}_{\mathcal{M}} \in \mathcal{F}_{\mathcal{M}}} \mathbf{e}_{\mathcal{T}}^T \mathbf{\Phi}_1^{(\mathcal{M})}(\mathbf{f}_{\mathcal{M}})$$

where $\mathbf{F}_{\mathcal{M}}$ is the set of strategies available to the grand coalition $\mathcal{M}$. It is easy to see that the superadditivity property of the characteristic function v:

$$v(\mathcal{C}_1) + v(\mathcal{C}_2) \geq v(\mathcal{C}_1 \cup \mathcal{C}_2), \ \forall\, \mathcal{C}_1, \mathcal{C}_2 \subset \mathcal{M}, \ \mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$$

holds directly from the minmax definition (1) of v.

A. Algorithm for computing coalition values

The values $v(\mathcal{C})$ may be *infinite*. In Appendix A it is shown that $v(\mathcal{C})$ is the value of the game at Nash equilibrium. If $v(\mathcal{C}) = +\infty$, then the optimal strategies for the players, i.e. the strategies at Nash equilibrium, impede at least one source-destination path by causing a loop in the network. In practice, $v(\mathcal{C}) = +\infty$ is not the cost that coalition $\mathcal{C}$ has to bear; anyway, it shows well that any service provider cannot accept to lose its own packets.
The theory of stochastic games provides an approach to *avoid infinities in the computation of coalition values*. The details are illustrated in Lemma A.8, Appendix A-A. The idea is to compute the optimal strategies $(\mathbf{f}_{\mathcal{M}/\mathcal{C}}^*, \mathbf{f}_{\mathcal{C}}^*)$, for coalitions $\mathcal{M}/\mathcal{C}$ and $\mathcal{C}$ respectively, for *all* the discount factors sufficiently close to 1. Then, we adopt the strategy that is still optimal in the limit for $\beta \to 1$.

In the following, we illustrate the proposed approach. Fix a pure strategy $\mathbf{f}_{\mathcal{M}/\mathcal{C}}$ for coalition $\mathcal{M}/\mathcal{C}$. We say that the pure strategy $\mathbf{f}_{\mathcal{C}}'$ is an improvement for coalition $\mathcal{C}$ with respect to $\mathbf{f}_{\mathcal{C}}'$ for the discount factor β iff

$$\mathbf{\Phi}_{\beta}^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}') \leq \mathbf{\Phi}_{\beta}^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$$

where the relation $\leq$ is component-wise and $<$ is valid for at least one component. Let $\Gamma_{\mathcal{M}/\mathcal{C}}(\bar{\mathbf{f}}_{\mathcal{C}})$ be the optimization problem that $\mathcal{M}/\mathcal{C}$ faces when $\mathcal{C}$ fixes its own strategy $\bar{\mathbf{f}}_{\mathcal{C}}$. Then, the optimum strategy for $\mathcal{M}/\mathcal{C}$ in $\Gamma_{\mathcal{M}/\mathcal{C}}(\bar{\mathbf{f}}_{\mathcal{C}})$ maximizes $\mathbf{\Phi}_{\beta}^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \bar{\mathbf{f}}_{\mathcal{C}})$ component-wisely.

Algorithm 1.

1) *Pick a pure routing strategy $f_{\mathcal{C}}$ for coalition $\mathcal{C}$.*
2) *Find the best strategy $f_{\mathcal{M}/\mathcal{C}}$ for coalition $\mathcal{M}/\mathcal{C}$ in the optimization problem $\Gamma_{\mathcal{M}/\mathcal{C}}(f_{\mathcal{C}})$, for all the discount factors close enough to 1.*
3) *Find the first node controlled by coalition $\mathcal{C}$ in which a change of strategy $f_{\mathcal{C}}'$ is a benefit for coalition $\mathcal{C}$ for all the discount factors close enough to 1. If it does not exists, then set $(f_{\mathcal{M}/\mathcal{C}}^*, f_{\mathcal{C}}^*) := (f_{\mathcal{M}/\mathcal{C}}, f_{\mathcal{C}})$ and go to step 4. Otherwise, set $f_{\mathcal{C}} := f_{\mathcal{C}}'$ and go to step 2.*
4) *If $\lim_{\beta\to 1} \mathbf{e}_{\mathcal{T}}^T \mathbf{\Phi}_{\beta}^{(\mathcal{C})}(f_{\mathcal{M}/\mathcal{C}}^*, f_{\mathcal{C}}^*) = l < +\infty$ then set $v(\mathcal{C}) = l$. Otherwise, set $v(\mathcal{C}) = +\infty$.*

We remark that the optimal strategy in step 2 and the strategy refinement in step 3 are found with the help of simplex tableaux in the non-archimedean ordered field $F(\mathbb{R})$ of rational functions with real polynomial coefficients (for all details, see [1]).

B. Transient case

Suppose now that the following assumption holds.

Assumption 1. *For any couple of pure strategies $(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$ for $\mathcal{M}/\mathcal{C}$ and $\mathcal{C}$ respectively, and for all $i \in V$, there exists a path*[1] *$\tau_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$ of finite length*[2] *$L_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$ and without loops linking node i to the destination node N.*

The following result shows that the assumption above ensures $\mathbf{\Phi}_1^{(\mathcal{C})}$ to be finite, for any couple of strategies.

Proposition III.1. *Suppose that assumption 1 holds. Then, for all the pure strategies $\mathbf{f}_{\mathcal{M}/\mathcal{C}} \in \mathbf{F}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}} \in \mathbf{F}_{\mathcal{C}}$:*

(i) *the path $\tau_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$ is unique;*
(ii) $\mathbf{\Phi}_1^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) < +\infty$.

Proof: Let $\tau_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) = \{i_0 = i, i_1, \ldots, i_{L_i} = N\}$ be the nodes crossed by the path τ_i when $\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}$ are fixed. If there existed more than one path linking two nodes then there would exist at least one node in which more than one arc go out of it. This is impossible since the strategies are pure. Then, *(i)* is proved. Therefore, we can say that

$$\begin{cases} p_t(j|i_0 = i, \mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) = \mathbb{1}(j = i_t), & \forall\, t \in [1; L_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})] \\ p_t(j|i_0 = i, \mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) = 0, & \forall\, t > L_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) \end{cases}$$

where $p_t(j|i_0)$ is the probability that the t-th node crossed by the packets starting in node i_0 is j. Thus, $\forall\, i \in V$, the i-th component of $\Phi_1^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$ is bounded by

$$L_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) \, |\, \mathcal{C} \,| \, \max_{k,i,j} c_k(i,j) < +\infty$$

■

[1]a path is a sequence of connected nodes
[2]the length of the path is the number of edges that it is composed of.

1) Adapted algorithm for finite coalition values: If Assumption 1 holds, then the algorithm 1 can be adapted as follows (see Lemma A.4).

Algorithm 2.

1) *Pick a pure routing strategy $f_{\mathcal{C}}$ for coalition $\mathcal{C}$.*
2) *Find the best strategy $f_{\mathcal{M}/\mathcal{C}}$ for coalition $\mathcal{M}/\mathcal{C}$ in the optimization problem $\Gamma_{\mathcal{M}/\mathcal{C}}(f_{\mathcal{C}})$, for $\beta = 1$.*
3) *Find the first node controlled by coalition $\mathcal{C}$ in which a change of strategy $f'_{\mathcal{C}}$ is a benefit for coalition $\mathcal{C}$, for $\beta = 1$. If it does not exists, then set $(f^*_{\mathcal{M}/\mathcal{C}}, f^*_{\mathcal{C}}) := (f_{\mathcal{M}/\mathcal{C}}, f_{\mathcal{C}})$ and go to step 4. Otherwise, set $f_{\mathcal{C}} := f'_{\mathcal{C}}$ and go to step 2.*
4) *Set $v(\mathcal{C}) = \mathbf{e}_T^T \mathbf{\Phi}_1^{(\mathcal{C})}(f^*_{\mathcal{M}/\mathcal{C}}, f^*_{\mathcal{C}})$.*

We remark that the algorithm 2 is analogous to the one described by Raghavan and Syed in [6] when $\beta = 1$ and restricted to the transient case, with the difference that in step 2 the search is not necessarily lexicographic for coalition $\mathcal{M}/\mathcal{C}$. Indeed, at each iteration $\mathcal{M}/\mathcal{C}$ is allowed to find its own temporarily optimal strategy with *any* Markov Decision Process solving method.

IV. Network design

The main contribution of this paper consists in describing how to compute the coalition values, and the network design is not our purpose. Nevertheless, we suggest which steps could be followed in this direction.

An eventual network designer should aim at devising both the routing decisions α_i available to each provider in each node $i \in V$ and the cost of the links $c_k(i,j)$, in order to ensure that each coalition of providers has an interest in not deviating from the global optimum policy $\mathbf{F}^o$. Formally, a network designer should ensure the non-emptiness of the *core* of the TU (transferable utility) coalition game (M, v), i.e. that set of cost $Co(v) = \{g_1, \dots, g_M\} \in \mathbb{R}^M$ that providers can share among themselves through binding agreements, such that

$$\begin{cases} \sum_{k=1}^M g_k = v(\mathcal{M}) \\ \sum_{\{k\} \in \mathcal{C}} g_k \leq v(\mathcal{C}), \quad \forall \mathcal{C} \subset \mathcal{M}. \end{cases}$$

We see from the former equation that the core is globally *efficient* for the network and from the latter that it is also *stable* with respect to the formation of greedy coalitions.

V. Hacker-Provider routing game

The routing game with just two players described in section II can also be re-interpreted in the framework of the conflicts between one service provider and one hacker.
There is a set $V_1 \subseteq V$ of vulnerable nodes, where the routing control may be got hold by a hacker. V_0 is the set of nodes in which the routing is handled by a service provider. The set $V_2 = V_0/V_1$ is the set of unattackable nodes among the ones controlled by the service provider. Each link $i \to j$ is assigned $c(i,j) > 0$, that in this case can be also interpreted as a *delay*, i.e. the time that a packet of provider k spends to go from node i to node j. In such a case, let us assume that the nodes are capable to re-direct all the incoming packets as soon as they receive them, without any additional delay due to the buffering. The service provider here wants to find the routing rule that *jointly* minimizes the packet delay $\mathbf{\Phi}_1$ for all the sources; conversely, the hacker wants to slow down the network.

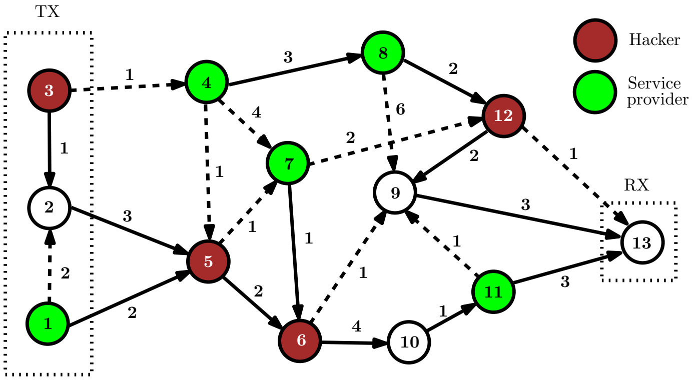

Figure 1. Nash equilibrium in routing game. The continuous arrows are the activated links. The costs are specified next to each arrow. Green and red nodes are controlled by the service provider and by the hacker, respectively. In white nodes there are no routing choice.

As in section II, there may be some couple of strategies for the two players for which there exist loops in the network, that cause the packet delay from some sources to be infinite. Note that the hacker can also disrupt some nodes, by forcing a loop on them. Hence, here we deal with the general case of undiscounted stochastic games described in Appendix A-A. The undiscounted optimal strategies can be computed by the algorithm 1, in which player 1 is now the hacker which controls nodes V_1, and player 2 is the service provider, which controls nodes V_2.

Note that in this case, in contrast to the coalition game, we are more interested in the computation of the optimal strategies, and not in the value of the game at the Nash equilibrium. Indeed, the optimal strategy for the service provider is the pure routing policy it should adopt in order to minimize the source-wise packet delay in the worst case. For Lemma A.3, the worst situation for the provider is when the hacker is able to control all the vulnerable nodes V_1 and has at its disposal as many routing policies as possible. Note that the optimal strategies for both players are pure, i.e. the routing policy is deterministic in each node.

An example of optimal strategies for both players in a delay routing game is shown in Figure 1.

VI. Natural disaster

Let us reformulate the model described in section V, where player 1 is now a natural agent that can put out of order some nodes $V_1 \subset V$ of the network, independently of the routing action taken by the service provider in such nodes. This addresses the practical situation in which nodes V_1 are located in areas subject to catastrophic natural phenomena. It is straightforward to see that the computation of the optimal strategies for the service provider boils down to the calculation of a Markov Decision Process uniform optimal solution (see

[3]), in which the set of nodes of interest is reduced to V_2, that is the collection of nodes controlled by the service provider.

VII. Epidemic network

In this section we model an epidemic network with N nodes; $N-1$ possibly infected individuals are located in nodes $\{1, \ldots, N-1\}$ respectively. Each individual can infect, with some probability, only one among a subset of other individuals in its neighborhood. There is a probability $\boldsymbol{\mu}_i$ that the infection process starts from the i-th individual. The infection spread terminates when the virus reaches the healer, located in node N. Hence, there is a probability $\boldsymbol{\mu}_N$ that the epidemic spread is averted. There are two player: player 2, the "good" one, wants to design and force the connections among the individuals such that the lowest expected number of individuals are infected, while player 1 has the opposite goal. The assumption of perfect information still holds, i.e. the set of nodes in which player 1 and player 2 have more than one action available are disjoint.

The formulation of the problem is analogous to the two-player game described in section II, in which the cost of the link (i, j) is 1 for all nodes i, j. The nodes are substituted by the individuals, the destination with the healer, the sources become the first infected entity, the packet routing is replaced by the virus transmission. In this context, we wish $\boldsymbol{\mu}^T \boldsymbol{\Phi}_1$ to represent the average number of infected individuals. Therefore, for each couple of routing strategies, *no loops* in the network are allowed, i.e. we suppose that the Assumption 1 holds. Hence, thanks to Propostion III.1, for every couple of pure stationary strategies $(\mathbf{f}, \mathbf{g})$, $\boldsymbol{\mu}^T \boldsymbol{\Phi}_1(\mathbf{f}, \mathbf{g})$ is actually the expected number of infected individuals.

Thanks to Corollary A.1, we can use the algorithm 2 to find the optimal strategy for the "good" player, who is interested in minimizing the objective function $\boldsymbol{\mu}^T \boldsymbol{\Phi}_1(\mathbf{f}, \mathbf{g})$. If $(\mathbf{f}^*, \mathbf{g}^*)$ are the undiscounted optimal strategies, then the value $\boldsymbol{\mu}^T \boldsymbol{\Phi}_1(\mathbf{f}^*, \mathbf{g}^*)$ is the most pessimistic estimate for player 2 for the expected number of infected individuals.

VIII. Conclusions

Several providers share the same network and control the routing in disjoint sets of nodes. There are several information sources and one destination. By using the framework of stochastic games, we provided algorithms to compute the minimum costs that each coalition of providers can ensure for itself. This helps the optimum design of a network, which should guarantee the existence of an efficient and stable costs partition among the providers. We also modeled situations in which there are two players with conflicting interests, like a hacker against a service provider, or in which a service provider wants to reduce the damages to the network caused by a natural disaster. An epidemic spread network model was shown as well. From a theoretical perspective, we extended some results on uniform optimal strategies in stochastic game to the case of undiscounted criterion.

Acknowledgment

This reasearch was supported by "Agence Nationale de la Recherche" with reference ANR-09-VERS-001 and by the European research project SAPHYRE, partly funded by the European Union under its FP7 ICT Objective 1.1 - The Network of the Future.

Appendix A
Stochastic games

In a two-player stochastic game Γ we have a set of states $S = \{s_1, s_2, \ldots, s_N\}$, and for each state s the set of actions available to the *i*-th player is called $A^{(i)}(s)$, $i = 1, 2$. Under the zero-sum assumption, each triple (s, a_1, a_2) with $a_1 \in A^{(1)}(s)$, $a_2 \in A^{(2)}(s)$ is assigned an immediate reward $r(s, a_1, a_2)$ for player 1, $-r(s, a_1, a_2)$ for player 2, and a transition probability distribution $p(.|s, a_1, a_2)$ on S.

A stationary strategy $\mathbf{u} \in \mathbf{U}_S$ for the *i*-th player determines the probability $u(a|s)$ that in state s player *i* chooses the actions $a \in [a_1^{(i)}, \ldots, a_{m_i(s)}^{(i)}]$.
We assume that both the number of states and the overall number of available actions are finite.
Let $p(s'|s, \mathbf{f}, \mathbf{g})$ and $r(s, \mathbf{f}, \mathbf{g})$ be the expectation with respect to the stationary strategies $(\mathbf{f}, \mathbf{g})$ of $p(s'|s, a_1, a_2)$ and of $r(s, a_1, a_2)$, respectively.

In this paper we consider stochastic games with perfect information, i.e. in each state *at most* one player has more than one action available. Let $S_1 = \{s_1, \ldots, s_{t_1}\}$ be the set of states controlled by player 1 and $S_2 = \{s_{t_1+1}, \ldots, s_{t_1+t_2}\}$ be the set controlled by player 2, with $t_1 + t_2 \leq N$. Therefore, $r(s, \mathbf{f}, \mathbf{g}) = r(s, \mathbf{f})$ if $s \in S_1$ or $r(s, \mathbf{f}, \mathbf{g}) = r(s, \mathbf{g})$ if $s \in S_2$; the same simplification can be carried out for the transition probabilities.
Let $\boldsymbol{\Phi}_\beta(\mathbf{f}, \mathbf{g})$ be a column vector of length N defined as

$$\boldsymbol{\Phi}_\beta(\mathbf{f}, \mathbf{g}) = \sum_{t=0}^{\infty} \beta^t \mathbf{P}^t(\mathbf{f}, \mathbf{g}) \mathbf{r}(\mathbf{f}, \mathbf{g}) \tag{2}$$

such that its i-th component equals the expected β-discounted reward when the initial state of the stochastic game is s_i. In (2), $\beta \in [0; 1)$ is the discount factor, $\mathbf{P}(\mathbf{f}, \mathbf{g})$ and $\mathbf{r}(\mathbf{f}, \mathbf{g})$ are the *N*-by-*N* transition probability matrix and the *N*-by-1 average reward vector associated to the couple of strategies $(\mathbf{f}, \mathbf{g})$, respectively. If $\beta = 1$, then the reward is called **undiscounted**.
Let ρ be such that $\beta(1+\rho) = 1$. Note that if $\beta \uparrow 1$, then $\rho \downarrow 0$.
Let us give some definitions useful for our purpose.

Definition 1. *The β-discounted value of the game Γ is such that*

$$\boldsymbol{\Phi}_\beta(\Gamma) = \sup_{\mathbf{f}} \inf_{\mathbf{g}} \boldsymbol{\Phi}_\beta(\mathbf{f}, \mathbf{g}) = \inf_{\mathbf{g}} \sup_{\mathbf{f}} \boldsymbol{\Phi}_\beta(\mathbf{f}, \mathbf{g}). \tag{3}$$

An optimal strategy $\mathbf{f}_\beta^*$ ($\mathbf{g}_\beta^*$) for pl. 1 (2) assures to him a reward which is at least (at most) $\boldsymbol{\Phi}_\beta(\Gamma)$.

Corollary A.1. *The optimal strategies $\mathbf{f}_\beta^*$, $\mathbf{g}_\beta^*$ are also optimal when the scalar objective $\mathbf{a}^T \boldsymbol{\Phi}_\beta$ is considered, where $\mathbf{a}_i \geq 0,\ \forall\, i = 1, \ldots, N$.*

Definition 2. *A stationary strategy $\mathbf{h}$ is said to be **uniformly discount optimal** for a player if $\mathbf{h}$ is optimal for every β close enough to 1.*

A strategy is pure if the action choice is deterministic in each state. The following Theorem (see [2]) ensures the existence of such strategies among the optimal ones.

Theorem A.2. *For a stochastic game with perfect information, both players possess uniform discount optimal pure stationary strategies* $(\mathbf{f}^*, \mathbf{g}^*)$ *and moreover* $\mathbf{\Phi}_\beta(\Gamma) = \mathbf{\Phi}_\beta(\mathbf{f}^*, \mathbf{g}^*)$.

Let s_t be a state controlled by player $i = 1, 2$ and $X \subset A_i(s_t)$. Let us call Γ^t_X the stochastic game which is equivalent to Γ except in state s_t, where player i has only the actions X available. In the next sections, we make use of the following result in [1].

Lemma A.3. *Let* $i = 1, 2$ *and* $s_t \in S_i$, $X \subset A_i(s_t)$, $Y \subset A_i(s_t)$, $X \cap Y = \emptyset$. *Then* $\mathbf{\Phi}^*_\rho(\Gamma^t_{X\cup Y}) \in F(\mathbb{R})$, *which is the uniform value of the game* $\Gamma^t_{X\cup Y}$, *equals*

$$\mathbf{\Phi}^*_\rho(\Gamma^t_{X\cup Y}) =_l \max\nolimits_l \{\mathbf{\Phi}^*_\rho(\Gamma^t_X), \mathbf{\Phi}^*_\rho(\Gamma^t_Y)\} \qquad \text{if} \quad i = 1$$
$$\mathbf{\Phi}^*_\rho(\Gamma^t_{X\cup Y}) =_l \min\nolimits_l \{\mathbf{\Phi}^*_\rho(\Gamma^t_X), \mathbf{\Phi}^*_\rho(\Gamma^t_Y)\} \qquad \text{if} \quad i = 2.$$

Let us introduce another special class of stochastic games.

Definition 3. *Let* $p_t(.|s)$ *be the transition probability from state* s *after* t *steps. A stochastic game* ***transient*** *if*

$$\sum_{t=0}^{\infty} \sum_{s' \in S} p_t(s'|s, \mathbf{f}, \mathbf{g}) < +\infty \qquad (4)$$

for each $s \in S$ *and all pure stationary strategies* $\mathbf{f}$ *and* $\mathbf{g}$.

Note that the stochastic game in section III-B is transient.

Lemma A.4. *Algorithm 2 provides the undiscounted optimal strategies for transient stochastic games.*

Proof: In transient stochastic games with bounded instantaneous payoffs, the undiscounted reward is also bounded, for each couple of stationary strategies (see [2]). Furthermore, under the transient condition, the uniform optimal strategies are optimal under the undiscounted criterion as well (see [1]). It is straightforward to prove that all the elements, belonging to $F(\mathbb{R})$, of the simplex tableaux built throughout the algorithm 1 are right continuous in $\rho = 0$ (or, equivalently, left continuous in $\beta = 1$). Therefore, we are allowed to shift the ordered field on which the algorithm works from $F(\mathbb{R})$ to $\mathbb{R}$, with $\beta = 1$. ∎

A. Undiscounted criterion with positive rewards

In this section we analize the concept of optimal strategies for the undiscounted criterion. Before, let us state two important Theorems.

Theorem A.5 (Abel's Theorem on power series)**.** *Let the power series* $f(x) = \sum_{n=0}^{\infty} a_n x^n$ *have radius of convergence* r *and still converge for* $x = r$. *Then,* $\lim_{x\uparrow r} f(x) = f(r)$.

Theorem A.6 ([4])**.** *Let* $\sum_{k\geq 0} c_k$ *be a divergent series of positive terms. Then*

$$\lim_{x\uparrow 1} \sum_{k\geq 0} x^k c_k = +\infty$$

Now we can we state as follows.

Corollary A.7. *Let* $\sum_{k\geq 0} c_k$ *be a series of positive terms and* $\xi \in \mathbb{R}$. *Then*

$$\begin{cases} \lim_{x\uparrow 1} \sum_{k\geq 0} x^k c_k = \xi \iff \sum_{k\geq 0} c_k = \xi \\ \lim_{x\uparrow 1} \sum_{k\geq 0} x^k c_k = +\infty \iff \sum_{k\geq 0} c_k = +\infty \end{cases}$$

Proof: For the *if* conditions, see Theorems A.5, A.6. About the *only if* conditions, we know [4] that a positive term series either converges or diverges to $+\infty$. If $\sum_{k\geq 0} c_k = \xi_1 \neq \xi$, then $\lim_{x\uparrow 1} \sum_{k\geq 0} x^k c_k = \xi_1$ for Theorem A.5. Hence, both the ($\Leftarrow$) relations are proved by contradiction. ∎

Now we are ready to state the following result.

Lemma A.8. *Suppose that all the instantaneous rewards are nonnegative. Let us utilize the extended line of real numbers, i.e. treat* $\pm\infty$ *as a number (*$\pm\infty = \pm\infty$, $-\infty < a \in \mathbb{R} < +\infty$*). Then, the uniform optimal strategies are optimal in the undiscounted criterion as well, i.e.*

$$\mathbf{\Phi}_1(\mathbf{f}, \mathbf{g}^*) \leq \mathbf{\Phi}_1(\mathbf{f}^*, \mathbf{g}^*) \leq \mathbf{\Phi}_1(\mathbf{f}^*, \mathbf{g}) \quad \forall\, \mathbf{f}, \mathbf{g} \qquad (5)$$

Proof: By definition, the saddle point relation (5) is valid $\forall\, \beta \in [\beta; 1)$ and hence also for the limit $\beta \uparrow 1$. Then, it is still valid for $\beta = 1$ for Corollary A.7. ∎

Appendix B
Formulation of Routing Game as a Stochastic Game

Let us formulate the routing model in section II as a stochastic game. Player 2 is the coalition $\mathcal{C} \subset \mathcal{M}$, while player 1 is the rest of the providers $\mathcal{M}/\mathcal{C}$. There exist a bijective association between the network nodes V and the states S. Let S_1 and S_2 be the set of states associated to the set of nodes $\bigcup_{\{k\}\in\mathcal{M}/\mathcal{C}} V_k$ and to $\bigcup_{\{k\}\in\mathcal{C}} V_k$, respectively. The network link $i \to j$ is activated if and only if player k selects the action $a^{(k)}_j(s_i)$, where $j \in \alpha_i$, $k : s_i \in S_k$. The instantaneous reward $r(s_i, a^{(k)}_j(s_i)) = \sum_{\{p\}\in\mathcal{C}} c_p(i,j)$, where k is the player that controls the node i. The transition probability is $p(s_w|s_i, a^{(k)}_j(s_i)) = \mathbb{I}(w = j)$, where $\mathbb{I}$ is the indicator function. Note that $\sum_{s'\in S} p(s'|s, \mathbf{f}, \mathbf{g}) = 1$, $\forall\, s \in S/\{s_N\}$ and for each couple of stationary strategies $(\mathbf{f}, \mathbf{g})$. The destination node is a "sink", i.e. $p(s_i|s_N) = 0$, $\forall\, i \in [1; N]$, and no actions are available in it for both players.

References

[1] K. Avrachenkov, L. Cottatellucci, L. Maggi, *Algorithms for uniform optimal strategies in two-player zero-sum stochastic games with perfect information,* http://hal.archives-ouvertes.fr/docs/00/50/63/90/PDF/RR-7355.pdf, (2010).

[2] J. Filar, K. Vrieze, *Competitive Markov Decision Processes,* Springer (1996).

[3] A. Hordijk, R. Dekker, L.C.M. Kallenberg, *Sensitivity Analysis in Discounted Markov Decision Processes,* OR Spektrum, Vol. 7, No. 3, pp. 143-151 (1985).

[4] K. Knopp, *Theory and Application of Infinite Series,* Dover (1990).

[5] R.B. Myerson, *Game Theory - Analysis of conflict,* Harvard University Press (1991).

[6] T.E.S. Raghavan, Z. Syed, *A policy-improvement type algorithm for solving zero-sum two-person stochastic games of perfect information,* Mathematical Programming, Vol. 95, No. 3, pp. 513-532 (2003).

[7] T.Roughgarden, *Routing Games,* chap. 18, pp. 461-486 in: N. Nisan, T.Roughgarden, E. Tardos, V. Vazirani, *Algorithmic Game Theory,* Cambridge University Press (2007).