Computing Stackelberg Strategies in Stochastic Games

JOSHUA LETCHFORD[1], LIAM MACDERMED[2], VINCENT CONITZER[1], RONALD PARR[1], and CHARLES L. ISBELL[2]

[1]Duke University and [2]Georgia Institute of Technology

Significant recent progress has been made in both the computation of optimal strategies to commit to (Stackelberg strategies), and the computation of correlated equilibria of stochastic games. In this letter we discuss some recent results in the intersection of these two areas. We investigate how valuable commitment can be in stochastic games and give a brief summary of complexity results about computing Stackelberg strategies in stochastic games.

---

1. INTRODUCTION

Computing game-theoretic solutions has long been one of the main research topics in the intersection of computer science and economics. Starting with an EC paper in 2006, much attention has been focused on the computation of optimal strategies in two-player *Stackelberg* games, in which player 1 is able to *commit* to a strategy before player 2 moves. Commitment has the potential to increase the utility of player 1, and, when commitment to *mixed* strategies is possible, it never decreases it [von Stengel and Zamir 2010]). The computation of Stackelberg mixed strategies has already found application in a number of real security problems, such as airport security [Pita et al. 2009], assigning Federal Air Marshals to flights [Tsai et al. 2009], and Coast Guard patrols [Shieh et al. 2012].

Most research on computing Stackelberg mixed strategies so far has focused on games where neither player learns anything about the other's actions until the end of the game, with the exception of player 2 learning player 1's mixed strategy before acting. This includes work on computing Stackelberg mixed strategies in normal-form games [Conitzer and Sandholm 2006; von Stengel and Zamir 2010; Conitzer and Korzhyk 2011], Bayesian games [Conitzer and Sandholm 2006; Paruchuri et al. 2008; Letchford et al. 2009; Pita et al. 2010; Jain et al. 2011], and security games (games inspired by the applications above, whose normal form would be exponentially large) [Kiekintveld et al. 2009; Korzhyk et al. 2010; Jain et al. 2010]. The only exceptions of which we are aware concern the computation of Stackelberg mixed strategies in extensive-form games [Letchford and Conitzer 2010] and, most recently, in stochastic games [Letchford et al. 2012; Vorobeychik and Singh 2012]. Vorobeychik and Singh focus on Markov stationary strategies, though these are

---
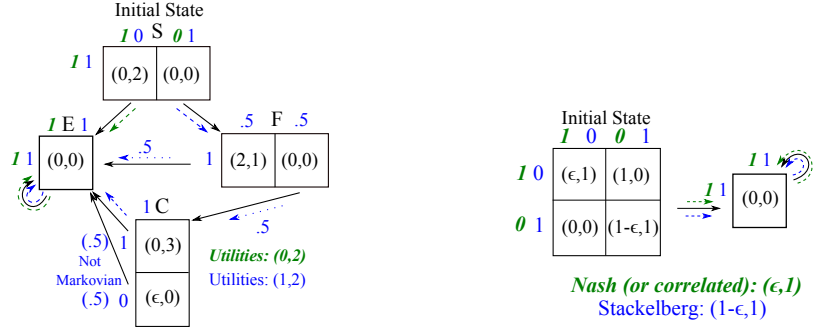
(a) Example game where signaling must oc-
cur early, but not too early.

(b) Game where commitment offers an un-
bounded advantage over correlation.

Fig. 1

generally not optimal. Indeed, various unintuitive phenomena occur when apply-
ing the Stackelberg model to stochastic games, and we hope to give some insight
into this in this brief article that summarizes our results.

The results in our paper [Letchford et al. 2012] fall in three main categories:
complexity results, theoretical results on the value of being able to commit and
the value of being able to correlate, and an approximation algorithm for finding
approximate Stackelberg strategies. Below, we highlight a few results on how valu-
able commitment can be to the leader in stochastic games, and summarize our
complexity results.

2.   COMMITMENT AND CORRELATION IN STOCHASTIC GAMES

A two-player stochastic game is defined as follows. There are two players, 1 and 2;
a set of states, T; and, for each state $t \in T$, a set of actions A_t^i for each player i. For
each state $t \in T$ and each action pair in $A_t^1 \times A_t^2$, there is an outcome, consisting
of two elements: (1) the immediate payoff that each player obtains in that round,
and (2) a probability distribution for the next state that the game transitions to.
Finally, there is a discount factor γ that is used to discount the value of future
payoffs.

In two-player normal-form games, it is known that correlation, by means of the
leader signaling to the follower before play, is of no value to the leader [Conitzer
and Korzhyk 2011]. This is no longer the case in stochastic games, and moreover,
the timing of the signaling matters. To illustrate this, consider the game pictured
in Figure 1a. This game has four states: S, F, C, and E. We assume state S is
our initial state; thus, play begins with one possible action for player 1 (the row
player, who is able to commit to a strategy ahead of time) and two possible actions
for player 2 (the column player), L_S and R_S. The state transitions in this example
are deterministic and are expressed using arrows (e.g., if player 2 chooses R_S then
play will transition to state F in the next round). Because E is an absorbing state,
we can set $\gamma = 1$. Suppose that player 1 is able to signal what she will play in state
C (if it is reached) before player 2 chooses his action in state F, but after he acts
in state S. Suppose that player 1 commits to drawing her play for state C from the

distribution $(.5 + \epsilon)U_C + (.5 - \epsilon)D_C$, and to sending the following signal to player 2 right before play in state F: R_F when she will be playing U_C, and L_F when she will be playing D_C. Then, in state F, player 2 will be best off playing according to the signal received from player 1. Moreover, in state S, player 2 will be best off playing R_S, resulting in an expected utility of $2 + 2\epsilon$ for him, and an expected utility of $1 - 2\epsilon$ for player 1.

In contrast, if player 1 only sends the signal after the transition to state C, then player 2 will prefer to play R_F with probability 1 in F for an expected utility of 1.5, and hence to play L_S in the first state. On the other hand, if player 1 signals what she will play in state C too early, namely before player 2 makes a choice in S, then player 2 will prefer to choose L_S when the signal is to play L_F. In both cases, player 1 receives 0.

In the figure, the resulting profile (for the limit case $\epsilon = 0$) is shown in blue, and the unique correlated (and, hence, Nash) equilibrium (without commitment) is shown in bold green itallics. This example shows that a combination of commitment and correlation (signaling) can be much more powerful than either alone.

3. AN ALGORITHM, AND EXPERIMENTS

For computing approximate correlated equilibria of a stochastic game (without commitment), an algorithm called QPACE [MacDermed et al. 2011] is available. It turns out that if we allow for the type of signaling discussed in the previous section, then QPACE can be modified to compute an optimal strategy for a Stackelberg leader in a stochastic game. By running both algorithms on randomly generated games, we obtain some insight into how valuable the ability to commit is in the "typical" case.

In the Stackelberg model, at the very least, the leader can commit to play according to the correlated equilibrium of (the non-Stackelberg version of) the stochastic game that is best for her. That is, Stackelberg leadership at least bestows the advantage of *equilibrium selection* on a player. On the other hand, as the example in the previous section shows, the benefit of Stackelberg leadership can go beyond this. But does it do so, to a significant extent, in the typical case?

The set of experiments represented in Figure 2a allows us to answer this question. Of course, to assess the benefit of equilibrium selection power, it is necessary to say something about which equilibrium would have resulted without this power. For this, we take the correlated equilibrium that corresponds to the Kalai-Smorodinsky bargaining solution, which favors equal gains to both parties. As the figure shows, the difference between the Stackelberg solution and the best correlated equilibrium is small compared to the difference to the bargaining solution, suggesting that most of the value comes from equilibrium selection, especially as γ grows.

We also examined how the number of actions affects the value of being able to commit. Figure 2b illustrates this value as the number of actions per player varies, over random games and with values of γ of 0.0 and 0.4. We observe that as the number of actions increases, the benefit of commitment decreases.
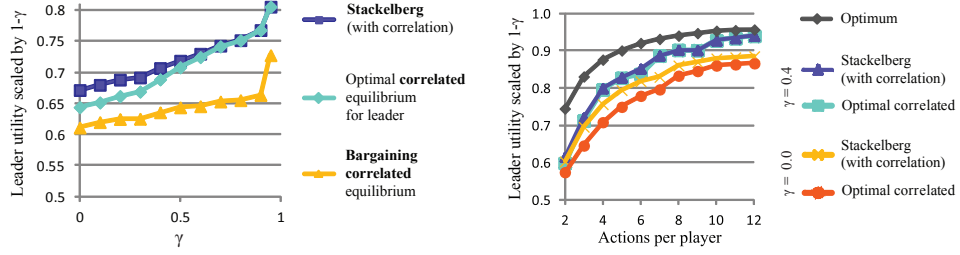
(a) The value of commitment compared to the value of equilibrium selection, as γ varies.

(b) The value of commitment compared to the value of equilibrium selection, as the number of actions varies.

Fig. 2

	$h = 0$	$0 < h < \infty$	$h = \infty$
Corr.	NP-hard (3SAT)	NP-hard (3SAT)	Modified QPACE (approximate)
No Corr.	NP-hard (3SAT)	NP-hard (3SAT)	NP-hard (Knapsack)

Fig. 3: Summary of hardness results. h represents the number of rounds that player 1 can remember.

4. COMPLEXITY RESULTS

For our complexity results for the Stackelberg model of stochastic games, we considered six different cases. First, we varied the amount of memory of previous rounds that the leader is allowed to use in her strategy, considering zero memory (i.e., stationary strategies), finite memory, and infinite memory. We considered each of these three cases both with and without correlation. Our results are summarized in figure 3.

5. ACKNOWLEDGEMENTS

REFERENCES

CONITZER, V. AND KORZHYK, D. 2011. Commitment to correlated strategies. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*. San Francisco, CA, USA, 632–637.

CONITZER, V. AND SANDHOLM, T. 2006. Computing the optimal strategy to commit to. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*. Ann Arbor, MI, USA, 82–90.

JAIN, M., KARDES, E., KIEKINTVELD, C., ORDÓÑEZ, F., AND TAMBE, M. 2010. Security games with arbitrary schedules: A branch and price approach. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*. Atlanta, GA, USA.

JAIN, M., KIEKINTVELD, C., AND TAMBE, M. 2011. Quality-bounded solutions for finite Bayesian Stackelberg games: Scaling up. In *Proceedings of the Tenth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. Taipei, Taiwan, 997–1004.

KIEKINTVELD, C., JAIN, M., TSAI, J., PITA, J., ORDÓÑEZ, F., AND TAMBE, M. 2009. Computing optimal randomized resource allocations for massive security games. In *Proceedings of the Eighth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. Budapest, Hungary, 689–696.

KORZHYK, D., CONITZER, V., AND PARR, R. 2010. Complexity of computing optimal Stackelberg strategies in security resource allocation games. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*. Atlanta, GA, USA, 805–810.

LETCHFORD, J. AND CONITZER, V. 2010. Computing optimal strategies to commit to in extensive-form games. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*. Cambridge, MA, USA, 83–92.

LETCHFORD, J., CONITZER, V., AND MUNAGALA, K. 2009. Learning and approximating the optimal strategy to commit to. In *Proceedings of the Second Symposium on Algorithmic Game Theory (SAGT-09)*. Paphos, Cyprus, 250–262.

LETCHFORD, J., MACDERMED, L., CONITZER, V., PARR, R., AND ISBELL, C. 2012. Computing optimal strategies to commit to in stochastic games. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*. Toronto, ON, Canada, 1380–1386.

MACDERMED, L., NARAYAN, K. S., ISBELL, C. L., AND WEISS, L. 2011. Quick polytope approximation of all correlated equilibria in stochastic games. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*. San Francisco, CA, USA.

PARUCHURI, P., PEARCE, J. P., MARECKI, J., TAMBE, M., ORDÓÑEZ, F., AND KRAUS, S. 2008. Playing games for security: An efficient exact algorithm for solving Bayesian Stackelberg games. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. Estoril, Portugal, 895–902.

PITA, J., JAIN, M., ORDÓÑEZ, F., PORTWAY, C., TAMBE, M., AND WESTERN, C. 2009. Using game theory for Los Angeles airport security. *AI Magazine 30*, 1, 43–57.

PITA, J., JAIN, M., ORDÓÑEZ, F., TAMBE, M., AND KRAUS, S. 2010. Robust solutions to Stackelberg games: Addressing bounded rationality and limited observations in human cognition. *Artificial Intelligence 174*, 15, 1142–1171.

SHIEH, E., AN, B., YANG, R., TAMBE, M., BALDWIN, C., DIRENZO, J., MAULE, B., AND MEYER, G. 2012. PROTECT: A deployed game theoretic system to protect the ports of the united states. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

TSAI, J., RATHI, S., KIEKINTVELD, C., ORDONEZ, F., AND TAMBE, M. 2009. IRIS - a tool for strategic security allocation in transportation networks. In *The Eighth International Conference on Autonomous Agents and Multiagent Systems - Industry Track*. 37–44.

VON STENGEL, B. AND ZAMIR, S. 2010. Leadership games with convex strategy sets. *Games and Economic Behavior 69*, 446–457.

VOROBEYCHIK, Y. AND SINGH, S. 2012. Computing stackelberg equilibria in discounted stochastic games. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*. Toronto, ON, Canada.