Algorithms, Games, and the Internet *

Christos H. Papadimitriou
University of California, Berkeley
christos@cs.berkeley.edu

ABSTRACT

If the Internet is the next great subject for Theoretical Computer Science to model and illuminate mathematically, then Game Theory, and Mathematical Economics more generally, are likely to prove useful tools. In this talk I survey some opportunities and challenges in this important frontier.

1. INTRODUCTION

Over the past fifty years, researchers in Theoretical Computer Science have sought and achieved a productive foundational understanding of the von Neumann computer and its software, employing the mathematical tools of Logic and Combinatorics. The next half century appears now much more confusing (half-centuries tend to look like that in the beginning). What computational artifact will be the object of the next great modeling adventure of our field? And what mathematical tools will be handy in this endeavor?

The Internet has arguably surpassed the von Neumann computer as the most complex computational artifact (if you can call it that) of our time. Of all the formidable characteristics of the Internet (its size and growth, its almost spontaneous emergence, its open architecture, its unprecedented availability and universality as an information repository, etc.), I believe that the most novel and defining one is *its socio-economic complexity:* The Internet is unique among all computer systems in that it is built, operated, and used by a multitude of diverse economic interests, in varying relationships of collaboration and competition with each other. This suggests that the mathematical tools and insights most appropriate for understanding the Internet may come from a fusion of algorithmic ideas with concepts and techniques from Mathematical Economics and Game Theory[1] (see [18, 23] for two excellent introductions in the respective subjects, and see the web site www.cs.berkeley.edu/~christos/cs294.html for many additional references to work in this interface.)[2]

In this talk I shall review some of the many important points of contact between Game Theory and Economic Theory, Theoretical CS, and the Internet. In doing so I am necessarily (and, to an observer, arbitrarily) selective, leaving out important areas such as combinatorial auctions [5], and computational learning in games [9].

2. NASH EQUILIBRIUM

Game theory was founded by von Neumann and Morgenstern (in fact, about the same time von Neumann designed the EDVAC...) as a general theory of rational behavior. Game theoretic concepts are already familiar to theoretical computer scientists: Proving lower bounds is often best seen as a game between an algorithm designer and an adversary [30], while strategic two-person games are important complexity paradigms [1] and tools in finite model theory (Ehrenfeucht-Fraissé), for example. There has been fertile interaction in the recent past between Game Theory and CS Theory in the context of bounded rationality and repeated games [25] as well as learning games [29]. Game Theory's sharp but pointedly faithful modeling, twisted cleverness, and often unexpected depth make it quite akin to our field; but this may also be deceptive, since Game Theory is also characterized by a cohesive and complex research tradition and a defiantly original point of view and norms that are often hard to get accustomed to.

In a game, each of n players can choose among a set of strategies $S_i, i = 1, \ldots, n$, and there are functions $u_i, i = 1, \ldots, n : S_1 \times \cdots \times S_n \mapsto \Re$ which assign to each such combined choice a payoff for each player. The fundamental question of Game Theory is, what constitutes rational behavior in such a situation? The predominant concept of rationality here (but my no means the only one) is the *Nash equlibrium:* A combination of strategies $x_1 \in S_1, \ldots, x_n \in S_n$ for which $u_i(x_1, \ldots, x_i, \ldots, x_n) \geq u_i(x_1, \ldots, x_i', \ldots, x_n)$ for all i and

*An extended abstract of this paper appears in the proceedings of the 2001 ICALP Conference

[1] A more radical point of view would be this: Since computation has moved over the past twenty years decisively closer to people, interfaces with *social sciences* such as Psychology and Sociology, besides Economics, have become increasingly important, and TCS needs to build bridges with the mathematical vanguards of those fields.

[2] But why, one may ask, should we embark on the foundational understanding of something that was not designed — and seems inherently undesignable? First, when faced with a novel and complex computational phenomenon, it seems to me that our community has no choice but to study it. Second, the Internet *is* being engineered —albeit in a subtle, diffuse, and indirect way. Foundational insights will be, with any luck, noted and appreciated.

STOC'01, July 6-8, 2001, Hersonissos, Crete, Greece.

$x_i' \in S_i$; a behavior, that is, from which no player has an incentive to deviate.

The Nash equilibrium concept is not without its problems. First, a game may not have one; here Nash himself provided an ingenious way out, by showing that Nash equilibria always exist if the S_i's are convex sets (the proof makes use of Kakutani's Theorem, a deep fact from topology with combinatorial origins, a generalization of Brouwer's Fixpoint Theorem); and a general way of making any set convex is to allow *distributions* (convex combinations) over the set. Thus, if we allow the x_i's to be randomized ("mixed" in the field's terminology) strategies, then a Nash equilibrium always exists. (In fact, this generalizes the well-known min-max theorem for zero-sum games, and thus linear programming duality.) Second, there is typically more than one Nash equilibrium in a game, and there is no useful way of choosing between them —it is a "declarative" concept containing no recipe for "getting there." This is an obvious invitation to algorithmic ideas, and some have been tried.

But the most interesting aspect of the Nash equilibrium concept to our community is that *it is a most fundamental computational problem whose complexity is wide open.* Suppose that $n = 2$ and S_1, S_2 are finite sets. *Is there a polynomial algorithm for computing a (mixed) Nash equilibrium in such a game?* Because of the guaranteed existence of a solution, the problem is unlikely to be NP-hard; in fact, it belongs to a class of problems "between" P and NP, characterized by reliance on *the parity argument* for the existence proof [24]. In a different direction, as we have already pointed out, this problem is a generalization of linear programming; in fact, there is an algorithm for it that is a combinatorial generalization of the simplex algorithm [2] (as a corollary, the solution is always a vector of rational numbers, something that is not true in general for $n \geq 3$ players).

Together with factoring, *the complexity of finding a Nash equilibrium is in my opinion the most important concrete open question on the boundary of P today.*

3. INTERNET EQUILIBRIA

Internet bandwidth is scarce; its allocation to individual end-to-end flows is achieved via the TCP/IP congestion control protocol: "If the previous batch of packets got through, then increase the batch size by one; if not, decrease it by half." This ingeniously simple scheme seems to work, and its users do not seem eager to abandon it for something more aggressive, but the origins of this apparent success and acquiescence are not well understood. One is justified to wonder: *Of which game is TCP/IP congestion control the Nash equilibrium?* [12][3]

If we see Internet congestion control as a game, we can be sure that its equilibrium is not achieved by rational contemplation, but by interaction and adaptation in an environment where conditions (and player populations) change rapidly (and in which changes in strategy incur costs). These considerations may lead to more sophisticated concepts of equilibria that are more appropriate for the context of the Internet; see [8] for initial work in this direction.

Games as defined above assume that players cannot negotiate with and compensate each other by side payments. *Coalitional game theory* [23] considers a game of n players as a set of possible $2^n - 1$ coalitions, each of which, call it S, can achieve a particular *value* $v(S)$ (the best possible sum of payoffs among players in S, against worst-case behavior of players in $[n] - S$). The problem is now how to divide the total payoff $v([n])$ among the n players. Many such notions of "fairness" have been proposed, defended and criticized over the past decades: The Shapley value, the kernel, the bargaining set, the nucleolus, the von Neumann-Morgenstern solution, and many others (see [23], and [3] for a complexity-theoretic treatment of the subject). The core is perhaps the most intuitive (and akin to the equilibrium concept); it is also the most conservative (as a result, games often have empty core): A vector $x \in \Re^n_+$ with $x[[n]] = v([n])$ (notation: $x[S] = \sum_{i \in S} x_i$) is in the core if $x[S] \geq v([S])$ for all S. That is, x, considered as a proposed splitting of the total payoff $v([n])$ among the n players, is fair according to the core school if no coaltion has an incentive to secede (because no coalition can make more by itself than it is allocated in x).

The Internet seems to me an intriguing theater of coalitional game theory. It is operated (and built) by thousands of large and small entities ("autonomous systems"), collaborating with admirable effectiveness to process and deliver end-to-end flows originating and terminating in any one of them, using an opaque protocol called BGP [27]. Consider the following abstraction of the situation: We are given a graph with n nodes (the autonomous systems); an $n \times n$ symmetric traffic matrix F, where f_{ij} is the total traffic requirements between customers of i and customers of j; and a capacity c_i for each node (a simplification attempting to capture the capacity of i's subnetwork to carry traffic). If S is a set of nodes, consider the subgraph induced by S as a multicommodity network with node capacities and commodity requirements given by the entries of F; let $v(S)$ be the maximum total flow in this network —notice that this defines a coalitional game.

The key problem here is this: Find an optimum solution in the multicommodity flow problem for the overall network, achieving a flow matrix $F' \leq F$, such that the corresponding payoffs for the nodes $x_i = \sum_j f'_{ij}$ are in the core of the coalitional game v (or abide by one of the other notions of fairness mentioned above). Here we assume that autonomous system i's payoff increases with the flow to and from i's customers.

4. THE PRICE OF ANARCHY

There is no central authority that designs, engineers and runs the Internet. [4] But what if there were such master puppeteer, a benevolent dictator who, for example, micro-managed its operation, allocating bandwidth to flows so as to maximize total satisfaction? How much better would the Internet run? *What is the price of anarchy?*

This question was posed (and partially answered in the restricted context of a network consisting of two nodes and parallel edges) in [16]. This is an instance of a more general

[3]In recent work, Kelly [10] establishes, by resorting to ODEs and Lyapunov functions, that TCP/IP is the control function that optimizes the sum of user utilities (assumed to be of the form $\arctan(cb)$ where b is the bandwidth allocated to the user and c a constant) minus the total number of packet drops.

[4]Recall David Clark's famous maxim: "We reject kings, presidents and voting. We believe in rough consensus and running code."

pattern, of a novel and timely genre of problems: Given a game-like situation, we seek the ratio between the worst-case Nash equilibrium and the optimum sum of payoffs. In other words, if competitive analysis reveals the price of not knowing the future, and approximability captures the price of not having exponential resources, the present analysis seeks the price of uncoordinated individual utility-maximizing decisions —that is to say, the price of anarchy. Since that paper there has been progress in this front (not least, in the present conference [19]), including a marvelous result [28] stating that, in the context of a multicommodity flow network in which message delays increase with edge congestion while flows choose paths so as to minimize delay, the price of anarchy is *two* (more precisely, the ararchistic solution is no worse than the optimum solution with double the bandwidth).

But, of course, in today's Internet flows cannot choose shortest paths. In the Internet, routers direct traffic based on local information, users respond to delay patterns by modifying their traffic, and network providers throw bandwidth at the resulting hot spots. How does this compare in efficiency with an ideal, *ab initio* optimum design? *What is the price of the Internet architecture?*

5. ROUGH MARKETS

As a further example of the possible interaction between Economic Theory and computational ideas, let me present an important economic model that transcends games, namely *markets,* a fundamental theorem about them, as well as a recent result by Deng Xiaotie and myself. We again have n agents, each of which possesses a nonnegative vector (its *endowment*) $e_i \in \Re^k_+$ of k *goods*, and a concave *utility function* u_i mapping $\Re^k_+$ to $\Re_+$. The agents may be all dissatisfied with their current endowments, in that there may be a reallocation of the same goods that is higher in everybody's utility (the set of local optimal allocations for which such overall improvement is impossible comprise the *Pareto set* of the market). Bilateral and multi-way exchanges and bartering may slowly improve the situation inching towards the Pareto set, but at considerable delay and communication cost.

Prices can be seen as an ingenious and efficient (even in the CS sense of low communication complexity) way for reaching the Pareto set. Suppose that there is a per unit price p_j for each good, a nonnegative real number. The only rational behavior for each agent would then be, to sell her endowment e_i at these prices, and to buy with the proceeds $p \cdot e_i$ a new vector of goods $\hat{x_i} \in \Re^k_+$ that is the "best vector she can afford," that is, the solution to the following optimization problem: max $u_i(x_i)$ such that $p \cdot x_i \leq p \cdot e_i$. But what guarantees do we have that there will be enough goods to fill everybody's "optimum affordable shopping cart?" Or that no goods will be left on the shelves?

Theorem (Arrow-Debreu, 1953): (Under some technical but reasonable assumptions about the u_i's,) there is always a price vector p called the *price equilibrium* such that *the market clears*, that is, the solutions $\hat{x_i}$ to the optimization problems above satisfy $\sum_{i=1}^n \hat{x_i} = \sum_{i=1}^n e_i$.

The proof uses Brouwer's fixpoint theorem. In a situation mirroring Nash equilibrium, there is no known polynomial algorithm for computing equilibrium prices in an economy (the corresponding problem is in fact *complete* for the parity class mentioned above [24]), even though there are empirically good algorithms.

Economists had been aware for half a century that the Arrow-Debreu theorem breaks down when the goods are discrete (bridges, days of work, airplanes, etc.). The following recent result captures this diffuse awareness in a computational context, and povides a remedy:

Theorem: [4] If (some of) the goods are integer-valued, then a price equilibrium may not exist, and it is in fact (strongly) NP-hard to tell if a price equilibrium exists (weakly NP-hard even when $n = 3$), even when the utilities are linear. However, (under some technical but reasonable assumptions about the u_i's,) there is a fully polynomial-time approximation scheme that computes a price equilibrium that is ϵ-*approximate in expectation* (definition omitted) if the number of goods is fixed.

The second (approximation) part uses randomized rounding [26].

6. MECHANISM DESIGN

If Game Theory strives to understand rational behavior in competitive situations, the scope of Mechanism Design (an important and elegant research tradition, very extensive in both scope and accomplishment, and one that could alternatively be called "inverse game theory") is even grander: Given desired goals (such as to maximize a society's total welfare), design a game (strategy sets and payoffs) in such a clever way that individual players, motivated solely by self-interest, end up achieving the designer's goals. There have been recently interesting interactions between this fascinating area and Theoretical CS, see e.g. [22, 7], and further opportunities abound. This area is too sophisticated and developed for a brief tutorial to be meaningful (see [18] for an excellent chapter, and [21] for a TCS-friendly introduction). Instead, I shall briefly develop an argument for its importance.

The complex socio-economic context of the Internet can have a deep influence on the design process in CS, and the research agenda of Theoretical CS. Traditionally, the "goodness" or "fitness" of a computational artifact (a new compiler, say) could be captured by its time and space performance, as well as its reliability, usability, etc. Such attributes were a fair approximation to the artifact's "fitness" (its chances for success), and theoreticians strived to develop methodologies for predicting and optimizing these attributes.

In the context of the Internet, such attributes only tell a small part of the story. If an artifact (a new congestion control protocol, a new caching scheme, a new routing algorithm, etc.) is demonstrated to have superior perfomance, this does not necessarily mean that it will be successful. For the artifact to be "fit," there must exist a *path* leading from the present situation to its prevalence.[5] This path must be paved with incentives that will motivate all kinds of diverse agents to adopt it, implement it, use it, interface with it, or just tolerate it. In the absence of such a path, the most

[5]This is not unlike biological systems, where it is known that successful genes and traits are the ones for which there is a continuously fitness-increasing path leading from the current phenotype and genotype to the target ones. See [20] for a fascinating interplay between between Game Theory and Biology.

clever, fast, and reliable piece of software may stay just that. All design problems are now mechanism design problems.

7. THE ECONOMICS OF PRIVACY, CLUSTERING, AND THE WEB GRAPH

There is no end to the list of computational matters for which the economic viewpoint leads to interesting insights, as well as novel algorithmic problems. I briefly discuss here three more examples from my recent work.

Privacy is arguably the most urgent concern and mission of Computer Science, and yet there is very little foundational work about it. In a recent paper [14] we argue that it has an important economic aspect: The problem with privacy is that decisions about the use of personal information are made by entities other than the person involved (such anomalies, called *externalities*, are known in Economics to be the root of most evil); to put it otherwise, personal information is intellectual property that bears negative royalty. Coalitional game theory is an interesting modeling tool here as well, since it can help determine the fair royalty due to the individual for any use of his or her private information. In [14] we study certain stylized versions of common situations (such as marketing surveys and collaboratie filtering) in which personal information is used, and the interesting algorithmic problems involved in computing fair royalties.

Clustering is one of the most practically important yet foundationally underdeveloped areas of CS (despite the steady stream of clever algorithmic ideas for approximately solving optimum clustering problems). There are far too many criteria for the "goodness" of a clustering (min-sum or min-max cluster diameter, min-sum or min-max distance from the center (which is either forced to be one of the original points or is not), or novel spectral parameters, to name only the ones most popular at STOC) and far too little guidance about choosing among them. In [15] we point out that economic considerations are crucial for understanding the issues here as well. For the purpose of clustering is to improve decision-making by allowing different segments of a space or population to be treated differently. The criterion for choosing the best clustering scheme cannot be determined unless the decision-making framework that drives it is made explicit. In [15] we show that this point of view gives rise to a host of novel optimization problems, called *segmentation problems.*

Consider the following hypothetical situation: A monopolist knows the demand curve of each of its customers: If the price is x, the ith customer will buy a quantity $y = b_i - a_i \cdot x$. The monopolist wants to cluster its customers into k segments, with a different price to each segment, in order to maximize revenue. How should this clustering be done? Which one of the two dozen or so criteria in the theoretical and experimental literature should be adapted here, and which approximation algorithm or heuristic should be used?

Theorem: The clustering that maximizes revenue subdivides the customers into segments of consecutive values of $\frac{a_i}{b_i}$. Thus, the optimum can be computed in $O(n^2)$ by dynamic programming.

The most attractive feature of the clustering algorithm suggested by this result is not that it is efficient, or that it finds the exact optimum. It is that *it optimizes the right thing.*

The Web Graph. It has been established recently [13, 11] that the world-wide web can be usefully considered as a directed graph with the documents as nodes and the hyperlinks as edges. Intuitively, the web is a huge "random" graph —except that it seems to violate every single prediction of the classical random graph models, such as $G_{n,p}$, so familiar to our community. For example, its indegrees and outdegrees of the nodes are distributed by a polynomial-tailed distribution (the x-th largest indegree or outdegree is about $c \cdot x^{-\alpha}$ for positive constants c and α) instead of the sharp Gaussian predicted by the law of large numbers, its giant strongly connected component covers about 40% of the nodes (instead of 0% or 100%), there are scores of $K_{3,3}$ subgraphs (instead of none), etc. Recently, there have been interesting efforts to model the web graph (see, for example, [17] for some of the latest). Despite much interest, we know of no model that predicts and explains to a satisfactory degree these and other features of the web graph, starting from primitive and credible assumptions about the world-wide web.

Heavy-tailed distributions were first observed in Economics: City populations are known to behave this way (the population of the x-th largest city of any country is eerily close to c/x, with c depending on the country), but also market shares (the market share of the xth largest manufacturer in any sector is often distributed as $c \cdot x^{-\alpha}$), as well as income distributions. Is it possible that the puzzling structure of the web graph has its origins in economic phenomena of this kind? This is not as implausible as it may seem at first: A document's indegree is determined by how *interesting* the document is, and interest is intuitively analogous to market share (and as much determined by fierce competition...), while outdegree depends on the entity's *attention*, which seems to me not entirely unlike income. With several students at Berkeley we are running experiments to determine the explanatory power of such considerations.

Last, heavy-tailed distributions are also observed beyond the world-wide web, in the Internet: The degrees of the routers and the autonomous systems are also heavy-tail distributed [6]. It would be interesting to expore if this can be the result of some rough local optimization heuristic used to allocate new links and routers, in the face of exponentially growing traffic.

8. SOME OPEN PROBLEMS

Great new areas are full of open problems (or, more likely, no area can get off the ground until a slew of cool and challenging open problems attracts the attention of smart, ambitious researchers). To recapitulate, here are the ones proposed in this talk:

- Is there a polynomial algorithm for computing a Nash equilibrium in a 2-person game? For $n \geq 3$ players? A polynomial-time approximation scheme?
- Ditto for a market equilibrium. Is there a PTAS for integer markets whose exponent does not involve the number of goods?
- Develop a reasonably faithful game-theoretic model of Internet congestion control for which (an approximation of) TCP/IP is a Nash equilibrium. Alternatively, develop a crisp notion of equilibrium appropriate for games modeling the Internet.

- Is there a polynomial algorithm for determining, given a network, a traffic matrix, and node capacities, whether there is a flow in the game's core? More interestingly, under what circumstances (conditions relating connectivity, capacities and traffic matrix) is such a flow guaranteed to exist?
- What is the price of the Internet architecture? (See Section 4.)
- Develop a graph generation model for the world-wide web, plausibly capturing key aspects at a primitive level (my conjecture: its economic aspects are indispensible), that predicts theoretically the observed characteristics of the web graph.
- Show that a particular simple local improvement heuristic for network design in the face of increasing flow demands results in heavy tails in node degrees, if driven by exponentially increasing flow demands.

9. REFERENCES

[1] Chandra, Kozen, Stockmeyer, "Alternation," *JACM, 28*, 1, 1981.
[2] Cottle *The Linear Complementarity Problem*, 1992.
[3] Deng, Papadimitriou, "On the complexity of cooperative game solution concepts," *Math. Oper. Res., 19*, 1994.
[4] Deng, Papadimitriou, in preparation.
[5] de Vries, Vohra, "Combinatorial Auctions: A Survey." http://www.kellogg.nwu.edu/research/math/Downpapers.htm
[6] Faloutsos, Faloutsos, Faloutsos, "On power-law relationships of the Internet topology," *Proc. 1999 SIGCOMM.*
[7] Feigenbaum, Papadimitriou, Shenker "Sharing the cost of multicast transmissions," *Proc. 2000 STOC.*
[8] Friedman, Shenker "Learning and implementation on the Internet," 1998.
[9] Fudenberg *Learning in Games*, M.I.T. Press, 1998.
[10] Kelly "Mathematical modelling of the Internet," in *Mathematics Unlimited - 2001 and Beyond* (Editors B. Engquist and W. Schmid), Springer-Verlag, 2001.
[11] www.google.com/technology/index.html
[12] Karp, Koutsoupias, Papadimitriou, Shenker, "Optimization problems in congestion control," *Proceedings of the 41st FOCS*, 2000.
[13] Kleinberg "Authoritative sources in a hyperlinked environment," *Proc. 1999 SODA.*
[14] Kleinberg, Papadimitriou, Raghavan "On the value of private information," *Proc. 2001 Conf. on Theoretical Aspects of Reasoning about Knowledge,* Sienna, July 2001.
[15] Kleinberg, Papadimitriou, Raghavan "Segmentation problems," *Proc. 1998 STOC.*
[16] Koutsoupias, Papadimitriou "Worst-case equilibria," *Proc. 1998 STACS.*
[17] Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkins, Upfal, "Random graph models for the Web graph," *Proc. 2000 FOCS.*
[18] Mas-Collel, Winston, Green *Microeconomic Theory,* Oxford University Press 1995.
[19] Mavronikolas, Spirakis, "The price of selfish routing," *this conference.*
[20] Maynard-Smith, *Evolution and the Theory of Games* Cambridge University Press, 1982.
[21] Nisan "Algorithms for selfish agents – Mechanism design for distributed computation," *Proc. 1999 STACS.*
[22] Nisan, Ronen "Algorithmic Mechanism Design," *Proc. STOC 1999.*
[23] Osborne and Rubinstein, *A Course in Game Theory,* MIT Press, 1994.
[24] Papadimitriou "On the complexity of the parity argument and other inefficient proofs of existence, *J.CSS, 48,* 3, 1994.
[25] Papadimitriou, Yannakakis "Complexity as bounded rationality," *Proc. 1994 STOC.*
[26] Raghavan, "Probabilistic construction of deterministic algorithms: approximate packing integer programs," *J.CSS, 37,* 1988.
[27] Rekhter, Li "RFC 1771, BGP-4," www.cis.ohio-state.edu/cgi-bin/rfc/rfc1771.html.
[28] Roughgarden, Tardos, "How Bad is Selfish Routing? (Extended Abstract)," *Proc. FOCS 2000.*
[29] Singh, Kearns, Mansour, "Nash Convergence of Gradient Dynamics in General-Sum Games," *Proc. UAI 2000.*
[30] Yao "Probabilistic computations: Toward a unified measure of complexity (extended abstract)," *Proc. 1977 FOCS.*