
DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU


HARVARD LIBRARY
Office for Scholarly Communication

Computational Challenges in E-Commerce

Citation
Feigenbaum, Joan, David C. Parkes, and David M. Pennock. 2009. Computational challenges in e-commerce. Communications of the ACM 52(1): 70-74.

Published Version
doi:10.1145/1435417.1435435

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:4039776

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

DOI:10.1145/1435417.1435435

Economic and social sciences will drive Internet protocols and services into the future.

BY JOAN FEIGENBAUM, DAVID C. PARKES, AND DAVID M. PENNOCK

Computational Challenges in E-Commerce

COMPANIES AND INDIVIDUALS are using computer networks to conduct increasing amounts of their daily business. Web search engines auctioned some $10 billion of ad space in 2007, accounting for almost half of all online advertising revenue. Sales at Amazon.com were $4.13 billion in the first quarter of 2008, including a fast-growing revenue stream from selling Web services to other e-commerce companies. At eBay, sales reached $15.7 billion in the second quarter, with 84.5 million active users.

This explosion of large-scale e-commerce poses new computational challenges that stem from the need to understand incentives. Because individuals and organizations that own and operate networked computers and systems are autonomous, they will generally act to maximize their own self-interest—a notion that is absent from traditional algorithm design. In this article, we provide an overview of four areas of computation in which incentives play a crucial role: resource allocation, knowledge integration, peer production and interaction, and security and privacy.

Resource Allocation

Allocating scarce resources—from bread to bytes—is a fundamental process that permeates economics and, indeed, society. Participants declare their perceived value for the resource and the market computes the best (for example, value-maximizing) allocation and the prices that participants should pay.

One decentralized prescription for resource allocation is an auction. Classical auctions emphasize simple rules for setting allocations and prices, which can be determined manually. Many of the largest marketplaces in the world, including financial exchanges, are at their core based on these centuries-old procedures. In one week of March 2008, the U.S. treasury sold, largely through manual means, more than $22 billion in three-month treasury bills.

Modern computer systems can support much richer and more flexible mechanisms. Governments use auctions to allocate property rights such as wireless spectrum (with worldwide proceeds exceeding $100 billion by the end of 2001). *Combinatorial* auctions allow bidders to express values for bundles of goods—for example, in assigning a higher value to two adjacent properties than the sum of the values assigned to each.[12] Generalized combinatorial auctions with rich and natural forms of expressiveness—volume discounts, side constraints, and bundling requirements, among others—are used to determine billions of dollars of spending within the supply chain, even though the problem is NP-hard. For example, they are used to source truckload-transportation logistics for Procter & Gamble, Walmart, and Target.[30]

Advertising is a business based on allocating *attention*, one of the scarcest and most valuable of resources. Media companies capture attention by providing information or entertainment and

typically sell a fraction of that attention to advertisers.

Historically, advertising sales featured straightforward allocation rules and manual negotiations. But now more aspects of advertising, including its sale, delivery, and measurement, are being automated. Web-search engines such as Google and Yahoo! have led the way, selling space beside particular search queries in continuous dynamic auctions worth billions of dollars annually.

Auctions and exchanges for all types of online advertising—including banner and video ads—are commonplace at present, and they are run by startups and Internet giants alike. Advertisers can buy not only space but also contextual events—such as clicks from a specific user on a specific property at a specific time—or, more generally, bundles of contextual events. An ecosystem of third-party agencies has grown to help marketers manage their increasingly complex ad campaigns.

The rapid emergence of new modes for selling and delivering ads is fertile ground for research, both from economic and computational perspectives.[25] Edelman et al.[15] and Varian[31] model how advertisers bid in search-ad auctions. Essentially, the advertisers raise their bids until they reach a point of indifference between staying where they are and swapping with the advertiser above them on the page. The authors show that this bidding strategy forms the basis of a symmetric Nash equilibrium and, in a nice example where theory aligns with practice, that real bidding behavior is largely consistent with the model.

A number of questions drive research in ad auctions and exchanges. What mechanisms increase advertiser value or publisher revenue? What user and content attributes contribute to variation in advertiser value? How can bids for different contingencies (impressions, clicks, or conversions) be integrated and optimized over time? What constraints on supply and budget make sense? How should advertisers and publishers bid? How can publishers and advertisers incorporate learning and optimization (while trying to balance exploration and exploitation)? How do practical constraints such as real-time delivery affect design? How is automation changing the advertising industry? More information can be found in the *Proceedings of the Workshop on Ad Auctions* series.[33]

Knowledge Integration

The eliciting and aggregation of information from diverse and frequently self-interested sources is in general called "knowledge integration," with a particular case being "price discovery"—a side effect of market-based resource allocation. The balance point of supply and demand reveals the negotiated value of the resource.

In some cases, the value revealed in prices can rival or eclipse the value of trade. For example, the price of an asset that pays $1 if a category-5 hurricane hits Florida in 2009 can be seen as a probabilistic forecast of this catastrophic event. The value of an actual and more accurate forecast could run into the millions of dollars.

A "prediction market" is a market designed primarily for price discovery rather than resource allocation, and this alternate focus leads to a different prioritization of design goals. For example, the market operator may be happy to pay for the information it seeks, instead of enforcing neutral or positive revenue. Trading is not the end goal but a means to the end of acquiring complete, accurate, and timely information. For example, the Iowa Electronic Market forecasts the outcomes of political elections, and intrade.com predicts events ranging from the outbreak of avian flu to Osama bin Laden's capture.

Liquidity and expressiveness play important roles in prediction markets. If a trader with information cannot reveal it to the market, either because illiquidity prevents matching with another trader or because the market does not support the way in which the trader wants to express information, then the mechanism may fail.

Designing prediction markets to improve liquidity and expressiveness poses substantial though not insuperable computational challenges.[25] Liquidity can be addressed through the use of automated market makers that are always willing to buy and sell at some prices and that adjust prices dynamically to ensure a bound on their worst-case loss.[18, 26] Expressiveness is gained at an often-severe computational cost, thus placing a difficult computational problem in the lap of the auctioneer striving to bring matched traders together or of the market maker trying to (implicitly) maintain an exponential number of prices.[10] In some instances, there is a reasonable and useful compromise between expressiveness and computational complexity.[11]

A more direct means of obtaining information is to pay an expert, though



payment of a fixed amount does not motivate the expert to be either truthful or careful, let alone to actively seek out new information. A "scoring rule" is a payment function that depends on the expert's prediction and the actual outcome in such a way as to motivate truthful participation.[32] Shared scoring rules can form the basis of self-financing (budget-balanced) wagering mechanisms to obtain multiple individual forecasts.[23] Indeed, the line between scoring rules and markets becomes blurred. For example, the most common automated market maker used for prediction markets can be viewed as a sequential shared scoring rule.[18]

Rating systems are forums for gathering subjective opinions on a variety of things, such as movies, restaurants, or trading partners. Unlike forecasts, rating and reputation systems have no fundamental truths on which to base rewards. Although rating systems may provide personalized recommendations or advice, the incentives of raters may not align with these goals. Real rating systems do provide considerable value, despite the persistence of spam and pollution. Designing rating systems that are resistant to manipulation is an important challenge, requiring both good algorithms and good economics.[28]

Peer Production and Interaction

"Peer production," a term coined by Benkler,[6] refers to large-scale collaboration that is not based on price signals and occurs outside of the typical hierarchies and reward structures provided by firms. Salient examples of successful artifacts of peer production include Wikipedia and Linux. *Social* production, a more general phenomenon, describes the output of social relations—for example, the videos that people upload to YouTube and the content on social-networking sites such as Facebook. Taken together, these activities make up an energetic swath of the e-commerce landscape, both because of the opportunities to promote nontraditional production through appropriate feedback, trust, and accounting mechanisms and because of opportunities for targeted advertising and monetization efforts.

How can it be that peer production seems to succeed despite the widely accepted economic model of people as selfish and rational actors? Nontraditional motivations, such as hedonic pleasure from doing useful work and civic pride in observing community and societal norms, seem to be a part of the story. Indeed, the established field of behavioral economics seeks to explain people's social or "other-regarding" preferences.[16] Monetary rewards can actually lead to the crowding out of social motivations and to increasingly selfish behavior. In one noted example, when a country moved from a system of voluntary blood donations to one with small payments it found that donation rates went down instead of up.[17]

Despite positive examples of peer production, there is little formal knowledge about the design of successful peer-production systems. Pertinent methodologies seem necessarily more indirect than those of economic mechanism design,[20] given that actions are intrinsically voluntary. One challenge is to design systems that observe behaviors with a view to learning (social) preferences. Can environments be usefully modulated through appropriate constraints and affordances (for example, with moderator rights and other tiers), the assignment of rewards (say, gold stars),[35] and the ability to aggregate and disseminate information about peers (such as through scoreboards and social context)?

Lessons learned from peer-to-peer file sharing suggest that incentive considerations will remain important even in the presence of other-regarding preferences. Early protocols failed to provide appropriate incentives for the uploading of files, and systems such as Gnutella suffered from a large amount of free-riding.[3] The BitTorrent protocol addressed this problem by limiting a user's download rate according to upload history, thus mitigating incentives to free-ride. But such tit-for-tat during bilateral peering introduces its own market inefficiencies; for example, users cannot contribute upload resources in return for credit for later downloads unless centralized trackers are employed. Hybrid systems that allow for accounting and some form of currency may be of interest, though they have associated challenges with regard to dynamic stability.[22] There are interesting directions as well in the introduction of social structure into peering societies.[27]

Reputation and trust metrics have an important role to play in Internet commerce,[29] yet finding the right design can be quite a subtle problem. For example, one study suggests that good reputations of sellers on eBay are sometimes the result of a badly designed feedback protocol. A seller, who has the "last move," can punish a buyer who leaves negative feedback; the seller may respond with negative feedback of his own about the buyer.[8]

Recent work has formalized the challenges of providing provably nonmanipulable trust metrics in graph-theoretic terms. Suppose that players are nodes and that player j can choose to lay down an edge (j, k) to another player k (possibly weighted), indicating a trust relationship; and suppose further that nodes can misrepresent trust information and create new ("fake") nodes. Various algorithms can be defined to compute pairwise trust between nodes, and their informativeness and manipulability can be compared. For instance, the EigenTrust algorithm[21] is vulnerable to Sybil attacks; one player can lay down multiple fake nodes that pump reputation flow in its direction. Other algorithms are more robust. Consider, for instance, defining pairwise trust i-j (i's trust of j) as the number of hops on the shortest path from i to j in an unweighted directed graph. Player j cannot reduce the i-j path length and improve the i-j trust by adding fake nodes and (directed) edges, as these nodes and edges can only affect shortest paths that flow *through* j and thus must leave the i-j path unaffected.[4] A similar argument establishes that node j cannot reduce the i-k trust for any node k that i trusts *more* than j. The only paths affected would be those that go through j and therefore ultimately reach nodes less trusted by i than j.

One outstanding challenge in the area of trust metrics is to find a satisfactory definition of informativeness; current axiomatic approaches appear unsatisfactory in this regard. With such a definition in hand, tensions between robustness and informativeness could be explored and perhaps also allow for mitigating factors such as the presence of other-regarding and altruistic actors within peering systems.

Security and Privacy

Research in security and privacy has been a central theme in academic computer science for more than 30 years. But although many clever algorithms, protocols, and devices have been rigorously analyzed, experimentally deployed, and even commercially developed, few of these solutions are in widespread use.

In 2001, Anderson initiated a new security-research direction by questioning the tacit assumption that security is primarily a technical problem. "[I]nformation insecurity is at least as much due to perverse incentives" as to technological failure, he said." Many of the problems can be explained more clearly and convincingly using the language of microeconomics: network externalities, asymmetric information, moral hazard, adverse selection, liability dumping, and the tragedy of the commons."[5] In the years since Anderson's seminal paper appeared, most aspects of security and privacy research have been suffused with economic considerations. We confine ourselves here to a brief discussion of the role of economics in user privacy, spam, and digital content distribution. More information can be found, for example, at the Web site of the Workshop on Economics of Information Security.[34]

That the increase in e-commerce has coincided with an erosion of consumer privacy seems paradoxical. After all, the ability to shop without having to stand face-to-face with others in a store seems privacy enhancing. Encryption, pseudonyms, electronic cash, and numerous other techniques purport to enable online consumers to protect their personal information and their identities. In practice, however, the trend has been for online merchants to ask for more information about their customers and for customers, even sophisticated ones who claim to value privacy, to provide this information when asked. Acquisti[1] offers the following explanation: "Behind a privacy intrusion there is often an economic trade-off. ... [I]ndividuals want to avoid the misuse of the information they pass along to others, but they also want to share enough information to achieve satisfactory interactions; organizations want to know more about the parties with which they interact, but they do not want to alienate them with policies deemed as intrusive."

That the increase in e-commerce has coincided with an erosion of consumer privacy seems paradoxical. After all, the ability to shop without having to stand face-to-face with others in a store seems privacy enhancing.

Behavioral economics combined with experimental psychology yields the following general explanation for privacy erosion. Even "sophisticated" consumers who value privacy will often compromise it to improve their position in an ongoing transaction. They know that loss of control over their private information poses a long-term threat, but they cannot assess the long-term risk accurately enough to compare it to the short-term gain in the ongoing transaction.[2]

Although not strictly an "e-commerce" issue, privacy in public databases[9] is relevant to this discussion. In publicly available, sanitized aggregations of sensitive data about individuals, the canonical example of which is a census database, there is a natural tension between *utility* of the collection and *privacy* of the individuals, and communities often opt for utility.

Among all forms of unwanted communication, email spam has received the most attention. But the phenomenon also includes link spam (which degrades search quality), shilling (which degrades reviewer and recommender systems), and click fraud. Given this rogue's gallery, there is widespread agreement that the incentive structure of next-generation network services must be designed more carefully.[13] Techniques for incentive-based prevention of unwanted communication include pricing via processing,[14] attention bonds,[24] and click-based learning algorithms.[19]

Nowhere is the inadequacy of a purely technical approach to information security clearer than in the realm of copyright enforcement. Digital-content distributors do have some control over the prices of their products and the flexibility with which purchased products can be used (for example, the extent to which they can be copied, shared, or modified), but although more flexible products may fetch higher prices they may also show dampened sales. Taking into account the fact that some digital environments are highly permeable (for example, they have higher rates of social contact or network bandwidth), Bergemann et al.[7] show the following:

▸ If users are homogenous, there is a critical permeability level up to which

all users buy the product, and the optimal price and flexibility levels decrease as permeability increases. When permeability rises above the critical value, the number of users who buy decreases, and the optimal price and flexibility stay constant.

▸ A platform vendor who also sells digital content (which he licenses from copyright owners) will find it optimal to allow very flexible use by platform customers and to set a low price for content; most of the vendor's profit is from platform sales. Copyright owners' revenue is low, because prices are low and sharing of content is common.

▸ Tension between the platform vendor and copyright owner is observed in the real world. In 2005, the *Financial Times* quoted a music executive as saying, "Our music is not something to be given away to sell iPods."

Conclusion

Nearly all mass-market services—email, Web search, social networks, recommendations, user content, and ad networks, for example—seem beset by attempts to manipulate and distort. New services, such as the semantic Web, are assumed to be immune only while they remain niche. Yet to date the design of Internet protocols and services, including monetization efforts, have often been guided by technology rather than economics; and users have been modeled in caricature as either cooperative or malicious, the later to be dealt with by security measures. In truth, few users are benevolent and fewer still are vandals. Even the most vilified of participants—spammers—are acting in response to rational economic considerations. Understanding how to align the incentives of participants with systemwide objectives is fundamental to the design of the next generation of Web-scale services. Increasingly, design teams will require dual expertise in social science and computer science, adding competence in economics, sociology, and psychology to more traditionally recognized requirements such as algorithms, interfaces, systems, machine learning, and optimization.

In this article we focused on computational challenges in Internet-based (specifically Web-based) commerce. Yet people are starting to use smartphones and other handheld devices to do much of what they used to do with Internet-connected desktop or laptop PCs. To what extent these handheld devices will ultimately behave like networked PCs is unclear, and thus we leave to a future article the needed discussion of computational challenges in mobile commerce.

Acknowledgments

Joan Feigenbaum was supported in part by NSF grants CNS-0331548, CNS-0428422, and IIS-0534052. David Parkes was supported in part by NSF grants DMS-0631636, IIS-0534620, IIS-0238147, an Alfred P. Sloan Research Fellowship, and Yahoo! Research and Microsoft Research awards; he gratefully acknowledges helpful conversations with Yochai Benkler, Chris Kelty, Johan Pouwelse, and Sven Seuken.

References

1. Acquisti, A. *The Economics of Privacy*; www.heinz.cmu.edu/~acquisti/economics-privacy.htm.
2. Acquisti, A. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th Electronic Commerce Conference*. ACM (2004), 21–29; www.heinz.cmu.edu/~acquisti/papers/privacy-gratification.pdf.
3. Adar, E. and Huberman, B. Free riding on Gnutella. *First Monday 5* (2000), 2.
4. Altman, A. and Tennenholtz, M. An axiomatic approach to personalized ranking systems. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence* (2007), 1187–1192.
5. Anderson, R. Why information security is hard—An economic perspective. In *Proceedings of the 17th Annual Computer Security Applications Conference* (2001), 358–365; www.cl.cam.ac.uk/~rja14/Papers/econ.pdf.
6. Benkler, Y. Coase's penguin, or, Linux and the nature of the firm. *The Yale Law Journal 112* (2002), 369–446.
7. Bergemann, D., Eisenbach, T., Feigenbaum, J., and Shenker, S. Flexibility as an instrument in digital rights management. In *Proceedings of the 2005 Workshop on Economics of Information Security*; http://infosecon.net/workshop/pdf/50.pdf
8. Bolton, G., Greiner, B., and Ockenfels, A. Engineering trust: Strategic behavior and the production of reputation information. Working paper, Harvard Business School, 2007.
9. Chawla, S., Dwork, C., McSherry, F., Smith, A., and Wee, H. Toward privacy in public databases. In *Proceedings of the 2nd Theoretical Cryptography Conference*. Springer LNCS 3378 (2005), 363–385; http://pages.cs.wisc.edu/~shuchi/papers/privacy-TCC.pdf
10. Chen, Y., Fortnow, L., Lambert, N., Pennock, D.M., and Wortman, J. Complexity of combinatorial market makers. In *Proceedings of the 9th Conference on Electronic Commerce. ACM* (2008), 190–199.
11. Chen, Y. Goel, S., and Pennock, D.M. Pricing combinatorial markets for tournaments. In *Proceedings of the 40th Symposium on Theory of Computation* (2008), 305–314.
12. Cramton, P., Shoham, Y., and Steinberg, R., Eds. *Combinatorial Auctions*. MIT Press, Cambridge, MA, 2006.
13. Dhajima, R., Tygar, D., and Heart, M. Why phishing works. In *Proceedings of the Conference on Human Factors in Computer Systems*. ACM (2006), 581–590.
14. Dwork, C. and Naor, M. Pricing via processing or combating junk mail. *Crypto '92*. Springer LNCS 740 (1992) 139–147; www.wisdom.weizmann.ac.il/~naor/PAPERS/pvp.ps.
15. Edelman, B., Ostrovsky, M., and Schwarz, M. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. *American Economic Review 97* (2007), 242–259.
16. Fehr, E. and Camerer, C. Measuring social norms and preferences using experimental games: A guide for social scientists. In *Foundations of Human Sociality*, Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., and Gintis, H. Eds. Oxford University Press, Oxford, U.K., 2004.
17. Frey, B.S. and Jegen, R. Motivation crowding theory. *Journal of Economic Surveys 15* (2001), 589–611.
18. Hanson, R. Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets 1* (2007), 3–15.
19. Immorlica, N., Jain, K., Mahdian, M., and Talwar, K. Click fraud resistant methods for learning click-through rates. In *Proceedings of the 1st Workshop on Internet and Network Economics 3828*. Springer LNCS (2005) 34–45; www.ece.northwestern.edu/~nickle/pubs/clickfraud.pdf.
20. Jackson, M.O. Mechanism theory. *Encyclopedia of Life Support Systems*. Derigs, U., Ed. EOLSS Publishers, Oxford, U.K., 2003.
21. Kamvar, S.D., Schlosser, M.T., and Garcia-Molina, H. The EigenTrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th International World Wide Web Conference*, (2003), 640–651; http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.3109.
22. Kash, I.A., Friedman, E.J., and Halpern, J.Y. Optimizing scrip systems: Efficiency, crashes, boarders, and altruists. In *Proceedings of the 8th Conference on Electronic Commerce*. ACM (2007), 305–315.
23. Lambert, N., Langford, J., Wortman, J., Chen, Y., Reeves, D., Shoham, Y., and Pennock, D.M. Self-financed wagering mechanisms for forecasting. In *Proceedings of the 9th Conference on Electronic Commerce*. ACM (2008), 170–179.
24. Loder, T., Van Alstyne, M., and Wash, R. An economic response to unsolicited communication. *Advances in Economic Analysis and Policy 1* (2006), Article 2; www.bepress.com/bejeap/advances/vol6/iss1/art2/.
25. Nisan, N., Roughgarden, T., Tardos, E., and Vazirani, V.V., Eds. *Algorithmic Game Theory*. Cambridge University Press, 2007.
26. Pennock, D.M. A dynamic pari-mutuel market for hedging, wagering, and information aggregation. In *Proceedings of the 5th Conference on Electronic Commerce*. ACM, 2004, 170–179.
27. Pouwelse, J.A., Garbacki, P., Wang, J., Bakker, A., Yang, J., Iosup, A., Epema, D.H.J., Reinders, M.J.T, van Steen, M.R., and Sips, H.J. TRIBLER: A social-based peer-to-peer system. *Concurrency and Computation: Practice and Experience 20* (2008), 127–138.
28. Resnick, P. and Sami, R. The influence-limiter: Provably manipulation-resistant recommender systems. *Recommender Systems Conference*. ACM, 2007, 25–32.
29. Resnick, P., Zeckhauser, R., Swanson, J., and Lockwood, K. The value of reputation on eBay: A controlled experiment. *Experimental Economics 9* (2006), 79–101.
30. Sandholm, T. Expressive commerce and its application to sourcing: How we conducted $35 billion of generalized combinatorial auctions. *AI Magazine 28* (2007), 45–58.
31. Varian, H.R. Position auctions. *International Journal of Industrial Organization 25* (2007), 1163–1178.
32. Winkler, R., Munoz, J., Cervera, J., Bernardo, J., Blattenberger, G., Kadane, J., Lindley, D., Murphy, A., Oliver, R., and Rios-Insua, D. Scoring rules and the evaluation of probabilities. *TEST 5* (1996), 1–60.
33. Workshops on Ad Auctions; http://research.yahoo.com/workshops/ad-auctions-2008/pastworkshops.html.
34. Workshops on Economics of Information Security; http://weis2008.econinfosec.org/.
35. Zhang, H. and Parkes, D.C. Value-based policy teaching with active indirect elicitation. In *Proceedings of the 23rd Conference on Artificial Intelligence*. (2008), 208–214.

Joan Feigenbaum (joan.feigenbaum@yale.edu) is the Grace Murray Hopper Professor of Computer Science at Yale University, New Haven, CT.

David C. Parkes (parkes@eecs.harvard.edu) is Gordon McKay Professor of Computer Science at Harvard University, Cambridge, MA.

David M. Pennock (pennockd@yahoo-inc.com) is a principal research scientist at Yahoo! Research, New York, NY.

© 2009 ACM 0001-0782/09/0100 $5.00