The Diffusion of Networking Technologies

Sharon Goldberg
Boston University
goldbe@cs.bu.edu *

Zhenming Liu
Princeton University
zhenming@cs.princeton.edu †

August 27, 2013

Abstract

There has been significant interest on the impact of cascade effects on the diffusion of networking technology upgrades in the Internet. Thinking of the global Internet as a graph, where each node represents an economically-motivated Internet Service Provider (ISP), a key problem is to determine the smallest set of nodes that can trigger a cascade that causes every other node in the graph to adopt the protocol. We design the first approximation algorithm with a provable performance guarantee for this problem, in a model that captures the following key issue: a node's decision to upgrade should be influenced by the decisions of the remote nodes it wishes to communicate with.

Keywords. Approximation algorithms, diffusion processes, networks.

Bibliographic note. An extended abstract of this work appeared in SODA'13.

1 Introduction

There has been significant interest in the impact of cascade effects on the diffusion of technology upgrades in the Internet [2, 6, 10, 14, 15, 18, 21, 24, 25, 33]. Thinking of the global Internet as a graph, where each node represents an independent, economically-motivated *autonomous system* (AS), *e.g.,* AT&T, Google, Telecom Italia, or Bank of America, a key problem is to determine the set of nodes that governments and regulatory groups should target as early adopters of the new technology, with the goal of triggering a cascade that causes more and more nodes to voluntarily adopt the new technology [6, 18, 21, 25]. Given the effort and expense required to target ASes as early adopters, a natural objective (that has appeared in both the networking literature [2,6,18] and also that of viral marketing [13,27]) is to find the smallest possible *seedset* of early adopters that could drive a cascade of adoption; doing this would shed light on how best to manage the upgrade from insecure routing [5] to secure routing [28,29], or from IPv4 to IPv6 [11], or the deployment of technology upgrades like QoS [22], fault localization [3], and denial of service prevention [38].

Thus far, the literature has offered only heuristic solutions to the problem of the diffusion of networking technologies. In this paper, we design the first approximation algorithm with a provable performance guarantee that optimizes the selection of early adopter nodes, in a model of that captures the following important property: the technologies we study only allow a pair of nodes to communicate if they have a *path* between them consisting of nodes that also use the new technology [3, 6, 18, 22, 23, 29, 38].

Model. Consider a graph $G(V, E)$ that represents the internetwork. We use the following progressive process to model the diffusion of a new technology: a node starts out as inactive (using an older version

*Supported by NSF grant S-1017907 and a gift from Cisco.
†Supported by NSF grants CCF-0915922 and IIS-0964473.

of the technology) and *activates* (adopts the new, improved technology) once it obtains sufficient utility from the new technology. Once a node is active, it can never become inactive. To model the cost of technology deployment, the standard approach [20, 27, 36] is to associate a threshold $\theta(u)$ with each node u that determines how large its utility should be before it is willing to activate. A node's utility depends on the *size of the connected components of active nodes adjacent to u in G*. Thus, node u activates if the connected component containing u in the subgraph induced in G by nodes $\{v : v \in V, \text{Node } v \text{ is active}\} \cup \{u\}$ has size at least $\theta(u)$. We study the following optimization problem:

> *Given G and the threshold function $\theta : V \to \{2, ..., |V|\}$, what is the smallest* feasible seedset *$S \subseteq V$ such that if nodes in S activate, then all remaining nodes in V eventually activate?*

This model of node utility captures two key ideas:

1. the traditional notion of "direct network externalities/effects" from economics [16, 26], marketing [4] and other areas [32], that supposes an active node that is part of a network of k active nodes has utility that scales with k, and
2. the fact that we are interested in networking technologies that only allow a pair of active nodes $u, v \in G(V, E)$ to communicate if there is path of active nodes between them in G.

Our model has much in common with the vast literature on diffusion of innovations, and especially the linear threshold model for diffusion in social networks, articulated by Kempe *et al.* [27] and extensively studied in many other works. Indeed, the two models diverge only in the choice of the utility function; ours is non-local, while theirs depends the (weighted) sum of a node's active *neighbors* in G. Meanwhile, the non-local nature of our utility function has much in common with the classic literature on "direct network externalities/effects" [4, 16, 26, 32] with the important difference that these classic models ignore the underlying graph structure, and instead assume that utility depends on only a *count* of the active nodes. We shall now see that these differences have a substantial effect on our algorithmic results.

1.1 Our results.

Our main result is an approximation algorithm based on linear programming that consists of two phases. The first is a linearization phase that exploits combinatorial properties to encode our problem as an integer program (IP) with a 2-approximate solution, while the second is a randomized rounding algorithm that operates by restricting our search space to *connected seedsets*, *i.e.,* seedsets that induce a connected subgraph of G . We have:

Theorem 1.1 (Main result)**.** *Consider a networking technology diffusion problem $\{G(V, E), \theta\}$ where the smallest seedset has size* opt*, the graph has diameter r (*i.e., *r is the length of "longest shortest path" in G), and there are at most ℓ possible threshold values,* i.e., *$\theta : V \to \{\theta_1, ..., \theta_\ell\}$. Then there is a polynomial time algorithm that returns a seedset S of size $O(r\ell \log |V| \cdot \text{opt})$.*

Relationship to the linear threshold model in social networks. Our main result highlights the major algorithmic difference between our work and the linear threshold model in social networks [27]. In the social network setting, Chen [7] showed that this problem is devastatingly hard, even when $r, \ell = O(1)$; to avoid this discouraging lower bound, variations of the problem that exploit submodular properties of the objective have been considered (*e.g.,* where thresholds are chosen uniformly at random [27] or see [8, 34] and references therein). Indeed, the ubiquity of these techniques seems to suggest that diffusion problems are tractable *only* when the objective exhibits submodularity properties. Our work provides an interesting counterpoint: our positive result does not rely on submodular optimization, and we show that the influence function in our problem, and its natural variations, lacks submodularity properties.

Dependencies on r, ℓ, and $\log|V|$ are necessary. Removing our algorithm's dependence on r, ℓ, or $\log|V|$ is likely to require a very different set of techniques because of the following barriers:

1. *Computational barrier.* We use a reduction from Set Cover to show that our problem does not admit any $o(\ln|V|)$-approximation algorithm, even if $r, \ell = O(1)$.
2. *Combinatorial barrier.* We present a family of problem instances that prove that any algorithm that returns a connected seedset must pay an $\Omega(r)$-increase in the size of the seedset in the worst case.
3. *Integrality gap.* The linear program we use has an integrality gap of $\Omega(\ell)$ so that our rounding algorithm is asymptotically optimal in ℓ.

Quality of approximation. We interpret the quality of our approximation for typical problem instances. The motivation for our problem is to help centralized authority (*e.g.,* a government, a regulatory group) determine the right set of autonomous systems (ASes) in the Internet to target as early adopters for an upgrade to a new networking technology [17, 31]. We comment on the asymptotic order of r and ℓ when a centralized authority executes this algorithm. The graph G is the Internet's AS-level graph, which is growing over time, with diameter r that does not exceed $O(\log|V|)$ (see, *e.g.,* [30]). We remark that the empirical data we have about the Internet's AS-level topology [1, 9, 12, 37] is the result of a long line of Internet measurement research [35]. On the other hand, obtaining empirical data on ASes' thresholds is still subject to ongoing research [17,19]. The following natural assumption and practical constraint restrict the threshold granularity ℓ: (a) ASes should not be sensitive to small changes in utility (*e.g.,* 1000 nodes vs. 1001 nodes), and that (b) in practice, it is infeasible for a centralized authority to obtain information about $\theta(u)$ from every AS u in the Internet, both because this business information is kept private and because, perhaps more importantly, many of these nodes are in distant and possibly uncooperative countries. Thus, thresholds should be chosen from a geometric progression $\{(1+\epsilon), (1+\epsilon)^2, ..., (1+\epsilon)^\ell\}$ or even restricted to a constant size set $\{5\%, 10\%, 15\%, 20\%, 30\%, 50\%\}$ as in [6, 18, 33] so that $\ell = O(\log|V|)$. Our approximation ratio is therefore polylogarithmic in $|V|$ in this context.

New heuristics. Finally, we remark that our IP formulation might also be a promising starting point for the design of new heuristics. Indeed, we ran a generic IP solver to find seedsets on problem instances of non-trivial size; the seedsets we found were often substantially better than those returned by several natural heuristics (including those used in [2, 6, 18]).

Bibliographical notes. An extended abstract of this work appeared at the Symposium for Discrete Algorithms (SODA'13), and the full version of this work is available as arXiv report 1202.2928 `http://arxiv.org/abs/1202.2928`. Video of a presentation of this work is available at `http://crcs.seas.harvard.edu/2012/11/02/monday-february-25-2013-sharon-goldberg-boston-university-on-tba/`.

References

[1] B. Augustin, B. Krishnamurthy, and W. Willinger. IXPs: Mapped? In *IMC*, 2009.

[2] I. Avramopoulos, M. Suchara, and J. Rexford. How small groups can secure interdomain routing. Technical report, Princeton University Comp. Sci., 2007.

[3] B. Barak, S. Goldberg, and D. Xiao. Protocols and lower bounds for failure localization in the Internet. In *IACR EUROCRYPT*, 2008.

[4] F. Bass. A new product growth model for consumer durables. *Management Science*, pages 215–27, 1969.

[5] K. Butler, T. Farley, P. McDaniel, and J. Rexford. A survey of BGP security issues and solutions. *Proceedings of the IEEE*, 2010.

[6] H. Chang, D. Dash, A. Perrig, and H. Zhang. Modeling adoptability of secure BGP protocol. In *Sigcomm*, 2006.

[7] N. Chen. On the approximability of influence social networks. In *ACM-SIAM Symposium on Discrete Algorithms*, 2008.

[8] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincón, X. Sun, Y. Wang, W. W., and Y. Y. Influence maximization in social networks when negative opinions may emerge and propagate. In *SDM*, 2011.

[9] Y.-J. Chi, R. Oliveira, and L. Zhang. Cyclops: The Internet AS-level observatory. *ACM SIGCOMM CCR*, 2008.

[10] D. D. Clark, J. Wroclawski, K. R. Sollins, and R. Braden. Tussle in cyberspace: defining tomorrow's Internet. *Trans. on Networking*, 2005.

[11] S. Deering and R. Hinden. RFC 2460: Internet Protocol, Version 6 (IPv6) Specification. http://www.ietf.org/rfc/rfc2460.txt, 1998.

[12] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, and kc claffy. AS relationships: Inference and validation. *ACM SIGCOMM Computer Communication Review*, JAN 2007.

[13] P. Domingos and M. Richardson. Mining the network value of customers. In *Proc. 7th Conf on Knowledge discovery and data mining*, KDD '01, pages 57–66, New York, NY, USA, 2001. ACM.

[14] B. Edelman. Running out of numbers: Scarcity of ip addresses and what to do about it. Technical report, Harvard Business School, 2009.

[15] H. A. Elmore, L. J. Camp, and B. P. Stephens. Diffusion and adoption of ipv6 in the arin region. In *Workshop on the Economics of Internet Security*, 2008.

[16] J. Farrell and G. Saloner. Standardization, compatibility, and innovation. *The RAND Journal of Economics*, pages 70–83, 1985.

[17] FCC. The communications security, reliability and interoperability council iii working group 6: Secure bgp deployment. Technical report, March 2012.

[18] P. Gill, M. Schapira, and S. Goldberg. Let the market drive deployment: A strategy for transistioning to BGP security. *SIGCOMM'11*, 2011.

[19] P. Gill, M. Schapira, and S. Goldberg. Modeling on quicksand: dealing with the scarcity of ground truth in interdomain routing data. *ACM SIGCOMM Computer Communication Review*, 42(1):40–46, 2012.

[20] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, May 1978.

[21] R. Guérin and K. Hosanagar. Fostering ipv6 migration through network quality differentials. *SIGCOMM Comput. Commun. Rev.*, 40:17–25, June 2010.

[22] M. Howarth, P. Flegkas, G. Pavlou, N. Wang, P. Trimintzios, D. Griffin, J. Griem, M. Boucadair, P. Morand, H. Asgari, and P. Georgatsos. Provisioning for inter-domain quality of service: the MESCAL approach. *IEEE Communications Magazine*, June 2005.

[23] G. Huston. Stacking it up: Experimental observations on the operation of dual stack services. In *NANOG'52*, 2011.

[24] Y. Jin, S. Sen, R. Guerin, K. Hosanager, and Z.-L. Zhang. Dynamics of competition between incumbent and emrging network technologies. *NetEcon*, 2008.

[25] D. Joseph, N. Shetty, J. Chuang, and I. Stoica. Modeling the adoption of new network architectures. In *CoNEXT'07: Conference on emerging Networking EXperiments and Technologies*, 2007.

[26] M. Katz and C. Shapiro. Network externalities, competition, and compatibility. *The American economic review*, 75(3):424–440, 1985.

[27] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *ACM SIGKDD*, 2003.

[28] S. Kent, C. Lynn, and K. Seo. Secure border gateway protocol (S-BGP). *JSAC*, 2000.

[29] M. Lepinski, editor. *BGPSEC Protocol Specification*. IETF Network Working Group, Internet-Draft, Mar. 2011. Available from `http://tools.ietf.org/html/draft-lepinski-bgpsec-protocol-00`.

[30] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2005.

[31] C. D. Marsan. White house issues ipv6 directive. *Network World*, September 28, 2010.

[32] B. Metcalfe. Metcalfe's law: A network becomes more valuable as it reaches more users. *InfoWorld*, 1995.

[33] A. Ozment and S. E. Schechter. Bootstrapping the adoption of internet security protocols. In *The Fifth Workshop on the Economics of Information Security (WEIS 2006).*, 2006.

[34] M. G. Rodriguez and B. Schölkopf. Influence maximization in continuous time diffusion networks. In *29th International Conference on Machine Learning (ICML)*, 2012.

[35] M. Roughan, W. Willinger, O. Maennel, D. Perouli, and R. Bush. 10 lessons from 10 years of measuring and modeling the internet's autonomous systems. *Selected Areas in Communications, IEEE Journal on*, 29(9):1810–1821, 2011.

[36] T. C. Schelling. *Micromotives andMacrobehavior*. Norton, 1978.

[37] Y. Shavitt and E. Shir. Dimes: Let the internet measure itself. *ACM SIGCOMM Computer Communication Review*, 35(5):71–74, 2005.

[38] A. Yaar, A. Perrig, and D. Song. SIFF: a stateless internet flow filter to mitigate ddos flooding attacks. *IEEE Symposium on Security and Privacy*, 2004.