AGENT-BASED MODELING OF OPEN SOURCE USING SWARM

Greg Madey
Computer Science & Engineering
University of Notre Dame
gmadey@nd.edu

Vincent Freeh
Computer Science & Engineering
University of Notre Dame
vin@nd.edu

Renee Tynan
Department of Management
University of Notre Dame
rtynan@nd.edu

Abstract

The open source software (OSS) development phenomenon appears to be a self-organizing process with emergent properties. Such processes are difficult to understand because emergent properties are by definition difficult to predict using traditional modeling and analytical techniques. An approach under evaluation is to use agent-based simulation techniques to study the OSS phenomenon. We are using the Swarm library and the Java programming language to model the self-organizing processes seen in the OSS phenomenon.

Keywords: Agent-based modeling, open source software, swarm, simulation, emergence

Introduction

The Open Source Software (OSS) movement is a prototypical example of a decentralized self-organizing process. At the highest level of the phenomenon, there is no central control or central planning. It challenges conventional economic assumptions, it turns conventional software engineering and project management principles inside out, it threatens traditional proprietary software business strategies, and it presents new legal and government policy questions regarding software licensing and intellectual property. Moreover, OSS is a major component of the IT infrastructure enabling global e-Commerce. Open source software including BIND, sendmail, Apache, Linux, INN, GNU utilities, MySQL, PostgreSQL, and Perl are critical components of the Internet. They enable major services hosted on the Internet, e.g., e-mail, WWW, e-Commerce, domain name lookup. The *Netcraft.com* survey of 36.6 million web servers worldwide reports an over 60% market share for the open-source web-server ApacheENRfu(Netcraft, 2002).

Open Source Software Development

Open source software, by its name, implies that users have access to the source code. This distinguishes it from the recent common practice by commercial software publishers of only releasing the binary executable versions of the software. Most open source software is also distributed at no cost with limited restrictions on how it can be used; hence the term "free" when used to describe open source carries two meanings: 1) free of cost and 2) free to do with the software as you wish (i.e., most importantly — free to read, modify, and re-distribute the code).

Case studies documenting the open source software development model, although often sympathetic to that model, point to potential lessons and benefits that may be of value to IT managers ENRfu(O'Reilly, 1999; Wu, 2001). It is claimed that open source development produces more bug-free code, faster, than closed proprietary code. Open source software development teams are generally comprised of volunteers working not for monetary return, but for the enjoyment and pride of being part of a successful virtual software development project ENRfu(Bollinger, 1999; Edwards, 1998; Feller, 2000; Fielding, 1999; Hars, 2001; O'Reilly, 1999; Scacchi, 2002; Wang, 2001; Wu, 2001). Team members often come from around the world and rarely meet one another face-to-face. The open source projects are self-organized, decentralized, employ extremely rapid code evolution, massive peer code review, and rapid releases of prototype code. Some innovative firms are beginning to support open source development internally — still releasing for free — as a new business strategy ENRfu(Hecker, 1999; Ousterhout, 1999; Pavelick, 2000; Sharma, 2002). Many of these practices are counter intuitive and the opposite of what conventional software engineering held as the correct processes for the production of high quality code or what business strategists considered viable practices ENRfu(Bollinger, 1999; Charles, 1998; Edwards, 1998; Fielding, 1999; Hecker, 1999; Lawrence, 1998; O'Reilly, 1999; Ousterhout, 1999; Raymond, 1999; Sanders, 1998; Torvalds, 1999)

Several important studies on the OSS phenomenon have been conducted and provide important foundations for this study. Feller and Fitzgerald ENRfu(Feller, 2000) developed a research framework and analyzed the OSS phenomenon. Hars and Ou ENRfu(Hars, 2001) surveyed OSS developers and reported on their motivations for participation in OSS projects. Scacchi ENRfu(Scacchi, 2002) has an extensive ongoing study of the socio-technical processes associated with OSS development work practices. Wolf, et al recently released the results of a survey of 526 OSS developers from SourceForge, and 134 participants on the Linux kernel mailing list, reporting on developer motivations and attitudes ENRfu(Wolf, 2002).

Agent-Based Modeling of OSS

Prior research suggests that the OSS phenomenon can be considered a complex, self-organizing system ENRfu(Axelrod, 1999; Barabasi, 2002; Barabasi, 2000; Faloutsos, 1999; Holland, 1998; Huberman, 1999; Johnson, 2001; Kuwabara, 2000). These systems are typically comprised of large numbers of locally interacting elements. Although the rules describing those local interactions may be few and simple, often unexpected and difficult to predict global properties emerge. Many investigators of such systems have found that they can only be understood through modeling, and specifically through what some researchers call iconological modeling and structural modeling ENRfu(Eve, 1997; Harvey, 1997; Kiel, 1997; Smith, 1997). The goals of these simulation approaches can be achieved using the agent-based approach pioneered by Schelling ENRfu(1978), and advanced by Axelrod ENRfu(1984), Epstein and Axtell ENRfu(1996), Resnick ENRfu(1994), Cohen et al ENRfu(1998), and many others. Using such models enables social science researchers to address the modeling difficulties of this field; social processes are complex, they have sensitive dependence on behaviors of individual heterogeneous agents (which are not omniscient rational utility optimizers), and they cannot always be modeled as systems assumed to be in static equilibrium ENRfu(Axelrod, 1997a; Byrne, 1997; Epstein, 1996; Gaylord, 1998; Goldspink, 2002). The goal of these models is not to predict, but to develop an understanding of how and why the elements of the system are able to produce emergent behavior ENRfu(Axelrod, 1997a; Axelrod, 1997b; Harvey, 1997; Holland, 1998). This understanding can be obtained by discovering the rules and mechanisms that control agent interactions. In the simulations of OSS processes, the agents could be the developers, the projects, and clusters of projects.

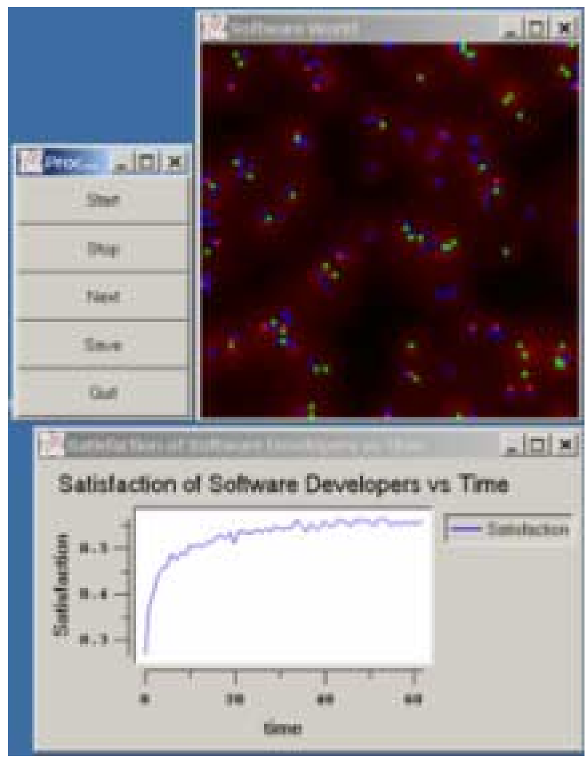

Figure 1. Agent-Based Swarm Simulation of OSS Developers; Developer Satisfaction

The simulations will be built using the Java programming language and the agent-based modeling library Swarm ENRfu(Minar, 1996; Swarm_Development_Group, 2000; Terna, 1998). The Swarm library enables discrete event, multi-agent simulations. In these simulations, agents are organized into "swarms". Several swarms can be nested into another swarm in a hierarchical manner. Thus, OSS developers can be grouped into a "project swarm", which in turn can be grouped into a "cluster swarm", which would be part of the entire "OSS development swarm". Both the Swarm library and the Java language are object-oriented, providing the object-oriented programming benefits of attribute and behavior encapsulation, information hiding, and inheritance ENRfu(Epstein, 1996; Minar, 1996; Swarm_Development_Group, 2000; Terna, 1998). A screen shot of some of the GUI components associated with a prototype Swarm simulation of OSS developers self-organizing into groups is displayed in Figure 1. The simulation will model OSS developers as nodes and joint project membership as links. Other node-link representations are possible. Properties of the OSS movement, collected at SourceForge.org, such as distributions of project sizes and distributions of project membership (by developers) will be compared with distributions of the same relationships from the simulation. We report elsewhere that those relationships have power-law relationships (Madey, 2002). One goal would be to search for simulation parameters that generate similar power-law distributions.

References

ENBbuAxelrod, R. *The Evolution of Cooperation*, Basic Books, New York, 1984.

Axelrod, R. "Advancing the Art of Simulation in the Social Sciences," *Complexity* (3:2), 1997a, pp. 16-22.

Axelrod, R. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton University Press, Princeton, 1997b.

Axelrod, R., M. Cohen *Harnessing Complexity: Organizational Implications of a Scientific Frontier*, The Free Press, New York, 1999.

Barabasi, A.-L. *Linked: The New Science of Networks*, Perseus, Boston, 2002.

Barabasi, A.L., Albert, R., Jeong, H "Scale-free Characteristics of Random Networks: The Topology of the World Wide Web," *Physica A*, 2000, pp. 69-77.

Bollinger, T. "Linux and Open-Source Success: Interview with Eric. S. Raymond," *IEEE Computer*, 1999, pp. 85-89.

Byrne, D. "Simulation - A Way Forward," *Sociological Research Online* (2:2), 1997,

Charles, J. "Open Source: Netscape Pops the Hood," *IEEE Software*, 1998, pp. 79-82.

Cohen, M.D., Riolo, R.L., Axelrod, R. "The Emergence of Social Organization in the Prisoner's Dilemma: How Context-Preservation and other Factors Promote Cooperation," University of Michigan, http://www.pscs.umich.edu/RESEARCH/pscs-tr.html, 1998.

Edwards, J. "The Changing Face of Freeware," *Computer*, 1998, pp. 11-13.

Epstein, J.M., R. Axtell *Growing Artificial Societies: Social Science from the Bottom Up*, The MIT Press, Cambridge, MA, 1996.

Eve, R., S. Horsfall, M. Lee "Chaos, Complexity and Sociology," Thousand Oaks, 1997

Faloutsos, M., Faloutsos, P., Faloutsos, C. "On Power-Law Relationships of the Internet Topology," *Proceedings of the SIGCOMM'99*, Cambridge, MA, 1999, pp. 251-262.

Feller, J.F., B. "A Framework Analysis of the Open Source Software Development Paradigm," *Proceedings of the ICIS 2000*, Brisbane, Australia, 2000, pp. 58-69.

Fielding, R.T. "Shared Leadership in the Apache Project," *Communications of the ACM* (42:4), 1999, pp. 42-43.

Gaylord, R.J., L.J. L'Andria *Simulating Society: A Mathematica Toolkit for Modeling Socioeconomic Behavior*, Springer-Verlag (TELOS), New York, 1998.

Goldspink, C. "Methodological Implications of Complex Systems Approaches to Sociality: Simulation as a Foundation for Knowledge," *Journal of Artificial Societies and Social Simulation* (5:1), 2002, pp. 1-19.

Hars, A., Ou, S. "Working for free? Motivations of participating in Open Source Projects," *Proceedings of the Hawaii International Conference on Systems Sciences*, 2001

Harvey, D., M. Reed "Social Science as the Study of Complex Systems," In *Chaos theory in the Social Sciences: Foundations and Applications*, L. D. Kiel, E. Elliot (Ed.), The University of Michigan Press, An Arbor, 1997, pp. 295-323.

Hecker, F. "Setting up Shop: The Business of Open-Source Software," *IEEE Software*, 1999, pp. 45-51.

Holland, J. *Emergence: From Chaos to Order*, Addison-Wesley Publishing Company, Reading, MA, 1998.

Huberman, B.A., Adamic, L. A "Growth Dynamics of the World Wide Web," *Nature* (401), 1999, pp. 131.

Johnson, S. *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*, Scribner, New York, 2001.

Kiel, L.D., E. Elliot *Chaos Theory in the Social Sciences: Foundations and Applications*, The University of Michigan Press, Ann Arbor, 1997.

Kuwabara, K. "Linux: A Bazaar at the Edge of chaos," *First Monday* (5:3), 2000, pp. 1-68.

Lawrence, D. "Internetnews Server: Inside an Open Source Project," *IEEE Internet Computing*, 1998, pp. 49-52.

Madey, G., Freeh, V., Tynan R. *The Open Source Software Development Phenomenon: An Analysis Based on Social Network Theory*, AMCIS2002, Dallas, TX, 2002.

Minar, N., R. Burkhart, C. Langton, M. Askenzi "The Swarm Simulation System: A Toolkit for Building Multi-Agent Simulations,", 1996, pp. 1-11.

Netcraft "Netcraft.com Web Server survey," (2002:Feb. 16, 2002), 2002

O'Reilly, T. "Lessons from Open-Source Software Development," *Communications of the ACM* (42:4), 1999, pp. 33-37.
Ousterhout, J. "Free Software Needs Profit," *Communications of the ACM* (42:4), 1999, pp. 44-45.
Pavelick, R. *Embracing Insanity: Open Source Software Development*, SAMS Publishing, Indianapolis, IN, 2000.
Raymond, E.S. *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, O'Reilly, Sebastopol, CA, 1999.
Resnick, M. *Turtles, Termites, and Traffic Jams*, The MIT Press, Cambridge, MA, 1994.
Sanders, J. "Linux, Open Source, and Software's Future," *IEEE Software*, 1998, pp. 88-91.
Scacchi, W. "Understanding the Requirements for Developing Open Source Software Systems," *IEE Proceedings - Software* (In press), 2002
Schelling, T. *Micromotives and Macrobehavior*, W. W. Norton, New York, 1978.
Sharma, S., Sugumaram, V., Rajagopalan, B. "A Framework for Creating Hybrid-Open Source Software Communities," *Information Systems Journal* (12:1), 2002, pp. 7-25.
Smith, T. "Nonlinear Dynamics and the Micro-Macro Bridge," In *Chaos, Complexity and Sociology*, R. Eve, S. Horsfall, M. Lee (Ed.), Sage Publications, Thousand Oaks, 1997, pp. 52-78.
Swarm_Development_Group "Brief Overview of Swarm,", 2000
Terna, P. "Simulation Tools for Social Scientists: Building Agent Based Models with SWARM," *Journal of Artificial Societies and Social Simulation* (1:2), 1998, pp. 1-12.
Torvalds, L. "The Linux Edge," *Communications of the ACM* (42:4), 1999, pp. 38-39.
Wang, H., Whang, C. "Open Source Software: A Status Report," *IEEE Software*, 2001, pp. 90-95.
Wolf, B., Karim, R.L., Bates, J. "Hacker Survey," Boston Consulting Group, 2002.
Wu, M.W., Lin, Y. D "Open Source Development: An Overview," *IEEE Computer*, 2001, pp. 33-38.