Open access • Journal Article • DOI:10.2139/SSRN.290305

How Open Source Software Works: 'Free' User-to-User Assistance?

Eric von Hippel, Karim R. Lakhani

Institutions: Massachusetts Institute of Technology, Harvard University

Published on: 01 May 2000 - Social Science Research Network

Topics: User assistance, Software development, Software as a service, Software system and Crowdsourcing software development

ELSEVIER

Research Policy 32 (2003) 923–943

research policy

www.elsevier.com/locate/econbase

How open source software works: "free" user-to-user assistance

Karim R. Lakhani*, Eric von Hippel

MIT Sloan School of Management, 50 Memorial Drive, Cambridge, MA 02142, USA

Received 14 November 2001; received in revised form 10 July 2002; accepted 12 July 2002

Abstract

Research into free and open source software development projects has so far largely focused on how the major tasks of software development are organized and motivated. But a complete project requires the execution of "mundane but necessary" tasks as well. In this paper, we explore how the mundane but necessary task of field support is organized in the case of Apache web server software, and why some project participants are motivated to provide this service gratis to others. We find that the Apache field support system functions effectively. We also find that, when we partition the help system into its component tasks, 98% of the effort expended by information providers in fact returns direct learning benefits to those providers. This finding considerably reduces the puzzle of why information providers are willing to perform this task "for free." Implications are discussed.

Keywords: Open source software; User support; User innovation; Intrinsic motivation; Virtual community

1. Overview and problem statement

Some very successful "open source" software products have been and are being developed, distributed, and supported in the field on a voluntary basis by and for users themselves—no supplier required (von Hippel, 2002). The motives that induce users to contribute to an open source project "for free" and the mechanisms by which the various tasks can be effectively carried out are currently a subject of study for both practitioners and academics. To this point, explorations of the mechanics of and the incentives to participate in open source software projects has focused on the core tasks of developing and debugging and improving the open source software itself. Major motives used to explain why users would voluntarily work on these basic tasks include: (1) a user's direct need for the software and software improvements worked upon; (2) enjoyment of the work itself; and (3) the enhanced reputation that may flow from making high-quality contributions to an open source project. But a complete open source software development and diffusion system contains mundane but essential tasks as well—and the three motivations just described seem to apply relatively poorly to these. We, therefore, devote this empirical exploration to understanding why and how a task at the mundane but necessary end of the scale gets done.

The "mundane but necessary task" we have elected to examine is the delivery of high-quality "field support" to users of open source software. Field support involves provision of assistance to users having difficulties with a product—in this case, an open source software product—because of defects in the product itself or because of the state of the user's own

* Corresponding author. Tel.: +1-617-851-1224; fax: +1-617-344-0403.
E-mail addresses: karim.lakhani@sloan.mit.edu (K.R. Lakhani), evhippel@mit.edu (E. von Hippel).

0048-7333/02/$ – see front matter
PII: S0048-7333(02)00095-1

understanding. Commercial software vendors charge users for field support either directly or indirectly. Open source software does not generally involve a charge for field support. Instead, some product users voluntarily provide answers to the questions of other users—for free.

A number of possible explanations have been put forward as to how and why such a system might work, with the primary puzzle being why information providers expend the effort needed to help others who ask questions. Proposed motives include altruism; incentives to support one's community; reputation-enhancement benefits received by information providers; and expectations of benefits from reciprocal helping behavior by others ("I help today because I have been helped in the past and/or I expect to be helped in the future"). Our decomposition and examination of the Apache web software server help system reduces this puzzle by determining that this relatively mundane but necessary service is provided by volunteer effort at much lower cost than appears on the surface.

The Apache field support system involves information seekers posting their questions on a public website. Potential information providers log onto this website, read the questions and post answers if and as they choose to do so. Total annual time spent by information providers in our sample at the Apache help forum averages over 100 hours. In our analysis, we partition the overall task of information-providing into three subtasks: (1) the posting of a question by information seekers; (2) the reading of posted questions by potential information providers; and (3) the posting of answers. The latter two tasks are undertaken by information providers. We find that 98%, on average, of the time spent at the help website by an information provider is devoted to reading posted questions, and only 2% to providing answers. Information providers report that their motive for reading questions is primarily to learn about problems that other Apache users are experiencing. This learning helps them to manage and update their own Apache websites and software code. In other words, the major cost in providing help, matching of a posted question with a willing and able information provider, is carried out by providers because they directly receive a reward for this activity.

The cost of actually answering questions, task (3), is generally very low, because providers only transfer information they already know to questioners, and report that they expend only 1–5 min on that task per answer provided. The motives information providers report for undertaking this subtask vary. Thus, some answer to promote open source software/free software movement. Others report that they are motivated by an enhanced likelihood of receiving help ("If I answer question on CIWS-U, others are more likely to help me when I post a question in the future") or by a sense of obligation from having received help from others in the past.

In Section 2 of this paper, we describe the context of our empirical research. Next, we review extant literature (Section 3) and describe our research methods (Section 4). Then we report our findings under three headings: participation in the Apache help forum (Section 5); effectiveness of the Apache help forum (Section 6); cost and benefits to help forum participants (Section 7). Finally, in Section 8 we discuss the implications of these findings for open source help line design in particular, and user-based innovation systems in general.

2. Apache, an "open source" software program

Apache is web server software used on "web server" computers connected to the Internet. A web server's function is to "service" requests from Internet browsers for particular documents or content. A typical server waits for client requests, locates the requested resource, applies the requested method to the resource, and sends the response back to the client. Web server software began by offering relatively simple functionality. Over time, however, Apache and other web server software programs have evolved into the complicated "front end" for many of the technically demanding applications that now run on the Internet. For example, web server software is now used to handle security and authentication of users, provide e-commerce shopping carts and gateways to databases.

Apache, like most early web server software programs, was developed by a user—Rob McCool, who developed it for and while working at the National Center for Supercomputing Applications (NCSA) at the University of Illinois. (It was developed in conjunction with Mosaic, the first web browser and predecessor to Netscape, which was also developed at the

University of Illinois.) When McCool left NCSA in the middle of 1994, a small group of web masters who had adopted NCSA server software for their own websites decided to take on the task of continued development for themselves. A core group of eight individuals began the work by gathering all documentation and bug fixes that had been made for NCSA server software up to that point. They put this material together in the form of a consolidated patch. Over time, the name of this *patchy* web server software evolved into Apache. After extensive feedback and modification by users, Apache 1.0 was released on 1 December 1995. In the space of 4 years and in the face of strong competition from commercial competitors like Microsoft and Netscape, the Apache web server has become the most popular web server software on the Internet, used by more than 60% of the 8 million World Wide Web sites extant in early 2000 (Prettejohn, 2001). It has also received many industry awards for excellence.

Apache is open source software: anyone interested can download and have free access to program source code.[1] Given access to source code, technically skilled users of a program can easily make changes and improvements to it. In the case of Apache, this freedom has been exercised by many users and also by programmers working for companies such as IBM and Covalent, that 'package' and sell Apache software for particular applications. Although additions and improvements to Apache code can be made by anyone, additions to the "approved" version of Apache that can be downloaded from the official Apache website must be passed upon by the Apache Development Group, a committee of volunteers (currently 22 in number) who guide the further development and extension of Apache software (Fielding, 1999).

2.1. The Apache field support system

Apache is a relatively complex software program. One of the functions that somehow must be provided for users of such a complex product is "field support"—provision of assistance to users having difficulties with the program because of defects in the program itself or because of the state of their own understanding. Although such a system is needed, the Apache Development Group has made it very clear that they do not want to provide it:

> There is no official support for Apache. None of the developers want to be swamped by a flood of trivial questions that can be resolved elsewhere. Bug reports and suggestions should be sent via the bug report page. Other questions should be directed to the comp.infosystems.www.servers.unix or comp.infosystems.www.servers.ms-windows newsgroup (as appropriate for the platform you use), where some of the Apache team lurk, in the company of many other HTTPd gurus who should be able to help. (Apache Group, 1999)

Despite or because of this lack of "official support," a very effective online Apache field support system has evolved, operated by and for users themselves. The system takes the form of publicly accessible "newsgroup" discussion forums carried on a segment of the Internet called the Usenet. An Apache user with a question "posts" it on the appropriate Usenet discussion forum. Any interested user can read both the questions and answers that have been posted, and can provide answers or add to the discussion if he or she wishes to do so. Both questions and answers are typically signed and identified by the e-mail address of the person posting.

[1] Open source software has its roots in the "free software" movement started by Richard Stallman in the early 1980s. Stallman founded the Free Software Foundation (FSF) as a means to counter the trend towards proprietary development of software packages, and the release of software without the underlying source code. purpose of the foundation was to encourage development of software that would come with source code and be available to users for their own modification. A key feature of FSF based development is a licensing scheme called 'Copyleft'. Under Copyleft, the author of the program has the traditional and legal entitlements of copyright protection along with a license for users to redistribute and change software. The Copyleft license provides unique distribution terms that gives all users the rights to use, modify and redistribute the programs code or any program derived from it but only if the distribution terms are unchanged. Thus, the code and the freedoms become legally inseparable. The Copyleft concept prevents private hoarding of free software if it was just released under a public domain release. All users are compelled to leave copies behind for others to benefit. The philosophy of the FSF movement has been recently extended by a number of individuals who are promoting the 'open source' concept. These individuals are less concerned about the freeness of "free software" and are instead interested in encouraging software companies to release source code for their products. These individuals believe that companies that release source code, under any type licensing, are inherently preferential to closed and proprietary firms (Raymond, 1999).

A question posted on the Usenet initiates a new forum "thread" consisting of a question and associated answer(s). A typical example of such a thread (in this case with one answer only) is as follows.

Subject: Apache 1.3.1 and FrontPage 98 extensions. A small problem . . .

Information seeker:

Hi,

I've compiled and installed Apache 1.3.1 with mod_frontpage.c. That section seems to be working. I have installed the FrontPage 98 extensions, and that seems to almost be working, but I can't find any relevant information anywhere about how to solve this problem. I can look at a home page for a user, but I can't publish to it. Whenever FrontPage tries to connect to the server, this message appears in the error logs:

Thu Oct 8 10: 13:31 1998 [error] (104) Connection reset by peer: Incorrect permissions on webroot "/usr/local/httpd/htdocs/_vti_pvt" and webroot's _vti_pvt directory in FrontPageAlias().

Thu Oct 8 10: 13:31 1998 [error] File does not exist:/usr/local/httpd/htdocs/vti_bin/shtml.exe/_vti_rpc

I haven't a clue how to fix it. Any help will be very appreciated, and a reply by e-mail will be noticed more quickly (I'm terrible at remembering to check the newsgroups)

Thanks!

Information Provider 1:

Hi there,

There are two possible causes of your problem:

1. Make sure owner and group are the same and that the directories have the same set of permissions. /home/user/public_html user group/ home/user/public_html/_vit_bin www group1 should be: /home/user/public_html user group/ public_html/_vit_bin user group
2. Apache-fp utilizes fpexe and mod_frontpage to provide a higher level of security. Part of the mod_frontpage code sets LOWEST_VALID_UID and LOWEST_VALID_GID. Users with UIDs and GIDs less than these values will not be able to run the server extensions. These values are configurable. For more information please check the SERK documentation and the Apache-fp page.

Greetings.

Multiple sources of technical help for Apache users exist in addition to the Usenet help forum, ranging from books to online journals to an online collection of answers to frequently asked questions (FAQs). In order to reduce the volume of questions posted on the Usenet help forum, the Apache Development Group urges users who encounter problems with Apache software to perform two tasks before posting a question: (1) read the Apache FAQs and known bugs databases; and (2) search the Apache Usenet archives for related questions and answers that might solve the user's problem without need for a new Usenet posting. (Although there is no official Apache Usenet archive, all questions and answers have been and are being automatically indexed and preserved in Usenet archives by companies like Google.com.)

3. Literature review: motivations to contribute to open source

Lerner and Tirole (2002) phrase the central motivation question nicely: "Why should thousands of top-notch programmers contribute freely to the provision of a public good?" Raymond (1999), a very experienced participant in open source projects, argued that project participants have at least three basic motives for writing or contributing to the writing of open source software. First, they may directly benefit from the software code they develop, because they intend to use it themselves. Second, they may enjoy the work of programming itself. Third, they may gain an enhanced reputation in the eyes of peers from making high-quality contributions to an open source project.

Niedner et al. (2000) and Lakhani and Wolf (2002) conducted surveys that asked contributors to open source projects about their motivations for doing so. Their findings largely support Raymond's conjectures. Both find the contributors' own need for the software developed as the highest-ranking incentive. Somewhat lower but still strong incentives include improvements to programming skills and enjoyment of the programming work itself. Enhancements to reputation as an incentive was ranked significantly

lower in both surveys—possibly due to self-reporting bias on the part of respondents.[2]

Each of the major motivations listed as very important by contributors to open source projects has some support in the general literature. Thus, it has been shown that innovation users are frequent innovators in a number of fields, and that this course of action can "pay" (von Hippel, 1988). With respect to enjoyment of the work, the characteristics of tasks that individuals often carry out because they are intrinsically rewarding, such as rock climbing, have been explored by Csikszentmihalyi (1975, 1990, 1996). Tasks carried out by participants in open source software projects—writing or debugging software, for example, do fit a number of the characteristics identified by Csikszentmihalyi as associated with intrinsically rewarding tasks—a level of challenge somewhere between boredom and fear, for example. Finally, the fact that "reputation matters" and that seeking to maintain or enhance it can affect behavior has been explored by many.

Kollock (1999) discusses four possible motivations to contribute public goods online. Given that his focus is incentives to put online something that has already been created, his list does not include any direct benefit from developing the thing itself—either the use value or the joy of creating the work product. His list of motives to contribute does include the beneficial effect of enhancements to one's reputation. A second potential motivator he sees is expectations of reciprocity. Both specific and generalized reciprocity can reward providing something of value to another. When information providers do not know each other, as is often the case for participants in open source software projects, the kind of reciprocity that is relevant is called "generalized" exchange (Ekeh, 1974).[3] The third motivator posited by Kollock is that the act of contributing can have a positive effect on contributors' sense of "efficacy"—a sense that they have some effect on the environment (Bandura, 1995). Fourth and finally, he notes that contributors may be motivated by their attachment or commitment to a particular open source project or group. In other words, the good of the group enters into the utility equation of the individual contributor (ibid., pp. 228–289).

Kollock also points out that the kinds and quantities of contributions made online will be sensitive to the costs and benefits involved—and he notes that online costs for distributing a piece of information can be near zero. "While it may be the case that many people spend time and effort producing goods they intend to contribute to the group, another path to the production of public goods is as a simple side-effect of private behavior. People may need to write a particular computer program for their own use with no thought to anything other than solving their particular problem at hand. Having written the program, the costs of now sharing and distributing it with others may be near zero: they can simply post it in an appropriate discussion group or other online community." (Ibid., p. 229). More generally, Thorn and Connolly (1987) argue on the basis of theories of the economics of public goods that the rates and effectiveness of discretionary information sharing amongst employees in an organization will tend to

[2] Niedner et al. (2000) distributed their questionnaire to members of the Linux community. Among other matters, they asked the developers (code contributors) in their sample of respondents to rank the gains and losses associated with their participation on a five-point scale (1: very unimportant and 5: very important). "Facilitating my daily work due to better software" was ranked the highest gain at 4.7; "improving my programming skills" and "having fun programming" were ranked at (4.6); "personal exchange with other software developers" (4.2); "career advantages due to experience gained in Linux projects" (3.7); "gaining a reputation as an experienced programmer inside the Linux community" (3.5). The two losses listed were not regarded as very important. They were: "time loss due to my involvement in Linux projects" (2.6); and "lack of payment for my work in Linux projects" (2.2). Lakhani and Wolf (2002) conducted a questionnaire study of contributors to a range of open source projects listed on Sourceforge.net. Respondents were asked to list the "top three motivations for (your) contributing to an open source project. Fifty-nine percent rated work or non-work need for the software as one of their three top motivators. A progressively smaller proportion of respondents listed the following motives as among their top three: intellectually stimulating (44%); improves skills (41%); code should be open (33%); obligation felt from own use of open source code (29%); work with team (20%); enhance professional status (18%); increase reputation in the open source community (11%) and beat proprietary software (11%).

[3] In "generalized" exchanges, help given to a person is reciprocated by someone else in the group and not by the particular recipient of the original help. Generalized exchange is used to explain why, for example, stranded motorists get helped by strangers: the person helping is expecting that when they are stranded, someone will help them in turn. Kollock notes that "... indeed some observers (Wellman and Gulia, 1999; Rheingold, 1993) have reported that individuals who regularly offer advice and information seem to receive help more quickly when they ask for something." (Ibid., p. 227).

decrease as: (1) participation costs increase; (2) the size of the overall group increases; (3) lower value of information to participants; and (4) greater asymmetries in information values and benefits across participants.

3.1. Motivations to contribute to open source help lines

User participation in the major tasks of free and open source software projects—software writing and debugging—may in fact be motivated by personal benefit from the work product, by fun and learning associated with performing the work and by reputational considerations. However, "necessary but mundane" tasks carried out by volunteer effort in such projects do not appear to fit this set of motivations very well—at least on the face of it. Lerner and Tirole (2002) consider the net benefit that participants may obtain as consisting of immediate payoff (current benefit minus current cost) plus a delayed payoff. Immediate payoffs consist of the programmer's own use of the program improvement developed. Immediate cost consists of the opportunity cost of the time invested by the programmer, with the actual cost of this time depending upon how enjoyable the programmer finds the task. The delayed payoff consists of a career concern incentive (future job offers, etc.) and an ego gratification incentive stemming from a desire for peer recognition. Lerner and Tirole argue that both of these delayed payoff elements can usefully be seen as instances of what the economic literature calls signaling incentives (Holmstrom, 1999). As they observe, "... tasks aiming at helping the much-less-sophisticated end user—e.g. ... technical support—usually provide lower signaling incentives." (Ibid., p. 19).

If providing answers to users on a help line does not obviously involve a work product of immediate value to the information provider, and if signaling incentives are low for this task, the question we started remains unresolved: why do some users willingly carry out necessary but mundane tasks such as providing free help to others who pose questions on open source help lines? Constant et al. (1996) have carried out the only empirical study we are aware of that has some empirical data the motivations reported by participants in a computer "help line" system. The particular system they explored was the Tandem Computers Inc. internal corporate help line implemented upon that firm's internal computer network. Their sample was 55 information seekers and 295 information providers (most questions received several replies). Overall, they found that the system was effective: information seekers did get technical advice that they found useful, with 49% saying that replies received had solved their problem.

Table 1
Information providers reasons for answering questions on a corporate online help line[a]

Reasons for participating	Points assigned (mean)
Personal benefits	
I enjoy helping others	16
I enjoy solving problems	9.5
I enjoy earning respect	4.8
The company rewards information sharing	0.9
Total	31.2
Organizational motivation	
Being a good company citizen	17.8
The problem is important to the company	14.0
It is part of my job to answer questions like this one	12.6
I expect others to help me, so it is only fair to help them	11.8
Total	56.2

[a] Source of data: Constant et al. (1996), Table 5, p. 129.

To measure information providers motivations, the researchers asked each information provider in their sample to allocated 100 points among eight reasons they might have had for replying to the information seeker, with the results shown in Table 1. Of course, participating in an open source software help line is not the same as participating in a corporate one. However, on the face of it, these findings suggest that "being a good company (open source project?) citizen" and executing tasks "important to the company (project?)" may be important motives for participation. Enjoyment of the task of answering a question, "part of my job" and reputational gains ("I enjoy earning respect") also appear, but less strongly.

4. Research methods

The empirical exploration of the Apache help system we report upon here was preceded by a pilot study

of Apache help system behavior (Lakhani, 1999) and by several interviews held with several individuals who had very good first-hand knowledge of the Apache field support system.[4] The empirical data we collected for study was related to postings to the Apache Usenet help forum, CIWS-U (comp.infosystems.www.servers.unix). CIWS-U is one of two Usenet newsgroups that address questions related to Apache web server software. It was chosen for study because the questions posted to it are predominantly Apache-related. (Only a few postings on this site deal with questions about other varieties of Unix-based server software, and we excluded these from our analyses.)

Two basic types of empirical data were collected regarding postings to this Apache Usenet help site.

1. For data regarding long-term participation in CIWS-U—who participated, long-term trends, etc.—we examined Usenet posting patterns from 1996 to 1999. This 4-year period spans essentially the entire history of online Apache help (recall that Apache 1.0 was released only in December 1995). The Usenet log data was obtained from a World Wide Web service called Deja.com (since acquired by Google). This service archives all of the discussion groups on the World Wide Usenet and makes available advanced search and parsing capability through their website (http://groups.google.com).
2. We collected questionnaire data from people who posted either questions or answers to CIWS-U during the 4.5 months from 1 October 1999 to 15 February 2000 (see Appendix A for list of questions asked). During this time period, we monitored activity on CIWS-U near-continuously via computer. Within 3 days of when a question or an answer was posted on CIWS-U, our computer automatically detected whether the individual was posting a question (e.g. was starting a new "thread") or was providing information related to a previously posted question (e.g. was referring to an existing thread in his or her posting). It then sent the proper version of our questionnaire (one appropriate to information seeking or one appropriate to information providing) to the e-mail address of that individual. The e-mail contained a brief introduction to the study, a link to the individual's actual posting on CIWS-U and a link to a password protected website that contained the survey. This "automatic" data collection method had the advantage of allowing us to obtain information from posters on a near real-time basis—while recollections regarding what they did and why they did it was still fresh. Upon completion of each questionnaire, the individual answers were archived to a protected database as well as e-mailed to us.

[4] These individuals were: two current members and one emeritus member of the core Apache Group; one significant contributor to Apache, and two individuals who had participated frequently in the Usenet portion of the Apache field support system.

Table 2
Sample of individuals posting questions or answers on Apache Usenet help site from 1 October 1999 to 15 February 2000

	Total participants	Information seekers	Information providers
Sample queried	1709	1288	421
Usable responses	336	214	122
Response rate (%)	19.6	16.6	28.9

While designing our data collection methods, we sought advice from some Apache Group members regarding presentation and procedure. As finally implemented, each questionnaire was accompanied by a brief letter explaining who we were and what we were trying to do—that is, we were trying to learn about the Apache help system. To minimize intrusion on potential respondents, we did not follow up our initial request with any repeated requests to respond, and we only sent a questionnaire out to any individual once—in response to the first time that individual either posted a question or an answer during our period of data collection. We also provided an e-mail address for anyone who wanted to contact us to complain or comment. (In the end, we received only six comments, half favorable and half not.)

The sample size and response rates for this sample are as shown below. The data collection period for this sample included Christmas and New Year's vacations, and response rates during these times was about half of the average level shown Table 2.

An examination of posting histories on CIWS-U during the period 1996–1999 showed that some of our information seekers had sought information many more times than the mean for all seekers and that, similarly, some of our providers had provided many

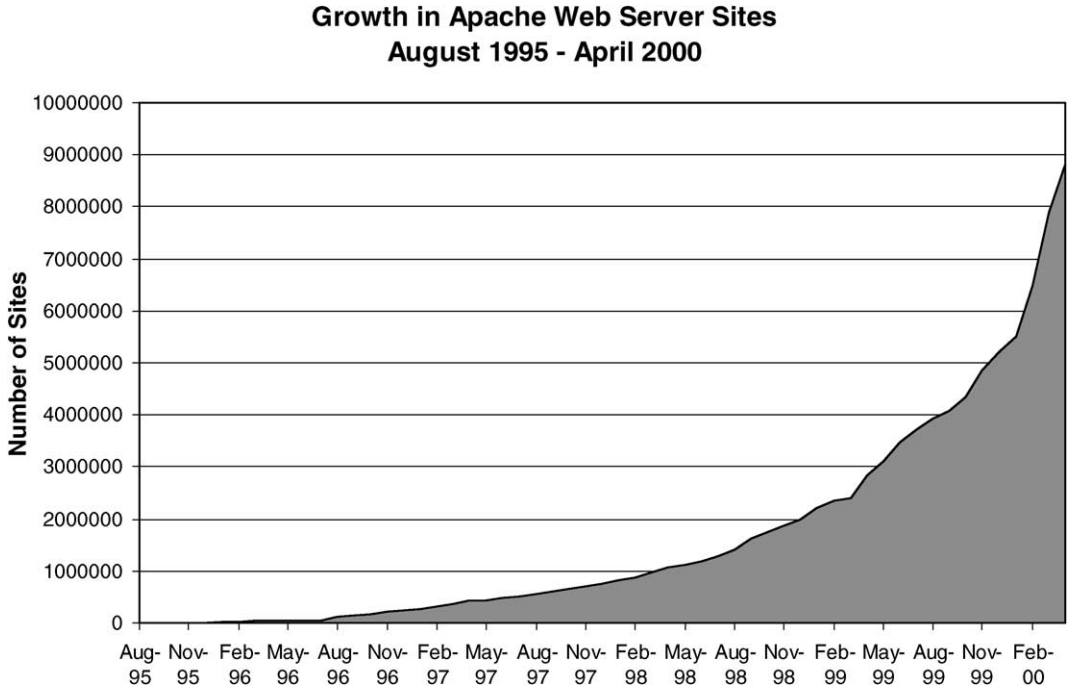

Fig. 1. Number of websites using Apache 1995–2000 (source: Netcraft web server survey, available at http://www.netcraft.com/survey).

more times than the mean for all providers. Preliminary data analyses showed it would be useful to contrast these individuals with more average seekers and providers on a number of variables. Accordingly, we divided our sample of information seekers into two subsamples. "Frequent seekers" were all information seekers who posted four or more questions during the period 1996–1999 (about the top 10% of our seeker respondents) and who had a ratio of information seeking to information providing posts greater than one. All other seeker respondents were placed into the subsample of "other seeker." Similarly, "frequent providers" were all information providers who posted 10 or more questions during the period 1996–1999 (about the top 10% of our provider respondents) and who had a ratio of provide to seek posts greater than one. All other provider respondents were placed into the subsample of "other providers."[5]

[5] The reason for the ratio test was that respondents were sorted into seeker or provider categories according to their role in the first (and sometimes only) posting they made in our sampling window of 4.5 months. If analysis of CIWS-U logs showed that they more typically were posting messages in the opposite role (e.g. seeker instead of provider) we did not want to include their data in our assessment of "characteristics of seekers versus providers." We could have gone the next step and shifted them into the category which was their typical role, but elected not to do this. Trial data analyses showed that such category shifting would affect only a few individuals and would not materially affect our findings. On the negative side, category shifting would make the analysis more difficult to follow.

5. Findings: nature of participation in the Apache Usenet help forum

Apache version 1.0 was released in December 1995. As Fig. 1 shows, the number of websites using Apache has increased dramatically since then, to over 60% of the web server software "market" and over 8 million sites active at the start of 2000 (Prettejohn, 2001).

The number of new "threads" initiated each month on the Apache help forum (a thread consists of a question) has also been growing, but not nearly so rapidly (Fig. 2). Participation in the Apache Usenet help forum is small relative to the number of sites (8 million in early 2000—run by perhaps 800,000

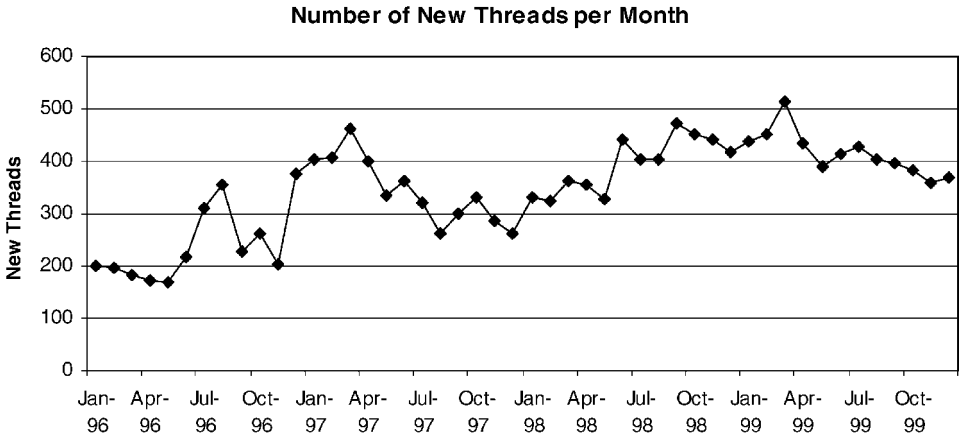

Fig. 2. Number of new questions asked per month from 1996 to 1999.

web masters) using Apache. However, interviewees inform us that this is not particularly surprising. Although the Usenet help site is the oldest one for Apache, other online sites have been established that perform a similar function for Apache users in a similar way, and that are used by many Apache help seekers.

During the 4-year period (1996–1999), there were 11,510 distinct participants in the Apache Usenet help site. Of these, 4902 only posted answers on CIWS-U (information providers), 8981 only posted questions, and 2372 did both. Information providing was relatively concentrated (Fig. 3). Approximately 50% of the answers on the system were provided by the 100 most prolific providers (2% of all providers; Gini coefficient 0.68). In contrast, 50% of the questions were provided by the 2152 most prolific posters of questions (24% of all information seekers).

The 100 most active information seekers posted an average of 10.43 questions and the 100 most active information providers posted an average of 83.63 answers during the 4-year period (1996–1999). Frequent participants also turned out to be long-term participants. We found that mean elapsed time between first and last posts during the 1996–1999 period was 674 days for frequent information providers; 168 days for other providers; 661 days for frequent information seekers and 107 days for other information seekers. (These periods of participation should be taken as "equal to or greater than" statements about length of

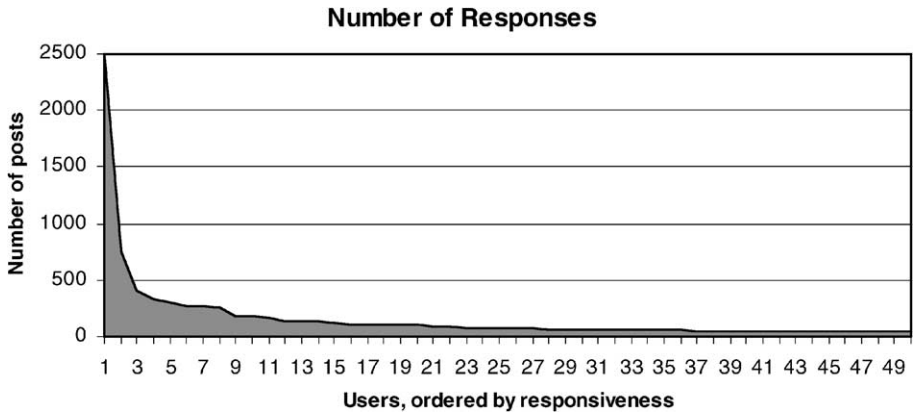

Fig. 3. Number of answers provided by the top 50 Apache Usenet help participants from January 1996 to September 1999.

Table 3
Attributes of respondent information seekers and information providers

Attribute	Frequent providers	Other providers	Other seekers	Frequent seekers
Mean Apache Usenet reading time/session (min)	12.48 (13.07)	18.09 (14.09)	18.52 (23.63)	17.69 (17.60)
Mean time using web servers (months)	47.71 (21.80)	43.94 (22.07)	29.70 (21.63)	50.31*** (20.97)
Mean Apache experience in months	33.86 (15.09)	31.99 (20.14)	21.13 (18.22)	41.54*** (20.09)
Mean percentage of work time dedicated to web server operations	51.19 (38.86)	36.85 (35.37)	29.38 (32.28)	24.39 (27.22)
Mean scale of website site in log of millions of hits per day	4.89** (1.43)	4.08 (1.37)	3.20 (1.58)	4.14* (1.81)
Percentage that have modified Apache source code	81	46	22	31
Percentage whose website is for professional purposes	48	60	47	69
Mean total posts as information seeker over 4 years (1996–1999)	3.81* (5.01)	1.44 (5.01)	1.71 (1.85)	4.77*** (1.09)
Mean total posts as information provider over 4 years (1996–1999)	169.29 (537.94)	2.53 (2.99)	1.98 (7.21)	2.08 (1.12)

Standard deviation is given in parenthesis. Statistical comparisons refer to differences between frequent providers vs. other providers, and to frequent seekers vs. other seekers.
* $P < 0.10$.
** $P < 0.05$.
*** $P < 0.01$.

participation, since it is likely that many will continue to post during year 2000 and beyond.)

Information seekers differed from information providers on a number of attributes (Table 3). In general, frequent information providers and frequent seekers as well appear to be more expert than "other" information seekers or providers, having on average have more months of experience with Apache, and with web servers in general. Frequent providers are much more likely to modify the Apache source code (81% have done this) than are other posters to Apache Usenet help.

6. Findings: effectiveness of the Apache help process

Web server users rank Apache technical support overall as somewhat better than that of its major commercial rivals in the server software field. Thus, participants in the 1999 ServerWatch, an Internet-based trade publication, poll[6] ranked Apache 4.5 out of 5 with respect to technical support. Commercial offerings from Netscape and Microsoft received a ranking of 4 out of 5. This general endorsement may or may not apply to Apache online help specifically since Apache technical support has a number of elements and, as our information seekers attest (Table 4), many are used.

Questions posted by information seekers varied in nature (Table 5). Only 9.6% of all information seekers said that the problem they posted online was extremely critical and that they needed an answer right away.

Data collected on response times from 1996 to 1999 Apache Usenet logs and also from our "real-time" sample showed that initial answers to publicly posted questions generally came quite quickly—at least 50% were answered on the day of or on the day after posting (Table 6).

As can be seen from Table 6, 39% of information seekers received no public reply (that is, a reply posted for all to read on Usenet) to their Usenet posting (true for both samples in Table 6). However, 40% of the respondents to our 4.5-month real-time sample who received no public reply to their query reported receiving one or more replies that were sent privately via e-mail instead. If this ratio holds for the historical data as well, then only about one-fourth of the questions posted on Usenet do not receive an answer. (Lakhani (1999) compared the content of a sample of messages that did receive public replies with a sample that did not, and found no obvious differences with respect to clarity, completeness or technical difficulty.)

[6] Available at http://ServerWatch.internet.com.

Table 4
Additional Apache help resources used by individuals posting questions on CIWS-U

Apache resource used	Frequent seekers		Other seekers	
	Using (%)	Mean time (min)	Using (%)	Mean time (min)
Apache FAQs	69	13.3	79	39.3
Usenet archives	77	23.5	78	30.2
Other online resources[a]	69	18.8	40	38.4
Books on Apache	54	65.8	39	140
Known bug data base	69	2.5	32	13.6

For frequent seekers, $N = 13$; for other seekers, $N = 201$.
[a] For example, online "journals" such as Apacheweek.com.

Table 5
Nature of questions posted on Usenet by information seekers

Type of problem asked about	Frequent seekers		Other seekers	
	Number	Percentage	Number	Percentage
Complete down	–	–	5	2.6
Functional—missing important features	4	33.3	48	24.7
Functional—missing optional features	5	41.6	109	56.2
Installation problems	2	16.6	26	13.4
Upgrade problems	1	8.3	6	3.1
Total	12	100	194	100

Table 6
Response to questions posted on Usenet

Sample	Got public reply same day (%)	Got public reply next day (%)	Got public reply after 2 days (%)	Got private e-mail reply only (%)	No reply received	Number
1996–1999 Usenet log data	32	17	12	NA	39%, no public reply	12,964
4.5-Month real-time sample	34	18	9	16	23%, no public or private reply	1,288

7. Findings: costs and benefits of participating on Apache help forum

To successfully complete an information transaction on the Apache Usenet help forum, three tasks must be completed: (1) a question must be posted; (2) the information sought must be matched to an appropriate and willing provider of information; and (3) an answer must be provided. Obviously, the burden of question-asking must be placed upon the information seeker, and the burden of information provision (both the time associated with providing it and any losses associated with sharing proprietary information) on the information provider. However, the burden of seeker and provider match-up varies according to the design of the information system. For example, in the case of an encyclopedia or a FAQ database (a list of answers to FAQs), the burden of match-up is placed upon the information seeker. However, in the case of the Apache help Usenet forum, the burden of matching up an information seeker and an information provider is placed on the information provider.

7.1. Costs and benefits of question posting

Members of the Apache community are very familiar with Usenet procedures. As a consequence, cost to information seekers posting a question to the Apache

Table 7
Information seekers' evaluation of the answers that they received to the question they posted on Apache Usenet

What was the value to you of the answers you received?	Frequent seekers (%)	Other seekers (%)
Solved my problem completely	23 ($n = 3$)	17 ($n = 34$)
Gave me information that helped solve my problem	69 ($n = 9$)	44 ($n = 87$)
Did not solve my problem	8 ($n = 1$)	39 ($n = 77$)

Table 8
Respondents' Usenet reading pattern

	Frequent provider ($n = 21$)	Other providers ($n = 68$)	Other seekers ($n = 195$)	Frequent seekers ($n = 13$)
How frequently do you read Apache Usenet?				
Daily (%)	76	32	11	23
Weekly (%)	24	43	22	30
Monthly (%)	–	7	8	–
Only when problem (%)	–	18	59	47
Time expended?				
Mean annual reading volume (min)[a]	4774	2774	1838	1816

[a] Annual reading volume was calculated by multiplying number of reading sessions reported times average length of session reported.

Usenet help site consists only of their time expenditure to prepare and post that question. Seekers report preparation and posting time to be a mean of 11.5 min (S.D. = 25.9, $n = 212$).

Benefits to seekers consist of the problem solving time saved due to answers received to their posted question. As can be seen from Table 7, a majority of both frequent seekers and other seekers who received replied to their questions judged the information contained in those replies to be useful. (Respondents who received both public and private replies generally judged both to be of equal value: 24% judged the private replies to be of higher value, 18% judged the public replies to be of higher value and 58% viewed them to be of equal value ($n = 106$).)

Seekers who received answers to their questions estimate the problem solving time they saved due to answers received to their questions at a mean of 115 min (S.D. = 225, $n = 187$).[7] Thus, the mean net time benefit information seekers receive from posting a question on CIWS-U is 103.5 min. Or, to put it another way, the benefit to cost ratio experienced by information seekers who post a question is about 9—quite a good return on investment!

7.2. Costs and benefits of question and answer matching

Potential information providers identify questions that they can and are willing to answer by simply reading or scanning the questions posted on the Apache help forum. In order to understand the extent of the match-up burden placed upon information providers, we asked our respondents about the time they spent reading CIWS-U. Table 8 indicates that the annual time spent, especially by information providers, is typically quite substantial.

If information providers incurred the substantial time expenditures devoted to reading CIWS-U only to identify questions they were able and willing to answer, they would indeed be spending heavily to help information seekers. But as Table 9 shows, information providers (and seekers) report that the most important reason they read CIWS-U is to learn: they gain valuable information from reading about problems other users are encountering, and how these might be solved.

[7] We excluded five "outlier" respondents that reported more than 1440 min (24 h) of time savings from the analysis. Including them in the analysis boosted the mean to 381 min and S.D. to 1923.

Table 9
Respondents' reasons for reading Usenet (seven-point scale: 1—strongly disagree, 7—strongly agree)

Reasons for reading	Frequent providers ($n = 21$)	Other providers ($n = 68$)	Other seekers ($n = 188$–191)	Frequent seekers ($n = 13$)
To learn	5.90 (1.58)	5.75 (1.29)	5.29 (1.81)	6.38*** (0.87)
To answer	4.95*** (1.02)	4.00 (1.44)	3.77 (2.04)	4.17 (1.83)
For fun	4. 29 (1.19)	3.97 (1.74)	2.91 (1.62)	3.46 (1.76)
For break	4.81** (1.33)	3.99 (1.33)	2.66 (1.88)	2.69 (1.65)

Standard deviation is given in parenthesis. Statistical comparisons refer to differences between frequent providers vs. other providers, and to frequent seekers vs. other seekers.

** $P < 0.05$.

*** $P < 0.01$.

7.3. Costs and benefits of information providing

In the Apache system, as we noted earlier, the cost of question and answer match-up falls upon the information provider. However, providers accomplish the match-up task by reading or scanning questions posted on Usenet. And, as responses in Table 9 showed, providers do this primarily in order to learn, rather than to answer questions. Given this finding, we reason that the task of question and answer match-up in the Apache Usenet system is effectively achieved as a costless side-effect of an activity undertaken for another reason by potential information providers. We, therefore, think it is reasonable to leave aside the cost of question and answer match-up in assessing the net benefit of posting to CIWS-U for information providers.

Leaving match-up costs aside, costs incurred by an information provider who answers a question on Usenet involves two elements: (1) value of proprietary information that may be lost when that information is publicly posted on the Apache Usenet forum; and (2) the costs and benefits associated with generating and posting an answer to a posted question. We assess each of these elements in turn.

Information held by information providers loses any proprietary value it might have had (unless it is protected by patent—a very unlikely circumstance) if it is publicly posted to the Apache help forum. However, *if* potential providers think that others know the same information and *if* they think those others will provide it if they do not, providers should assess the loss of intellectual property value associated with their choosing to answer a question at zero. (Indeed, under these conditions, a provider's best strategy may be to strive to be the first to reveal the information sought in order to reap any associated reputational advantages.)

Table 10
How many others do you think know the solution to the question you answered on Usenet?

How many others do you think knew the answer to the question you answered?	Frequent providers (%)	Other providers (%)
Many	38 ($n = 8$)	61 ($n = 41$)
A few with good Apache knowledge	38 ($n = 8$)	18 ($n = 12$)
A few with specific problem experience	24 ($n = 5$)	21 ($n = 14$)
No others	NA[a]	NA[a]

[a] See footnote 8.

On the basis of this reasoning, we asked the information providers in our sample "How many other readers of CIWS-U do you think also knew a solution?" to the question they had answered on the Apache forum. As can be seen below, all providers reported that they did think that some or many other readers also knew a solution and so could potentially furnish an answer Table 10.[8] Information providers potentially concerned about losses of valuable proprietary information incurred by answering a question posted on the Apache help forum have no logical reason to be concerned—if and as they think that others holding the same information would answer if they did not. We did not ask providers whether they in fact

[8] The level of this response is to some unknown degree inflated: we neglected to include an explicit response option of "no others" for this question on our questionnaire, and so the only way that a respondent could even indicate such a view was by not indicating agreement with any of the options presented—which none did.

Table 11
How long did it take you to answer the posted question?

Time spent to answer (min)	Frequent providers (%)	Other providers (%)
≤1	48 ($n = 10$)	19 ($n = 13$)
1–2	29 ($n = 6$)	21 ($n = 14$)
2–5	19 ($n = 4$)	40 ($n = 27$)
5–10	–	16 ($n = 11$)
>10	4 ($n = 1$)	4 ($n = 3$)

held this view. We did, however, ask a related question: "I answered the question because I thought the poster might not get a good answer if I did not." On a scale of 1–7, with 1 being strongly disagree and 4 being neutral, frequent information providers expressed a moderate level of agreement (a mean of 4.52) with this statement (Table 14, reason number 12).[9] This suggests that at least these information providers are not viewing answer-provision in terms of potential loss of value of proprietary information—whether or not they "should".

We next consider the costs and benefits associated with generating and posting an answer to a question posted on the Apache help forum. An important finding here is that the cost of carrying out this task is typically quite low. About half of frequent information providers spent 1 min or less answering a question on Usenet, and 80% of other providers spent 5 min or less at this task Table 11.

As we can see from Table 12, this small time expenditure was possible because providers generally already knew the answer to the posted question.

Providers were asked whether they knew the answer because of their general knowledge of Apache (32%, $n = 38$), or because they had experienced the same problem themselves (68%, $n = 82$). When information providers knew the answer due to their general expertise in Apache, their mean time expenditure was significantly shorter (3.2 min) than when they knew the answer because they had experienced the problem themselves (5.5 min mean time expenditure) ($P = 0.013$). Whatever their state of knowledge at the time information providers saw the posted question, they typically only provided information they already had in hand (Table 13).

[9] The level of agreement with the question (on a scale of 1–7, with 1 being strongly disagree) was: 1 = 10, 2 = 5, 3 = 20, 4 = 40, 5 = 24, 6 = 14, 7 = 9. Total $n = 122$.

Table 12
What was your state of knowledge when you first looked at the question you answered?

State of knowledge	Frequent providers (%)	Other providers (%)
Already knew solution	76 ($n = 16$)	64 ($n = 44$)
Knew where to find the solution	5 ($n = 1$)	2 ($n = 1$)
Some useful information but not solution	19 ($n = 4$)	28 ($n = 19$)
No solution but had ability to solve	–	6 ($n = 4$)

On average, information providers who only provided information they already had expended 4.0 min to provide an answer. Providers who either searched for more information or engaged in problem solving before answering expended 9.33 min to respond. This difference is significant at the 0.05 level ($P = 0.02$).

To this point we have found that the costs incurred by information providers to answer a question on Apache Usenet are typically quite small. Frequent providers typically take 2 min or less to generate and post an answer, and other providers spend 5 min or less to do this.

We next turn to consider the benefits potentially flowing to information providers from investing this small amount of time to answer a question posted on the Apache help forum. As was discussed in our review of the literature (Section 3), several types of benefit may be motivating information providers to respond.

- *I expect reciprocity* (statement nos. 1–3 in Table 14). Both specific and generalized reciprocity can reward providing something of value to another. Since,

Table 13
What did you do to answer the question

Activity undertaken	Frequent providers (%)	Other providers (%)
Provided information I already had	90 ($n = 19$)	82 ($n = 56$)
Searched for additional information	10 ($n = 2$)	15 ($n = 10$)
Engaged in problem solving	–	3 ($n = 2$)

Table 14
Providers' views regarding their motives for providing answers to help seekers on Apache Usenet (seven-point scale: 4—neutral, 7—strongly agree)

I was motivated to answer because	Frequent providers	Other providers	Percentage "strong" agreement (6–7 on scale; %)
(1) I help now so I will be helped in the future	4.52* (1.25)	5.16 (1.38)	15
(2) I have been helped before in CIWS-U—so I reciprocate	4.85 (2.08)	5.14 (1.52)	48
(3) I have been helped on Usenet before—so I reciprocate	4.61 (1.96)	5.16 (1.53)	45
(4) I answer to enhance my career prospects	3.76 (1.55)	3.57 (1.31)	6
(5) I want to enhance my reputation in OSS/Apache community	4.71 (1.35)	4.57 (1.42)	24
(6) I answer because its fun	4.81 (1.44)	4.38 (1.49)	28
(7) I answer to promote OSS	5.14 (1.35)	4.76 (1.47)	33
(8) I answer to take a break	4.65 (1.65)	4.22 (1.49)	20
(9) I answer because it is part of my job	2.23 (1.76)	2.52 (1.75)	5
(10) I have expertise in this area	4.47 (1.32)	3.92 (1.77)	18
(11) I am the authority in this area	2.47 (2.14)	2.01 (1.56)	4
(12) I answered because I thought the poster would not get a good answer if I did not	4.52 (1.57)	4.08 (1.50)	18

Standard deviation is given in parenthesis.
* $P < 0.10$.

as we will see shortly, the information providers did not know information seekers before providing help, the most relevant source of literature is that on "generalized" exchange (Ekeh, 1974). In such exchanges, help given to a person is reciprocated by someone else in the group and not by the particular recipient of the original help. Generalized exchange is used to explain why, for example, stranded motorists get helped by strangers: the person helping is expecting that when they are stranded, someone will help them in turn (Kollock, 1999).

- *I am "helping the cause"* (statement no. 7). Individuals involved in open source software projects often strongly identify themselves as belonging to a community (Raymond, 1999). Constant et al. (1996) demonstrated that people who have a strong attachment to an organization will be more likely to assist others with organization related problems. It has also been argued that people who develop a strong attachment to a virtual group are more likely to participate and provide assistance to others (Wellman and Gulia, 1999).
- *I will gain reputation or enhance career prospects* (statement nos. 4 and 5). The identity of information providers is preserved through their e-mail addresses, user names and the "signatures" to the answers they post. Thus, information providers may gain in reputation by answering frequently or well. Gains in reputation can be rewarding in and of itself, and may also lead to benefits such as enhanced career prospects.[10] A number of researchers have argued that gaining a reputation within a community, including a online community, is an important incentive for active participation (Constant et al., 1996; Lerner and Tirole, 2002; Raymond, 1999; Rheingold, 1993).
- *Answering questions is intrinsically rewarding* (statement nos. 6 and 8). Interviewees with expertise in Apache suggested to us that intrinsic rewards—induced feelings of competence, fun, or being rewarded by "taking a break" were important motivators for answering questions. This view finds support in the research of Csikszentmihalyi, who has explored the characteristics of activities

[10] Some Apache help forum users we interviewed suggested that an "alpha-male" variant of reputation building behaviors might be visible among information providers. Some providers, they said, wanted to be known as "the" expert in a particular aspect of Apache. To build and preserve such a reputation, these providers would strive to quickly answer all questions associated with "their" area. They would also seek to drive out other providers who offered answers in that area by quickly posting comments on the answers provided by those others in a way that, while outwardly cooperative, would also indicate their own technical superiority and prowess in the particular area. In other words, such a person acted like an "alpha-male" by attempting to drive out all other information providers from his chosen field of expertise. We saw no evidence of such behavior in our small sample—in the sense that we saw no clustering of answers by subject area.

individuals engage in because they offer the intrinsically rewarding experience of "flow". Answering questions on the Apache help forum does appear to fit a number of the characteristics of "microflow" activities that have been found to be intrinsically rewarding (Csikszentmihalyi, 1975, 1990, 1996).

- *It is part of my job* (statement no. 9). Several companies are now selling commercial versions of Apache software. Typically this entails offering a packaged distribution of Apache, plus documentation and support. It is possible that such companies might assign people to answer questions posted on the Apache help forum as part of their job responsibilities.

We asked information providers to express their agreement or disagreement regarding each of these possible motivations, with the results shown in Table 14.

In general, we can see that providers were in moderate agreement with most of the motivations listed in Table 14. Top providers differed in expectable ways from other providers, for example they felt that they had more expertise. In addition, we note that the statement that "it is part of my job" was strongly disagreed with by most (63% expressed disagreement and 27% indicated neutrality; only two respondents in the frequent provider category agreed with this statement). This makes it clear that helping is indeed voluntary for most respondents.

Of course, all self-reporting regarding motivations must be viewed with caution: respondents may be inclined to emphasize the "right" socially correct or conventional motivations (Drake et al., 1982). This concern is reinforced for us by an apparent contradiction between stated motives and related evidence with respect to reciprocity. In Table 14, the most agreed-with statements include the three statements having to do with reciprocity "I help because I have been helped and/or expect to be helped" (statements 1–3 in Table 14).[11] Information seekers do show a higher level of agreement than do information providers, but the level of agreement shown by providers is hard to square with rational expectations of specific, tit-for-tat reciprocity behaviors: 96.7% ($n = 116$) of the information providers reported that they did not know the individual they were helping. Also, it is unlikely that generalized reciprocity was at work here. Recall that seekers and providers had different characteristics. Recall also that, of the CIWS-U posters in the period 1996–1999, 57% sought information only, 22% provided information only, and only 21% did both (posting an average of 2.50 questions and 7.95 answers). Possibly respondents are really saying that they feel reciprocity is involved because they have gained by learning from reading the questions and answers posted by others on Usenet, and can reciprocate by answering questions.

[11] The exact text of each of these questions was as follows. (1) "Others have helped me in the past on other Usenet groups and I feel an obligation to reciprocate by answering questions on Apache Usenet." (2) "Others have helped me in the past on CIWS-U and I feel an obligation to reciprocate by answering questions on Apache Usenet." (3) "If I answer a question on CIWS-U others are more likely to help me when I post a question in the future."

8. Discussion

In this research we have explored provision of a "necessary but mundane" task—provision of online technical support—by and for users of Apache open source software. In net, we found that the Apache online Usenet help site works quite well for those who participate. Most questions posted are answered quickly and most answers received are judged to be valuable by information seekers.

In our analysis, we segmented an information transaction on Apache Usenet help into three subtasks: (1) a question must be posed; (2) the information sought must be matched to an appropriate and willing provider of information; (3) an answer must be provided. We noted that the case of the Apache Usenet help system, the burden of matching up an information seeker and an information provider and the actual provision of an answer has been placed on the information provider: each potential information provider finds questions he or she can and will answer by reading or scanning questions that have been posted on Apache Usenet help, and then posts an answer. A comparison of the time spent by information providers on tasks (2) and (3) shows that 98% of the time spent at the Apache online Usenet help site by providers is spent upon task (2)—reading questions and answers posted on the Usenet site. Apache information providers reported gaining a direct benefit from investing in this task—they learn valuable information

relevant to the management and upgrading of their own website.

We found that the actual answering of questions (task (3)) took up only 2% of a information provider's time on site, with providers reporting that they invested only l–5 min per question answered. We found that information providers were able to answer at this low cost because they only posted information they already knew "off the shelf"—they seldom did new problem-solving or searching in order to provide additional information to a help-seeker. (This low time investment by helpers matches findings by Constant et al. (1996). In their study of the Tandem Computers corporate help line, they found that the average time devoted to posting an answer to a question was 9 min (ibid., p. 124).)

What we have found has implications both in the very specific context of provision of technical support for Apache and other open source software, and the more general issue of the analysis and design of complete innovation systems run by volunteer effort.

8.1. Implications for the provision of online help for open source software

Apache has the same general development and support characteristics as other open source projects, such as a distributed development process driven by expert users along with voluntary participation and online user-based technical support. Thus, our findings with respect to provision of Apache online help should be relevant to the broad range of open source software projects that employ voluntary online support for users. However, there are some obvious issues with respect to the robustness of the Apache help system that we have studied given the variation in conditions that could reasonably occur in open source projects.

First, recall again that Apache Usenet help information providers have had 98% of their effort rewarded via the learning they gain from scanning the questions and answers posted by others. It is logical that this benefit could be higher or lower in some open source projects, or that in a given project, this benefit may decrease if and as there is "less to learn." In the specific case of Apache, this may happen if the rate of change in the environment faced by Apache or the rate of change in Apache itself decreases. A comment by Eric Raymond on his experience with help from users of his open source program, fetchmail, is suggestive in this regard.

> Actually ... the list [of fetchmail beta-testers] is beginning to lose members from its high of close to 300 for an interesting reason. Several people have asked me to unsubscribe them because fetchmail is working so well for them that they no longer need to see the list traffic! Perhaps this is part of the normal life-cycle of a mature, bazaar-style project. (Raymond, 1999, pp. 46–47).

Second, it is reasonable to ask whether the mechanism we have seen functioning well in the case of the Apache help website can also function effectively if question loads are much greater. The Apache Usenet site presently relies heavily on around 100 information providers who in aggregate post 50% of the messages, with the very top few frequent information providers answering hundreds of questions each (cf. Fig. 3). Would the number of providers go up in proportion if question volume rose, say 100×?

Third, we found that low cost provision of answers was possible in the case of Apache because some information providers could provide the requested information "off-the-shelf." These conditions may not hold for all problem types and user communities. Thus, in some communities the problems encountered by some users may be unique to them and no off-the-shelf solution may exist. Or, even if a solution does exist in the user community, a problem may not lend itself to a clear-enough description to allow a remotely-located expert to match up problem and solution at a low cost. For example, consider that there are some problems in fields ranging from machine diagnosis to medical diagnosis where experts find they must physically go to the problem site to make first-hand observations before they can understand the problem well enough to offer an appropriate solution.

If the present model of Apache help does get less effective for either the first or second reason mentioned, there is room for modifications that may still allow volunteer information providers to get the very important help task done. Under the current system, the benefit to cost ratio of information seekers is very favorable—currently they save 9× more time than they expend. This suggests that some system changes that partially or fully shift the match-up burden from providers to seekers might be acceptable. For exam-

ple, a partial shift could be made by the introduction of a filter that screened incoming questions and only forwarded those to each provider that matched that provider's expressed areas of interest. And/or, the system could gradually and seamlessly switch over to a system that completely shifts the costs of question and answer match-up to information seekers by an increased use of (improved?) FAQ and online help question and answer archives if and as provider willingness to respond to new posted questions declines.

8.2. General implications

Our analysis of the Apache online help system, in which help is provided by volunteer effort, presented an initial puzzle: why would information providers voluntarily help information seekers for free? We were able to reduce this puzzle considerably by disaggregating the total task of help provision into subtasks. This in turn allowed us to understand that 98% of the effort invested by help providers was intrinsically rewarding to those providers via a particular feature of the task setting. That is, we found that the public posting of both questions and answers created a site that potential information providers wanted to visit and study in order to gain valuable information for themselves. In addition, the public posting of answers with the names of providers attached created the possibility of gaining reputation and related benefits through helping. These specific features of help site design were probably the result of happenstance rather than intent—but they appear to be crucial to the successful functioning of the system we studied.

We draw a general conclusion from this result. We think that it is important to analyze the micro-level functioning of successful open source projects to really understand how and why they work. For example, we think it would be useful to conduct similar empirical studies to explore other puzzling aspects of how an open source project functions such as: how is coordination achieved among open source software contributors; how can problems be segmented into module of a size that fit the sources and incentives of individual users to effectively contribute?

The learning gained from such micro-studies of a range of tasks may well turn out to cumulate to some general principles. For example, it is interesting to discover that learning on the part of contributors is an important motivator in the case of the relatively "mundane" task of help-provision—just as it has been show to be for the task of code development. The learning gained can also help with the design of the next generation of open source projects and similar, user-based innovation systems.

Acknowledgements

We thank Brian Behlendorf, Matthew Gray, Ben Hyde and Robert Thau for sharing their deep insights into the functioning of the Apache free/open source community and thank Ben Ho and Starling Hunter for their helpful inputs during paper development. We also thank the participants on the Usenet newsgroup comp.infosystems.www.servers.unix who responded to our online survey. Doctoral student support from the Boston Consulting Group is gratefully acknowledged.

Appendix A. Questions asked in survey of Usenet newsgroup dedicated to Apache web server technical support

Our questionnaire was in three parts. Part 1 sent to information seekers differed from that sent to information providers.

A.1. Part 1—problem history and solution (this version sent to information seekers only)

1	What was the status of your problem when you posted it on the Usenet? (1—web server completely down; 2—web server functioning but missing important (necessary) features; 3—web server functioning but missing optional features; 4—installation problems; 5—upgrade problems)
2	Please rate how critical it was to get an answer to your problem (1—not so critical to 7—extremely critical)
3	How much time did it take you to formulate the question for posting on the Usenet? (minutes)

Appendix A.1. (*Continued*)

4	How many people (not listed in the posting URL) responded to your question via private e-mail?
5	Please describe the overall impact of your Usenet post and responses from both public and private authors in your problem solving process (1—did not solve my problem at all to 5—completely solved my problem)
6	Please estimate the problem solving time you saved due to the answers (public and private) provided for your post/question? (minutes)
7	Please rate the value and time spent (minutes) on the following Apache related resources in your problem solving process: resources listed in Table 4 (value ranged from 1—did not use to 7—high value)
8	Was the value to you of answers posted publicly significantly lower or higher than the value of those sent to you by private e-mail? (1—public lower than private; 2—public higher than private; 3—both had the same value)

A.2. Part 1—about your response to the posted question (this version sent to information providers only)

1	Did you previously know the person whose question you answered? (1—no previous contact; 2—yes, past interactions on Usenet; 3—yes, some other context)
2	What was the state of your information when you first saw the question? (1—knew solution; 2—knew where to find solution; 3—some (useful) information but not full solution; 4—no direct solution but had ability to solve the problem)

Appendix A.2. (*Continued*)

2A	If you already knew the solution to the post/question, why did you know it? (1—experienced same problem; 2—knew on the basis of general knowledge about Apache)
2B	If you already knew the solution to the post/question, how many other readers of questions posted on CIWS-U do you think also knew a solution? (1—many; 2—few, only people with good general expertise in Apache; 3—few, only people who had encountered a very similar problem)
3	What did you do to answer the post/question? (1—only provided information I already had; 2—searched for additional information that would be useful for the poster; 3—did some problem-solving to help the poster)
4	How much time did it take you to answer the question? (minutes)
5	There are many reasons as to why people choose to respond to posts on Usenet. Please indicate your level of agreement or disagreement to the following most common reasons people give for their motivation to post responses to questions related to Apache (1—strongly disagree to 7—strongly agree)
5A	If I answer question on CIWS-U, others are more likely to help me when I post a question in the future
5B	Others have helped me in the past on CIWS-U and I feel an obligation to reciprocate by answering questions on Apache Usenet
5C	Others have helped me in the past on other Usenet groups and I feel an obligation to reciprocate by answering questions on Apache Usenet

Appendix A.2. (*Continued*)

5D	Answering questions on Usenet enhances my career prospects
5E	Answering questions on Usenet can enhance my reputation in the Apache community and/or the open source community
5F	I answer questions on Usenet in order to promote the open source software/free software movement
5H	I answer questions on Usenet for fun
5I	I answer questions on Usenet as a break from other work
5J	I answered this question because I thought the poster might not get a good answer if I didn't
5K	I am an Apache Group member or module author and I want to assist people who are having problems with Apache or my module
5L	Part of my job description is to support the Apache software—answering questions is one way to do that part of my job
5M	I have an area(s) of expertise within Apache and try to answer all questions that come up in that area(s)

A.3. Part 2—about your CIWS-U reading pattern (information seekers and providers)

1	How often do you read CIWS-U? (1—daily; 2—weekly; 3—monthly; 4—only when I have an Apache related problem)
2	What is your approximate average time per session on CIWS-U? (minutes)
3	Please rate the following reasons for scanning CIWS-U (1—strongly disagree to 7—strongly agree)
3A	I browse CIWS-U to find and learn from message threads that contain information potentially relevant to my work
3B	I browse CIWS-U to find posted questions that I want to answer
3C	I browse CIWS-U because it is fun
3D	I browse CIWS-U as a break from other work

A.4. Part 3—backgrounds (information seekers and providers)

1	How long have you used HTTP servers? (months)
2	How long have you used Apache web server software? (months)
3	What percentage of your work-week is devoted to web server related functions? (percentage of work-week)
4	Is your Apache related work on a: (1—professional paid basis; 2—volunteer/hobby basis; 3—both)
5	Approximately, what is the average number of hits per day that your web server(s) gets? (hits per day)
6	Have you ever modified Apache source code or create new modules to suit your particular requirements? (yes/no)
7	Are you currently a student? (1—no; 2—yes, high school student; 3—yes, undergraduate student; 4—yes, graduate student)

References

Apache Group, 1999. Apache FAQ. http://httpd.apache.org/docs/misc/FAQ.html#support.

Bandura, A., 1995. Comments on the crusade against the causal efficacy of human thought. Journal of Behavior Therapy and Experimental Psychiatry 26 (3), 170–179.

Constant, D., Sproull, L., Kiesler, S., 1996. The kindness of strangers: the usefulness of electronic weak ties for technical advice. Organization Science 7 (2), 119–135.

Csikszentmihalyi, M., 1975. Beyond Boredom and Anxiety: The Experience of Play in Work and Games. Jossey-Bass, San Francisco.

Csikszentmihalyi, M., 1990. Flow: The Psychology of Optimal Experience. Harper & Row, New York.

Csikszentmihalyi, M., 1996. Creativity: Flow and the Psychology of Discovery and Invention. Harper Collins, New York.

Drake, A.W., Finkelstein, S.N., Sapolsky, H.M., 1982. The American Blood Supply. MIT Press, Cambridge, MA.

Ekeh, P., 1974. Social Exchange Theory: The Two Traditions. Harvard University Press, Cambridge, MA.

Fielding, R., 1999. Shared leadership in the Apache Project. Communications of the ACM 42 (4), 42–43.

Holmstrom, B., 1999. Managerial incentive problems: a dynamic perspective. Review of Economic Studies 66, 169–182.

Kollock, P., 1999. The economies of online cooperation: gifts and public goods in cyberspace. In: Smith, M.A., Kollock, P. (Eds.), Communities in Cyberspace. Routledge, London.

Lakhani, K.R., 1999. Sustaining the virtual commons: end user support for Apache web server software on the Usenet. Unpublished Masters thesis, Massachusetts Institute of Technology, Cambridge, MA.

Lakhani, K.R., Wolf, R., 2002. BCG Hacker Survey. Summarized study results available at http://www.osdn.com/bcg.

Lerner, J., Tirole, J., 2002. Some simple economics of open source. Journal of Industrial Economics 50 (2), 197–234.

Niedner, S., Hertel, G., Hermann, S., 2000. Motivation in Open Source Projects: An Empirical Study Among Linux Developers. Summarized study results available at http:/www.psychologie.uni-kiel.de/linux-study.

Prettejohn, M., 2001. Netcraft Web Server Survey. 20 September. http://www.netcraft.com.

Raymond, E., 1999. The Cathedral and the Bazaar: Musings on Linux and Open Source from an Accidental Revolutionary. O'Reilly and Associates, Sebastapol, CA.

Rheingold, H., 1993. The Virtual Community: Homesteading on the Electronic Frontier. Addison-Wesley, New York.

Thorn, B.K., Connolly, T., 1987. Discretionary Data Bases: A Theory and Some Experimental Findings. Communication Research 14, 512–528.

von Hippel, E., 1988. The Sources of Innovation. Oxford University Press, New York.

von Hippel, E., 2002. Open source projects as user innovation networks. MIT Sloan School of Management Working Paper 4366-02, June.

Wellman, B., Gulia, M., 1999. Virtual communities as communities. In: Smith, M.A., Kollock, P. (Eds.), Communities in Cyberspace. Routledge, New York.

Further reading

Mockus, A., Fielding, R.T., Herbsleb, J., 2000. A case study of open source software development: the Apache server. In: Proceedings of the Presentation of the Paper at the International Conference on Software Engineering (ICSE), Limerick, Ireland.

Tyre, M.J., von Hippel, E., 1997. The situated nature of adaptive learning in organizations. Organization Science 8 (1), 71–83.

von Hippel, E., 1994. Sticky information and the locus of problem solving: implications for innovation. Management Science 40 (4), 429–439.