A CHARACTERIZATION OF THE LINUX COMMUNITY OF PRACTICE USING

LINUX NEWSGROUPS AND

BALES' INTERACTION PROCESS ANALYSIS

by
Matthew Knuppel

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

May, 2000

Approved by

_____

Advisor

Matthew Knuppel. A Characterization of the Linux Community of Practice Using Linux Newsgroups and Bales' Interaction Process Analysis. A Master's paper for the M.S. in I.S. degree. May, 2000. 62 pages. Advisor: Diane H. Sonnenwald

The open source software community represents a new community of practice and approach to software development and distribution. It is important to study this community because their approach differs radically from the traditional development and distribution of proprietary software. One of the largest groups in the open source software community is the Linux operating system community. The community of users of the Linux operating system is a widespread group that frequently communicates through on-line newsgroups. These newsgroups serve as forums where the members may participate in ongoing discussions, or "threads," reading other members' messages and posting messages of their own. The participants can discuss issues relative to the Linux operating system, ask questions, provide technical support, and generally engage in the same activities that a face-to-face discussion group would engage in.

To increase our understanding of the Linux community, this study seeks to characterize the Linux community of practice through an analysis of these newsgroups. While quantitative studies of this community performed, there has not been much work done in the way of qualitative studies. The analysis is performed using categories developed by Bales' Interaction Process Analysis. After determining the newsgroups that we wanted to analyze, selecting specific discussions from these newsgroups, and coding the messages using Bales' categories of Interaction Process Analysis, we were able to establish a clear picture of the frequency with which specific behaviors occurred in the newsgroups.

We found that in general most of the behaviors that occurred in the messages posted to the newsgroups were attempted answers, and the second most frequently occurring behaviors in the messages were questions. There were a relatively small number of messages that displayed socio-emotional behavior.

Headings:

Social networks – Research

Computer networks – Social aspects

Internet (Computer network) – Social aspects

Cyberspace – Social aspects

Introduction

The goal of this paper is to relay the results of a statistical and content analysis of the Linux operating system user community. The purpose of these analyses is to provide a characterization of the Linux community of practice.

The open source software community represents a new community of practice and approach to software development and distribution. It is important to study this community because their approach differs radically from the traditional development and distribution of proprietary software. One of the largest groups in the open source software community is the Linux operating system community. The community of users of the Linux operating system is a widespread group that frequently communicates through on-line newsgroups. These newsgroups serve as forums where the members may participate in ongoing discussions, or "threads," reading other members' messages and posting messages of their own. The participants can discuss issues relative to the Linux operating system, ask questions, provide technical support, and generally engage in the same activities that a face-to-face discussion group would engage in.

To increase our understanding of the Linux community, this study seeks to characterize the Linux community of practice through an analysis of these newsgroups. While quantitative studies of this community performed, there has not been much work done in the way of qualitative studies. The analysis is performed using categories developed by Bales' Interaction Process Analysis. After determining the newsgroups that we wanted to analyze, selecting specific discussions from these newsgroups, and coding the messages using Bales' categories of Interaction Process Analysis, we were able to

establish a clear picture of the frequency with which specific behaviors occurred in the newsgroups.

We found that in general most of the behaviors that occurred in the messages posted to the newsgroups were attempted answers, and the second most frequently occurring behaviors in the messages were questions. There were a relatively small number of messages that displayed socio-emotional behavior.

However, the question arises as to how similar this mode of communication is to that of face-to-face communication. Literature in the field of computer-mediated communication (CMC) versus face-to-face communication provided us with some insight into this question.

Literature Review

The first section of the literature review will offer an historical overview of the articles, papers, and reports that have documented the Linux operating system to this point. The absence of academic papers is offered as further justification for the exploration of issues related to Linux.

The second section will discuss work that has been done by Bales' in the area of small group interaction, as well as other research related to face-to-face vs. computer aided communication, and the validity of using Bales' work in the analysis of on-line communities.

The Linux Community

A review of the literature about the Linux operating system and the specific areas of research related to it presents a problem. It is a relatively new phenomena, gaining general notoriety only recently with the initial public offering of Red Hat Software, Inc. stock, a company that distributes the Linux operating system along with documentation and technical support.

Because of its relative "newness," much of what is written about Linux exists in professional, not academic, journals. The technical aspects of the operating system are one area that is discussed, as well as the implications of Linux for the software industry and the business world in general. What is noticeably absent in the discussion of Linux is any in-depth analysis of the community behind it, the producers and users of the Linux kernel (the basic architecture of the operating system), and the various applications that it runs.

In one of the earliest writings about the Linux development community, "The Cathedral and the Bazaar," Eric Raymond (No date) discusses two very different approaches to the production of software code. In his introduction he states:

> Linus Torvalds's style of development   release early and often, delegate everything you can, be open to the point of promiscuity – came as a surprise. No quiet, reverent cathedral-building here – rather, the Linux community seemed to resemble a great babbling bazaar of differing agendas and approaches (aptly symbolized by the Linux archive sites, who'd take submissions from *anyone*) out of which a coherent and stable system could seemingly emerge only by a succession of miracles. (Raymond, No date, paragraph 4)

This touches upon the key issue behind our research, that of the need for an interaction process analysis of the "great babbling bazaar of differing agendas and approaches" that is the Linux collaborative community, a community that has heretofore been neglected.

A thorough history of the development of Linux is offered by Moody (1997) in "The Greatest OS That (N)ever Was." It presents some insights into the basic structure of the Linux development community:

> "The Linux and free software community can be thought of as true meritocracy," says Marc Ewing, who in 1994 founded Red Hat Software, which sells one of the most popular Linux distributions. "People in a traditional development group are assigned jobs that they may not know much about, or be best suited for." Bruno Haible, who has contributed to Linux's memory management code, puts it even more succinctly: "When the main author doesn't improve his code anymore, other people will." (Moody, 1997, paragraph 36)

Again, the Linux development community is portrayed as a true collaborative community, motivated by a desire to produce good code.

In analyzing this community, it would be desirable to have at least a general idea as to size of the community. Unfortunately, it would be impossible to determine the exact number of code producers and users worldwide, but there are metrics available to help estimate the number.

> Because Linux can be downloaded from hundreds of sites and users are encouraged to pass on CD-ROMs, it is impossible to accurately determine how many people use the OS. Based on an amalgamation of voluntary registration systems and market research, however, Linux distributor Red Hat estimates that there are between 3 million and 5 million users worldwide. (Moody, 1997, paragraph 48)

These are hardly accurate numbers, and hopefully not the last word in the determination of the size of the community. Other methods for determining the size of the community have been offered. According to Harald Alvestrand, project coordinator of The Linux Counter, "The more than 46,000 users who have chosen to register are less than 5 percent of the total number of Linux users …. I figure the true ratio is closer to 0.5 percent. That would indicate 9 million users" (Moody, 1997, paragraph 48). Regardless, we can be sure that the on-line Linux user community is not a small group.

There are critics of Linux, but even they must concede some of its merits. While it is not as user-friendly as other operating systems, this must be weighed against the greater community of other users who are willing to help with problems. After cautioning against a complete embrace of Linux in his article for salon.com, Leonard (1998) goes on to state:

> …Linux's unfriendliness to users who demand utter point-and-click simplicity should by no means rule it out for everyone …. Linux devotees regularly tout how readily help and support can be obtained on the Net, and I found that these claims are not exaggerated. Every time I posted a question, I received loads of immediate, friendly and exhaustive help. (Leonard, 1998, paragraph 13)

This illustrates again the presence of a huge community of Linux users, and their presence online. The use of newsgroups to exchange information about Linux and offer help to other users provides us with the content we need to perform an interaction process analysis. Newsgroup logs can be helpful in doing a content analysis of the data archives.

Web sites like Deja.com and RemarQ offer access to the archives of all the Linux

newsgroups.

Dempsey, Weiss, Jones, and Greenberg (1999) authored one paper that does seek

to analyze the Linux user community. As stated in the abstract:

> This paper contributes a baseline quantitative study of one of the oldest
> continuous repositories for the Linux open source project (the UNC MetaLab
> Linux Archives), including demographic information on its broad community of
> developers. Our methodology is a close examination of collection statistics,
> including custom monitoring scripts on the server, as well as an analysis of the
> contents of user-generated metadata embedded within the Archives. (Dempsey et
> al., 1999, 1)

This is an excellent starting point for a more thorough analysis of the Linux community.

While this paper analyzes the demographic of the community using the contributions to

the software archive at MetaLab, it is a purely quantitative analysis that does not get to

the heart of the collaborative process inherent in the Linux community.

An important point is made about the difficulty of finding hard data about the

Linux community:

> A prime difficulty in understanding and drawing conclusions about open source
> collaborative development has been the sketchy information available on exactly
> who participates in open source development and how their software archives
> evolve. This lack of information is understandable given the distributed, organic
> process of collaborative development in open source communities.
> (Dempsey et al., 1999, 2)

However, this is of course all the more reason to widen the amount of analysis that has

been performed on the Linux community, and pursue the question of the Linux

community's interaction process. This will provide a more qualitative set of data to add

to the quantitative work that has already been done.

Computer-Mediated Communication and Social Network Analysis

Bales' Interaction Process Analysis. The analysis of the communication process that

occurs in the Linux newsgroups is informed by the work of R.F. Bales in what he refers

to as the "interaction process." In his paper entitled "A Set of Categories for the Analysis

of Small Group Interaction," Bales (1950) proposes twelve different behaviors that can be

used to characterize the elements of a small group interaction:

> The heart of the method is a way of classifying behavior act by act, as it occurs in
> small face-to-face groups, and a series of ways of analyzing the data to obtain
> indices descriptive of group process, and derivatively, of factors influencing that
> process. (Bales, 1950, 258)

The behaviors characterized by Bales are broken into two main categories, *social-

emotional* areas and *task* areas. These are then broken down further. The social-

emotional behaviors consist of both *positive* and *negative*, while the task behaviors

consist of *attempted answers* and *questions* (see Appendix A for the entire table). By

using these behaviors to "score" an observed interaction process, patterns of behavior

emerge. The process of communication, as well as the occurrence of specific behaviors,

can be used to characterize the interaction. Bales intends this coding scheme to be used

on small, face-to-face group interactions, but suggests that it might be useful in other

arenas as well. Larger groups, and groups with a variety of purposes and goals could all

be analyzed using this scheme. Understandably, one scenario that Bales' did not suggest

at the time of his writing was the use of this coding scheme on groups that were not

interacting face-to-face, i.e., computer-mediated groups or groups interacting on-line.

Bales' Interaction Process Analysis and Computer-Mediated Communication. We must

ask ourselves, is the use of Bales' Interaction Process Analysis valid in the observation of

groups interacting in a CMC capacity? Are face-to-face and CMC interactions the same,

or at least similar enough to warrant the use of Bales' Interaction Process Analysis?

> Garton, Haythornthwaite, & Wellman (1999) state:
>
> When a computer network connects people or organizations, it is a social network. Just as a computer network is a set of machines connected by a set of cables, a social network is a set of people (or organizations or other social entities) connected by a set of social relations, such as friendship, co-working, or information exchange. (Garton et al., 1999, 75)

This supports our view that Bales' Interaction Process Analysis can be used for the

observation of an on-line community, but what about the small group aspect of Bales'

work? On-line communities can be quite large, and may be difficult to observe.

However, Garton et al. (1999) make the observation that:

> As widespread communication via computer networks develops, analysts need to go beyond studying single users, two-person ties, and small groups to examining the computer-supported social networks (CSSNs) that flourish in areas as diverse as the workplace and virtual communities. (Garton et al., 1999, 75)
>
> The best approach to studying on-line communities such as the Linux user

community would appear to be to treat the on-line community like any other social

network when performing an analysis of it. An analysis of an on-line community should

be informed by the same research, and performed with the same tools, as an analysis of

the small, face-to-face group:

> Because computer networks often are social networks, the social network approach gives important leverage for understanding what goes on in CMC – how CMC affects the structure and functioning of social systems (be they organizations, workgroups, or friendship circles) and how social structures affect the way CMC is used. (Garton et al., 1999, 100)
>
> Rice (1994), in "Network Analysis and Computer-Mediated Communication

Systems," makes the distinction between face-to-face and computer-mediated

communication:

> Text-based CMC systems may reduce the amount of social presence or information richness … and other cues … in the content of the communication, thus limiting the applicability of the CMC systems for more socioemotional communication activities. (Rice, 1994, 182)

While this does not seem to support our use of Bales coding scheme for the

analysis of the newsgroups, Rice goes on to say:

> However, the reduction of these nonverbal and social cues may improve the equality of participation and access by those otherwise constrained in interpersonal communication (such as employees with lower organizational status, who have speech difficulties, or who are members of minority groups). (Rice, 1994, 183)

While newsgroups may not reveal the social cues that face-to-face communication does,

this is made up for by the equality of participation that is promoted.

One of the earliest studies of computer-mediated communication to make use of

Bales' scheme of categories was that of Hiltz, Johnson, and Agle (1978), entitled

"Replicating Bales Problem Solving Experiments on a Computerized Conference: A Pilot

Study." They sought to analyze the different behaviors exhibited by groups performing

problem-solving activities in both face-to-face and computerized settings. There were

some differences noted between the two groups, as stated here:

> The main differences are in categories 3 (agreement) and 10 (disagreement). There was a lot more overt agreement communicated among the members of the face-to-face groups than was typed into the conferencing system. (Hiltz & Turoff, 1993, 110)

However, because of the small number of groups used in the analysis (12), other

differences were not evident:

> Most of the other differences that are statistically significant are substantively so small that we hesitate to say that they might mean anything. They might very well be due to lack of reliability in coding …. We do think that there is a good chance that a larger sample of groups would produce a statistically significant tendency toward more people giving more opinions in the CC condition than in the face-to-face condition. (Hiltz & Turoff, 1993, 111)

Extending this pilot study, Hiltz, Johnson, and Turoff (1986) continue to use Bales' interaction process analysis scheme to examine the differences between face-to-face and computer aided communication. "Experiments in Group Decision Making" by Hiltz et al. seeks to discover the differences in the process and outcome of groups using two different modes of communication, face-to-face and computerized. These groups were also given two different tasks to complete as a group. Bales' Interaction Process Analysis was used to code the interaction process:

> The explicit, verbalized content of communications in FtF conferences was compared with computerized conferences, using Bales's (1950) original categories for Interaction Process Analysis. Two types of problems were used: a complex ranking task with a criterion solution, and a qualitative, value-laden human relations task. (Hiltz et al., 1986, 242)

The literature that exists concerning the Linux community suggests that this is a field deserving of more analysis, especially qualitative analysis, as there is already some work in the way of quantitative analysis of the community. The literature that discusses the parallels between face-to-face and computer-mediated communication suggests that we may use Bales' Interaction Process Analysis effectively in our qualitative analysis of on-line newsgroups. In addition, the presence of literature that discusses the use of Bales' Interaction Process in a computer-mediated communication environment further strengthens this notion.

Methodology

Several steps needed to be taken in order to ensure that the data that we collected was valid. First, we needed to determine which tools we would use to identify the pool of newsgroups from which we would select the newsgroups for this study. There were several tools available to us, and each had its strengths and weaknesses. Second, after determining which tool to use, we had to determine which newsgroups we would analyze for this study. Third, we had to determine which discussions, or threads, we would analyze within the newsgroups we had selected. The final step was to code these threads using Bales' Interaction Process Analysis.

Determining the tools with which to identify potential newsgroups for analysis

There are a number of tools available for accessing both newsgroup archives and newsgroup statistics. Online archives such as Deja.com and RemarQ offer access to newsgroup archives going back several months in time. They also offer statistics on newsgroup membership and discussions. Advanced searching techniques can be used to pinpoint specific members of the community, specific messages, and any grouping of these based on subject matter, domain of the member, or dates of messages. Another available tool is the tin newsreader, which allows for access to newsgroups on a telnet session.

tin newsreader. Using the tin newsreader, one can subscribe to newsgroups using a wildcard pattern to return all of the newsgroups that contain a specific word in their title.

Figure 1. tin news reader

Figure 2. Deja.com web page

To return all of the newsgroups with linux in their title, one simply has to subscribe using

the wildcard expression '*linux*.' This returns 274 newsgroups, with only a fraction of

these being actively used. The number of messages currently in the archive of messages

is displayed along with the title of the newsgroup and occasionally a short description of

the newsgroup's content (see Figure 1).

Deja.com. Deja.com is a newsgroup archive that offers search capabilities for the

location of specific newsgroups. It is located at http://www.deja.com (see Figure 2).
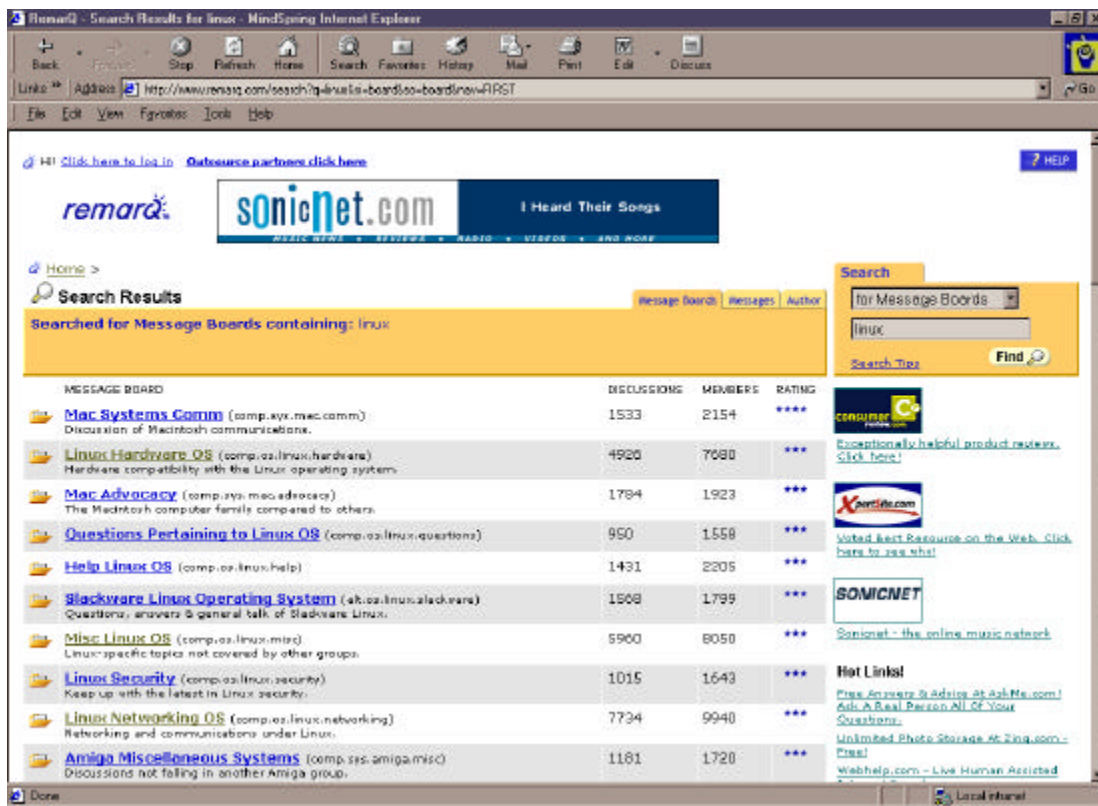

Figure 3. RemarQ web page

RemarQ. RemarQ is located at http://www.remarq.com and is similar to Deja.com in its

usage and features. However, RemarQ offers additional information of newsgroup

membership and available discussions. For each newsgroup, the number of members

currently subscribed and the number of discussions, or threads, that are contained in the

archive are displayed. The archive goes back roughly three months from the current date,

with older messages expiring and becoming unavailable (see Figure 3).

Final selection of newsgroup tool

The Deja.com and RemarQ tools contain more powerful features to locate specific

newsgroups and their statistics. It is for this reason that the tin newsreader was not

employed for this analysis. The same and additional information concerning Linux

newsgroups can be gleaned from the Deja.com and RemarQ sites with the added benefit

of having a graphical user interface.

Between these two tools, the RemarQ site is more useful for this analysis because

of the membership statistics that it offers. This allowed us to take the first step in our

analysis, which was to determine which newsgroups we should explore.

Determining which newsgroups to select for analysis

It should be noted here that all of these newsgroups and their discussion threads

and messages are in the public domain and easily accessible on the Internet. Using the

RemarQ site as our tool, we found 68 newsgroups that contained some reference to the

Linux operating system, and an additional 32 newsgroups that were conducted in a

language other than English. This made for a total of exactly 100 newsgroups.

While it is viable to use the number of actual *messages* currently in the newsgroup

archive to rank the newsgroups, we chose instead to rank the newsgroups by the number

of *members* currently subscribed. This guaranteed us the largest possible number of

different voices in the online communities that we investigated, rather than the largest

number of posted messages. It is conceivable, though unlikely, that one individual could

post thousands of messages to a newsgroup, thus giving the illusion that the newsgroup is very active and has a high number of contributing members. Of course, it is also possible that a newsgroup could be populated by members that only read messages and never post them ('lurkers'), leaving the majority of the communications to a handful of the members. However, by ranking the newsgroups by membership, we are at least moving in the direction of diversity.

We also considered ranking the newsgroups by the number of *threads* currently in the archive. However, the number of threads is closely correlated with the number of members in the newsgroup (Figure 4), so we did not consider this approach, as it would yield nearly the same ranking and not make a theoretical contribution to our sampling approach.

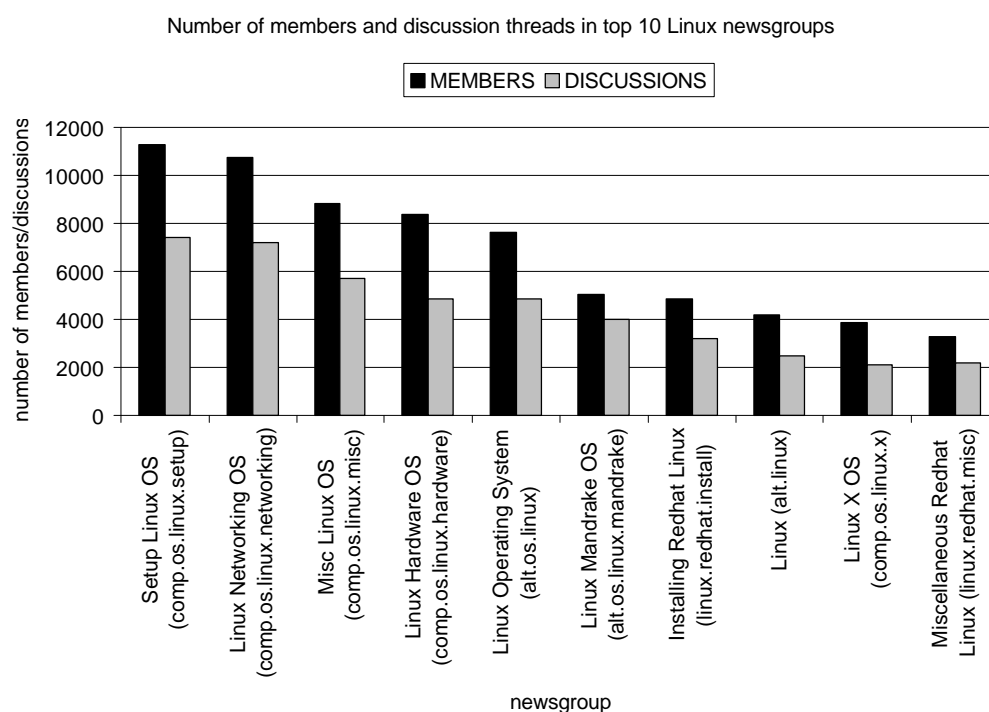Number of members and discussion threads in top 10 Linux newsgroups

Figure 4. Number of members and discussion threads in top 10 Linux newsgroups

Once we determined a ranking for the newsgroups, we needed to evaluate which newsgroups to include in our data analysis. For the purposes of this study we wanted as diverse a selection of discussion topics as possible. Newsgroups that deal with very specific topics will not display this diversity. To this end, newsgroups were not selected if they cover distributions of Linux, are intended for a particular nationality or community that uses a specific language (for example, the 32 non-English speaking newsgroups that we found), or are based on geographical location or locality. This allowed us only the most diverse collection of subject matter and geographic location (with the obvious and unfortunate exception of those members of the Linux community who do not speak English).

Determining which threads to analyze

Because the Linux newsgroups contain anywhere from 1 to over 7000 discussion threads, it is outside the scope of this paper to examine all of the threads.

Initially it would appear that using the five longest threads from each newsgroup would provide us with a good collection of data to analyze. However, this proved to be problematic from the standpoint of variety – many of the longer threads were discussions on the same subject (the Linux operating system vs. the Microsoft Windows operating system). Selecting the top five longest threads from each newsgroup would provide a very homogenous data set for analysis.

Therefore, to achieve some semblance of variety (which was not present in the five longest threads in each newsgroup) among the threads, threads of varying lengths were selected based on the length of the longest thread. This selection was made using

various percentages of the number of messages in the longest thread. To select five threads from each newsgroup, the longest thread from the newsgroup was used and considered to be valued at 100%. From there, threads were selected at values of 80% of the number of messages in the longest thread, 60%, 40%, and 20%. This approach yielded a satisfactory heterogeneity of thread lengths, and allowed for the inclusion in the analysis of various subjects.

Of course, the use of percentages did not yield a distinct group of five threads in each newsgroup. With the exception of the thread valued at 100%, the other percentages could include several threads that were valued in the vicinity of that percentage value. Rather than subjectively choose one of the threads that met the value of the percentage, we considered additional selection criteria.

First, threads that are cross-posted across multiple newsgroups were deleted from the pool of choices. Some threads were posted on several of the newsgroups, and it would not serve our purposes in this study to analyze the same thread multiple times because it occurs in different newsgroups (although this would be interesting in a related study to follow the path of a thread as it made its way through different segments of the community). However, for our purposes it seems more valuable to look at as many different subjects in the course of our analysis.

Second, because our goal is to assemble a variety of threads on a variety of subjects, it was decided not to consider threads that contained subject matter that was similar to that of other threads. For example, while all of the threads in comp.os.linux.networking would most probably have threads that addressed questions in the area of networking, it would not serve us well to analyze multiple discussions on the

same specific networking problem. That is, if different threads discussing the same topic are selected for analysis it is likely that we would continue to hear from the same members of the community, namely those members that are experts and/or interested in the question at hand. This is consistent with our goal of compiling a useful cross section of the community, with as many different participants as possible.

This approach yielded a selection of threads from a selection of newsgroups, a valid and varied group of participants, messages, and discussions. Every message in the selected threads was analyzed and coded using the categories outlined in Bales' Interaction Process Analysis (see Appendix B for a list of all newsgroups and threads analyzed in this study).

Content analysis using Bales' Interaction Process Analysis

After the sample of threads was collected, the individual messages within each thread were analyzed. Bales' method of Interaction Process Analysis (1950) was used to code each message in order to develop a deeper understanding of the characterization patterns.

As discussed previously, Bales' list of categorizations was not intended to address the interaction process as it occurs in an online community. Most obvious is the lack of face-to-face communication in this environment. However, Bales' categorizations will still be valuable in informing our analysis of this community because it is a comprehensive, well-validated coding scheme for interaction analysis.

For each of the messages, it was determined which of interaction behaviors in Bales' schema occur in the message. The coding was done in a binary fashion; when the

author exhibits a behavior in a specific message, that message is coded with a '1' to indicate that this behavior is present in the message. If a behavior is not observed, the behavior is coded with a '0' for that message. Multiple occurrences of the same behavior were not coded because we are simply seeking to determine if the behavior occurs in any given message, not the frequency with which it occurs in that message. Through this process of coding, a pattern of behaviors among the various newsgroup messages in this particular thread emerged. Both the prevalent social atmosphere of the thread and the social process of the thread are illustrated by the pattern of 1s and 0s in the coding table. Certain words and phrases served to indicate which behaviors were being exhibited in the messages (see Appendix C for a list of the various behavior coding indicators). In addition, the sum of the 1s in the table provides us with a clear picture of the number of messages that contain each of the twelve specific behaviors in Bales' schema. Each of these sums can be set aside the others to gauge their frequency as related to the others in the group. The sums were used to determine the percentage of messages in each thread that contained a specific behavior, and then the threads were set aside each other to gauge their similarities and differences. In addition, the results obtained from each of the five threads in each newsgroup were combined to yield an overall picture of the newsgroup itself. We were provided with the overview we were seeking of the general behaviors of the participants in each thread, and more importantly, the members of the different newsgroups.

The Coding Process

A general description of the coding process will help put the methodology described above in perspective. It should be noted, the threads were analyzed for the number of messages that contain any given behavior categorized by Bales in the Interaction Process Analysis. They were not analyzed based on the order in which the behaviors occurred in the thread or the process that the discussion group underwent in interacting with one another.

While the discussions were not analyzed for their interaction process, it is still useful to see the coding schema put into practice:

IPA categories

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	1	1	0	0	0	0
2	0	0	0	0	1	0	0	0	0	0	0	0
3	0	0	0	0	0	0	1	0	0	0	0	0
4	0	0	0	0	1	1	0	0	0	0	0	0
5	0	0	0	1	1	1	0	0	0	0	0	0
6	0	0	0	0	1	1	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	1	1	0	0	0	0	0	0
total	0	0	0	1	6	4	3	1	0	0	0	0

Figure 5. Example of categorization scheme for newsgroups

The grid shown (Figure 5) is an example of the coding scheme that was utilized in the categorization of the messages in the discussions. Across the top of the grid are numbers representing the 12 categories in Bales' Interaction Process Analysis. Along the side of the grid are numbers representing the 9 messages in this particular discussion. 0s are essentially placeholders, indicating that there is no occurrence of that behavior in the message. 1s indicate that the behavior does occur in the message. Using this approach,

we can see the general process of the discussion. A step-by-step description of the above grid will help clarify the coding scheme and what it reveals.

For the complete set of messages analyzed in the following coding process, see Appendix D. In the first message someone asks the group for orientation or opinion (7 and 8). The second message indicates that a reply is given that contains an opinion (5). The third message indicates that someone again requests orientation (7). The fourth message contains the expression of an opinion (5), and orientation (6). The fifth message contains a suggestion (4), the expression of an opinion (5), and orientation (6). The sixth message also contains the expression of an opinion (5), as well as orientation (6). The seventh message contains an opinion (5). The eighth message contains a request for orientation (7). Finally, the ninth message expresses an opinion (5) along with orientation (6).

After determining the newsgroups that we wanted to analyze, selecting specific discussions from these newsgroups, and coding all the messages in the twenty-five discussions in the five newsgroups using Bales' categories of Interaction Process Analysis, we were able to establish a clear picture of the frequency with which specific behaviors occurred.

Discussion

To increase our understanding of interaction in the open source software community and the Linux operating system open source software community in particular, each of the five discussions from each of the five newsgroups will be discussed, as well as the five newsgroups in general. The graphs illustrate the percentage of messages in each discussion that contain at least one instance of the behaviors described in Bales' Interaction Process Analysis schema. Because each discussion contains a different number of messages, the graphs have been normalized to indicate the percentage of messages as they relate to the total number of messages for that discussion. The five graphs from each of the five newsgroups have been categorized by length, ranging from "longer" (while some of the discussions were indeed the longest in their newsgroup, they were not all necessarily so, so the category "longer" has been used) through "long," "medium," and "short," to "shorter" ("shortest" was not considered because none of the discussions were the shortest in their newsgroup). The graphs will be discussed in order from "longer" to "shorter" within each newsgroup. The newsgroups will be discussed in order of membership, from the most members to the least.
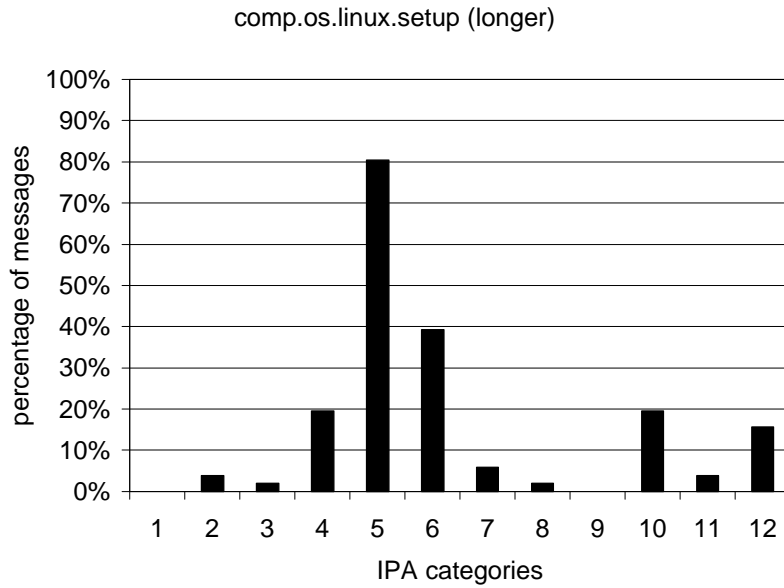
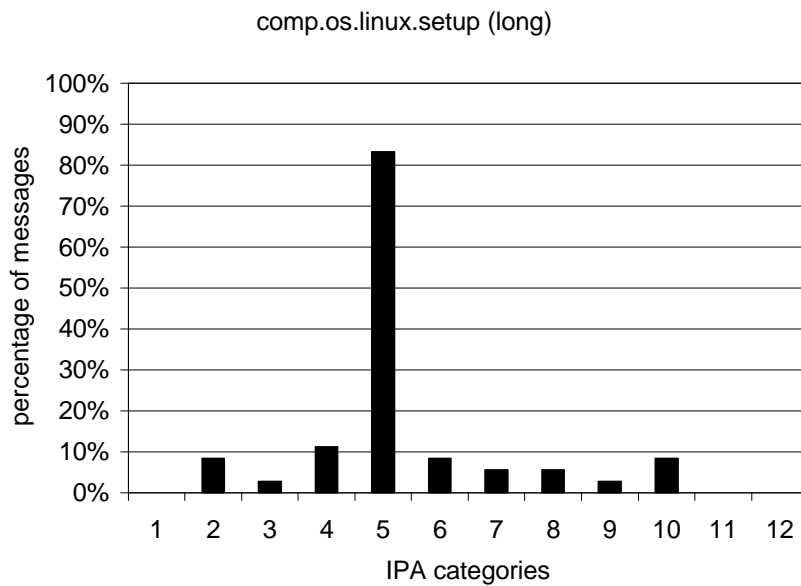<u>comp.os.linux.setup Newsgroup</u>

comp.os.linux.setup (longer)

<u>Figure 6-a.</u> comp.os.linux.setup "longer" discussion

comp.os.linux.setup (long)

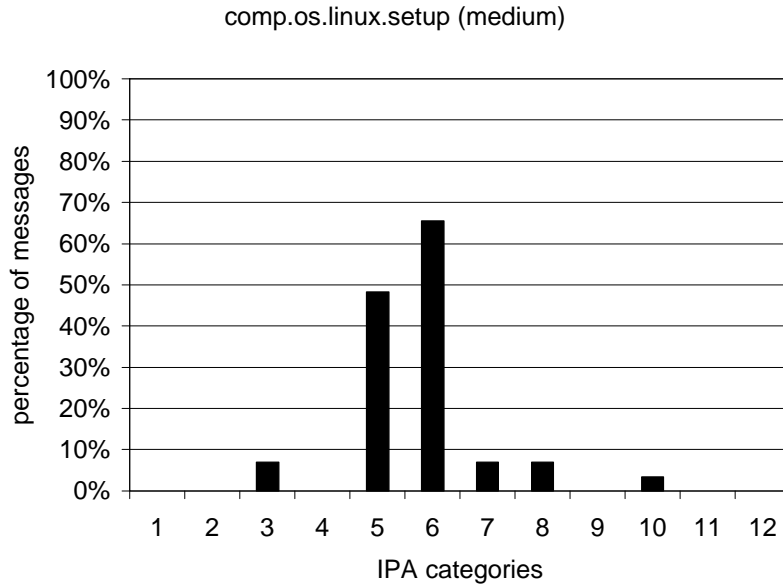<u>Figure 6-b.</u> comp.os.linux.setup "long" discussion

comp.os.linux.setup (medium)

Figure 6-c. comp.os.linux.setup "medium" discussion

comp.os.linux.setup (short)

Figure 6-d. comp.os.linux.setup "short" discussion
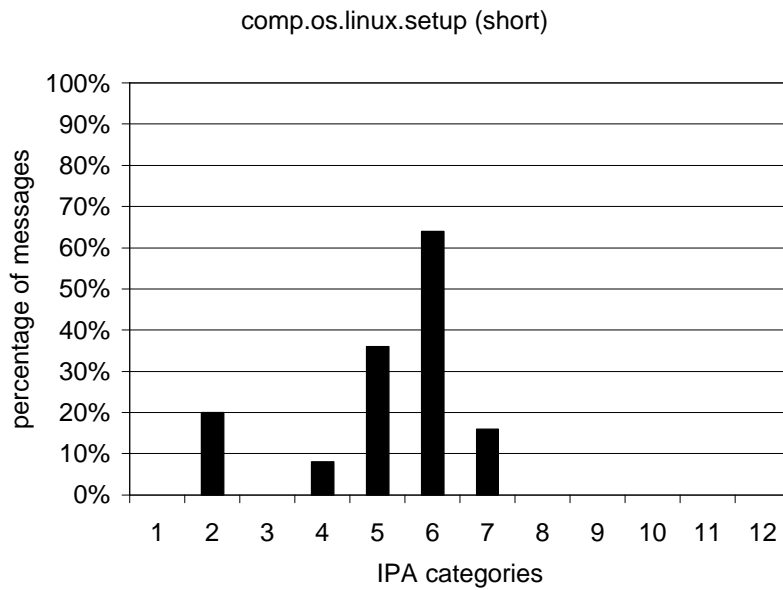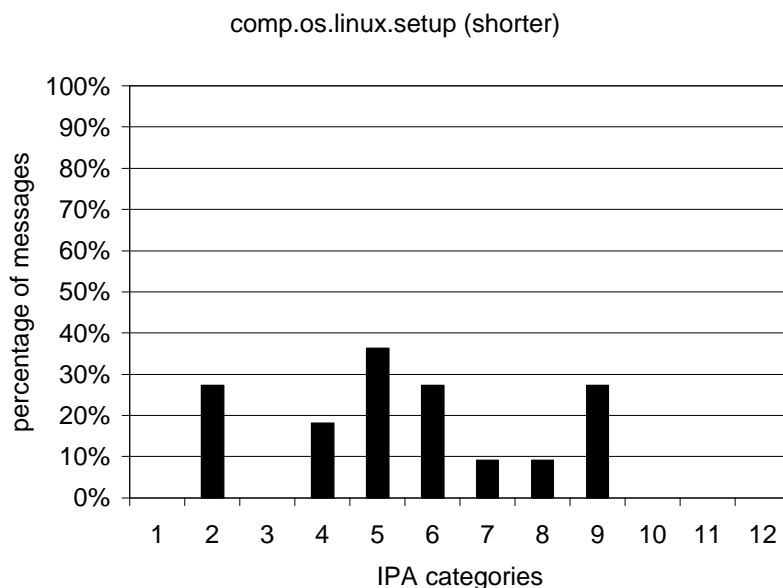
comp.os.linux.setup (shorter)

Figure 6-e. comp.os.linux.setup "shorter" discussion

The "longer" discussion (Figure 6-a) in this newsgroup consisted of 51 messages.
This discussion covered the relative merits of the Linux operating system. Discussions of
this type seemed to bring out an array of behaviors. The most frequent behavior by far is
category 5 ("Gives opinion"). Category 6 ("Gives orientation") was next in frequency,
with about half the messages exhibiting that behavior as category 5. The next two most
frequent message types were category 4 ("Gives suggestion") and category 10
("Disagrees") behavior. Category 12 ("Shows antagonism") was also noticeable in this
discussion.

The "long" discussion (Figure 6-b) consisted of 36 messages. The subject of this
discussion was again the merits of the Linux operating system, but this time focused on a
specific distribution of Linux (Red Hat Linux). There is a dramatic percentage of
messages (over 80%) that exhibit category 5 ("Gives opinion") behavior. The other
categories were much lower, all save one (category 4) occurring in less than 10% of the

messages. Unlike the "longer" discussion, there was a complete lack of antagonistic behavior (12) in this discussion.

The "medium" discussion (Figure 6-c) consisted of 29 messages. This discussion covered a technical matter, and it shows an interesting departure from the previous two discussions. Category 6 ("Gives orientation") is the prevalent behavior in this discussion, with category 5 ("Gives opinion") following. Perhaps because of the straightforward nature of the discussion, there were less opinionated answers and more informational answers. There is a complete lack of the extreme behaviors on either end of the scale, and a small amount of both agreement (3) and disagreement (10).

The "short" discussion (Figure 6-d) consisted of 25 messages. The subject of this discussion was again a technical matter, and shows even more concentration toward the center of the scale. Category 6 ("Gives orientation") was again the prevalent behavior, with category 5 ("Gives opinion") following. The questions in this discussion only seek orientation (7), not opinions (8) or suggestions (9). There is also a noticeable amount of tension release and joking (2) in this discussion.

The "shorter" discussion (Figure 6-e) consisted of 11 messages. This discussion covered a rather simple technical matter, and apparently did not require an extended discussion. Category 5 ("Gives opinion") is the most frequently occurring behavior, though not by a significant amount. The reason for this is most likely because of the short length of this discussion. No category had the opportunity to occur more frequently than any of the others.

comp.os.linux.networking Newsgroup

comp.os.linux.networking (longer)

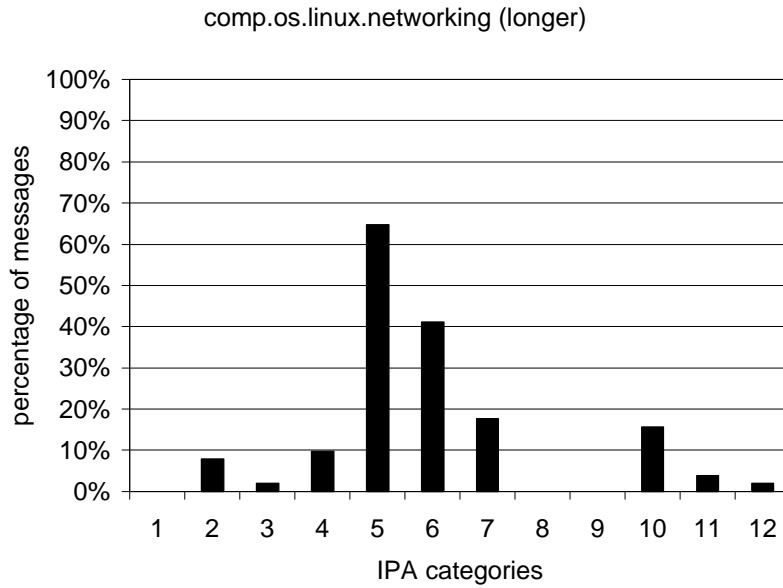Figure 7-a. comp.os.linux.networking "longer" discussion

comp.os.linux.networking (long)
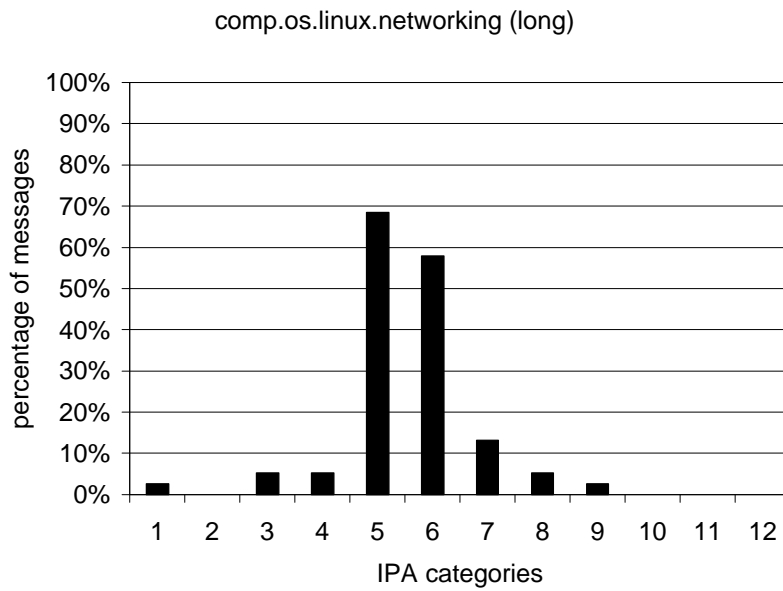

Figure 7-b. comp.os.linux.networking "long" discussion

comp.os.linux.networking (medium)
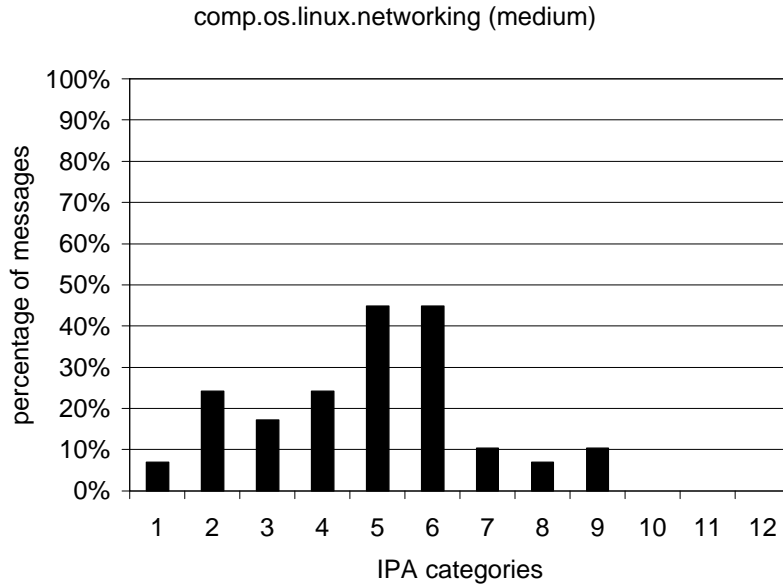

Figure 7-c. comp.os.linux.networking "medium" discussion

comp.os.linux.networking (short)

Figure 7-d. comp.os.linux.networking "short" discussion
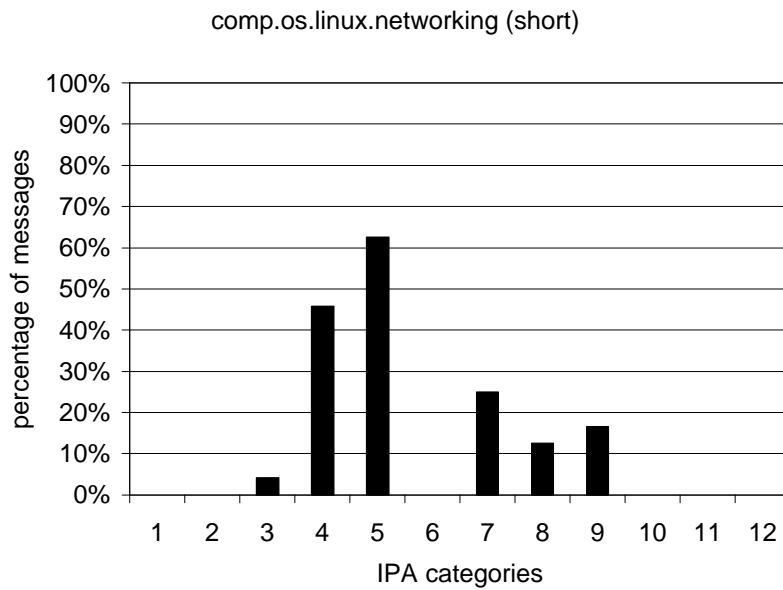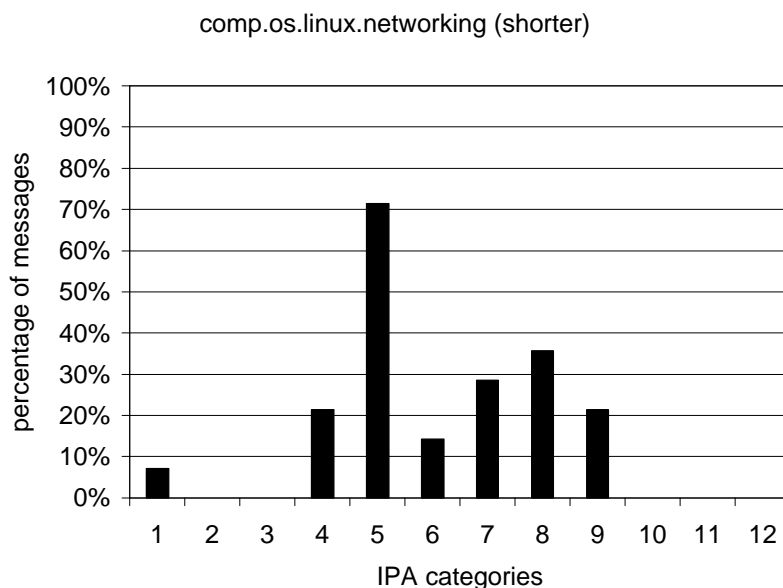
comp.os.linux.networking (shorter)

Figure 7-e. comp.os.linux.networking "shorter" discussion

The "longer" discussion (Figure 7-a) consisted of 51 messages. The

subject of this discussion was a combination of both technical issues and newsgroup

members' opinions on a non-technical topic. The prevalent behavior was category 5

("Gives opinion"), with category 6 ("Gives orientation") following. While the questions

in the discussion only sought orientation (7), the answers provided opinions (5),

orientation (6), and suggestions (4) in that frequency. The initial question that started the

discussion was indeed a simple request for information, but it began a rather spirited,

opinionated discussion (as evidenced by the occurrences of disagreement (10), tension

(11), and antagonism (12), as well as some tension release (2)).

The "long" discussion (Figure 7-b) consisted of 38 messages. This discussion

covered a technical matter requiring a great deal of clarification (6), as seen in the other

discussions covering technical matters. However, it was category 5 ("Gives opinion")

that was the most prevalent in the messages of this discussion. The other behaviors do

not occur to any significant degree, though this is the first time we see occurrences of solidarity expression (1).

The "medium" discussion (Figure 7-c) consisted of 29 messages. The subject of this discussion was a technical problem that one of the newsgroup members was experiencing. Questions seeking orientation (7), opinion (8), and suggestion (9) were all asked, and were met with an equal number of expressions of orientation (6) and opinion (5). Suggestions (4) were offered as well. There was a complete lack of negative social-emotional reactions, but a noticeable amount of positive ones (agreement (3), joking (2), and expressions of solidarity (1)). This could be explained by the successful outcome of the problem-solving tasks that occurred in this discussion. When the problem was solved the group displayed signs of satisfaction.

The "short" discussion (Figure 7-d) consisted of 24 messages. This discussion covered a technical problem that one of the newsgroup members was experiencing. There is an interesting dynamic in this discussion, as there are questions seeking information (7), opinions (8), and suggestions (9), yet there are no answers given that simply state facts. All of the answers are either opinions (5) or suggestions (4). This could be explained by the obtuse nature of the problem. There did not seem to be any simple answer to the questions at hand, so the newsgroup members were left to speculate. There was a small amount of agreement (3) in the discussion, but otherwise no behaviors in the social-emotional area.

The "shorter" discussion (Figure 7-e) consisted of 14 messages. The subject of this discussion was again a simple technical matter. The questions in the discussion sought orientation (7), opinions (8), and suggestions (9), while the related answers

provided these same things. Of the behaviors, category 5 ("Gives opinion") was again

the most prevalent. Save for some expressions of solidarity, there were no occurrences in

the social-emotional area.

comp.os.linux.misc Newsgroup

comp.os.linux.misc (longer)
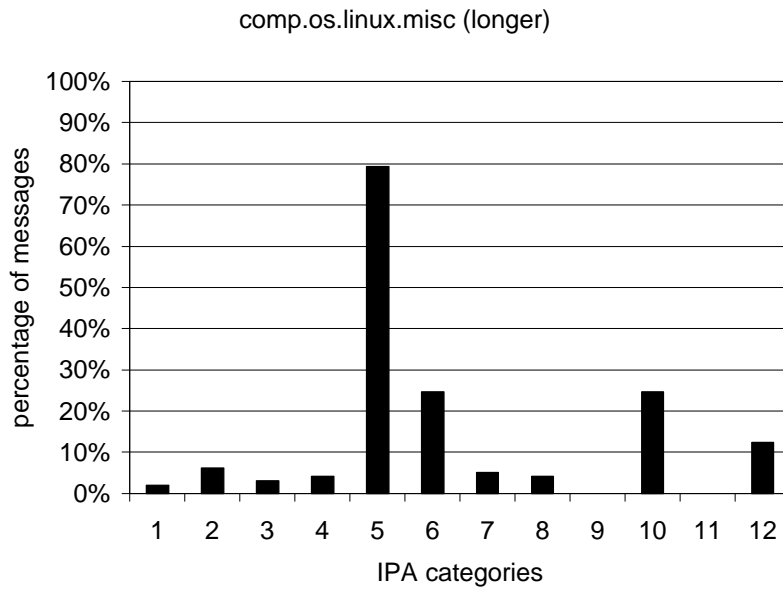
Figure 8-a. comp.os.linux.misc "longer" discussion

comp.os.linux.misc (long)

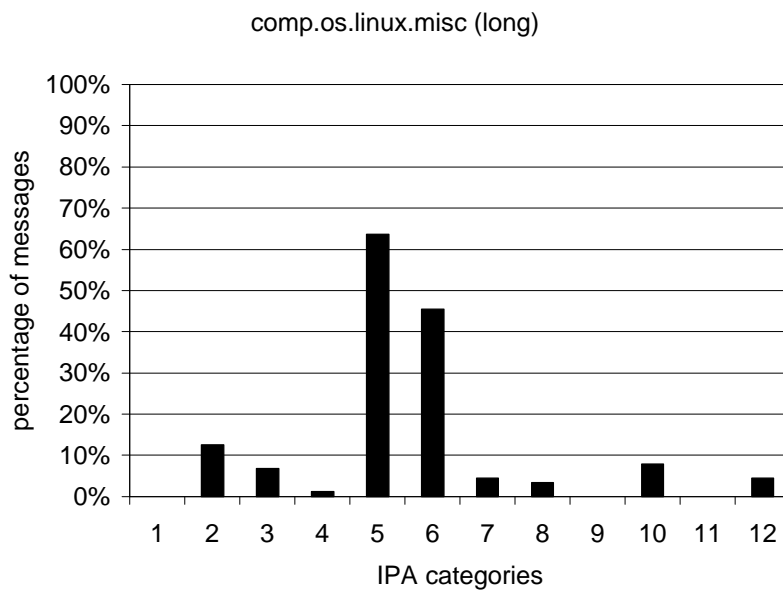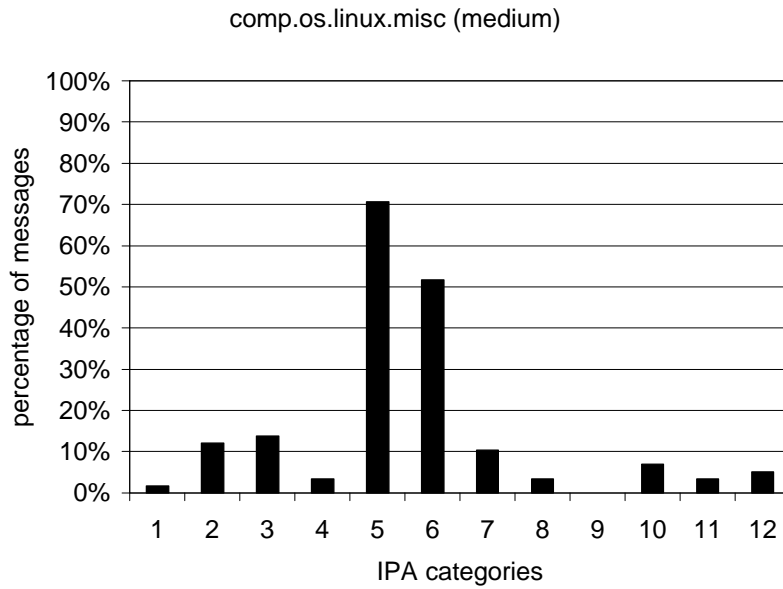

Figure 8-b. comp.os.linux.misc "long" discussion

comp.os.linux.misc (medium)

Figure 8-c. comp.os.linux.misc "medium" discussion

comp.os.linux.misc (short)
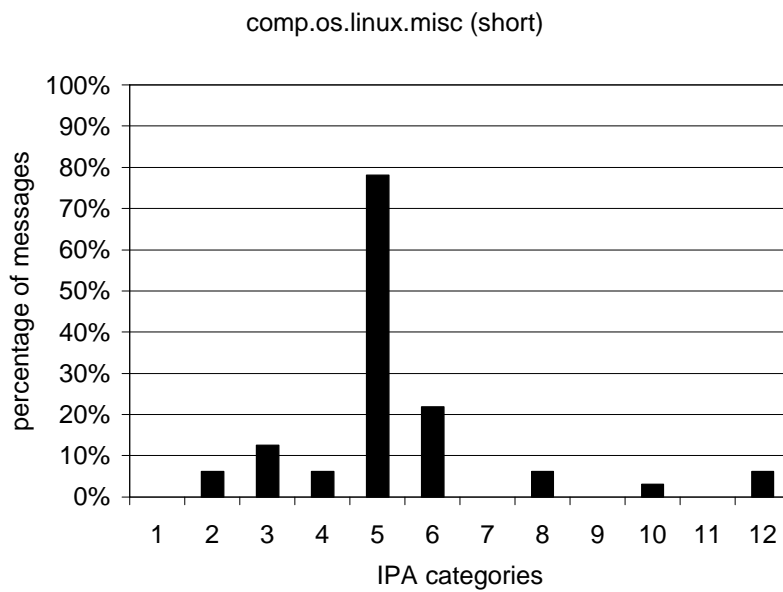
Figure 8-d. comp.os.linux.misc "short" discussion

comp.os.linux.misc (shorter)
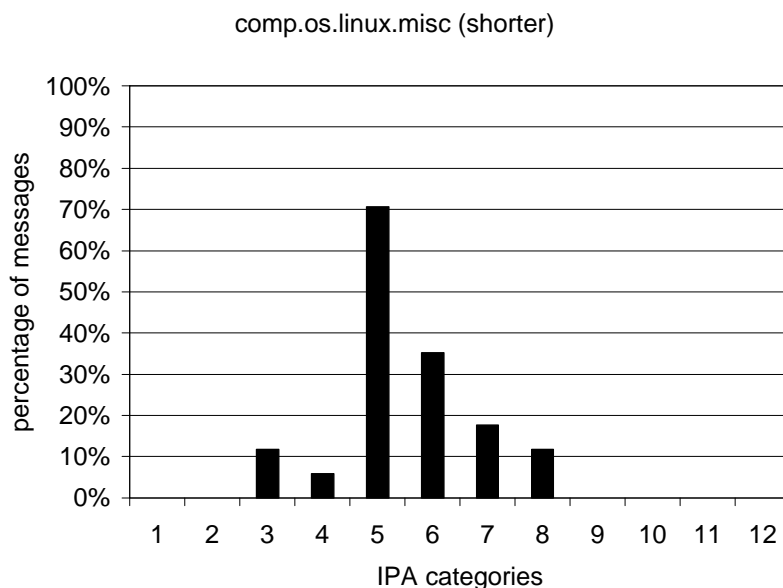
Figure 8-e. comp.os.linux.misc "shorter" discussion

The "longer" discussion (Figure 8-a) consisted of 97 messages. The subject of

this discussion was actually a departure from the norm, discussing the Java programming

language more than the Linux operating system. It appears this brought more

programmers into the discussion than usual. In this discussion we see a prevalence of

category 5 behavior ("Gives opinion"), but also the most disagreement (10) that we have

seen thus far in any discussion. The number of messages that contain disagreement is

equal to the number that provide orientation (6). This is explained by the atmosphere of

the discussion, which divided the group into two camps that each felt their approach to

the problem at hand was the correct one. The amount of antagonism (12) that this

elicited was also noticeable.

The "long" discussion (Figure 8-b) consisted of 88 messages. This discussion

covered a technical matter with some room for opinion, and contained what now appears

to be the prevalent distribution of behaviors in the discussions. Category 5 ("Gives

opinion") occurred in most of the messages, with category 6 ("Gives orientation")

following. The only other category that occurred in over 10% of the messages was category 2 ("Shows tension release"), and this was balanced on the other end of the scale with a combination of disagreement (10) and antagonism (12).

The "medium" discussion (Figure 8-c) consisted of 58 messages. The subject of this discussion was denial of service attacks and security issues related to Linux. Again, there was a predictable distribution of behaviors among the messages of the discussion. Category 5 ("Gives opinion") and category 6 ("Gives orientation") occurred in the most messages. Of the remaining behaviors, only expressions of agreement (3) and tension release (2) occurred in more than 10% of the messages. There were, however, noticeable occurrences of negative behaviors in the social-emotional area.

The "short" discussion (Figure 8-d) consisted of 32 messages. This discussion covered a very specific realm of Linux (text editors), and was started by a newsgroup member seeking others' opinions (8). These opinions were then offered (5) in a significant amount of the messages (nearly 80%). Some of the messages also displayed antagonism (12) along with the opinions.

The "shorter" discussion (Figure 8-e) consisted of 17 messages. The subject of this discussion was web browsers. This was a straightforward discussion, with a small number of questions seeking information and opinions balanced by a larger number of messages offering that information and opinion. With the exception of some agreement (3), there were no behaviors in the social-emotional area. As noted earlier, it is possible that short discussions like this one never provided the newsgroup members participating in the discussion with the reason or opportunity to express strong opinions combined with

more extreme social-emotional behaviors, and therefore reveal behaviors concentrated in

the center of the scale.

comp.os.linux.hardware Newsgroup

comp.os.linux.hardware (longer)

Figure 9-a. comp.os.linux.hardware "longer" discussion

comp.os.linux.hardware (long)
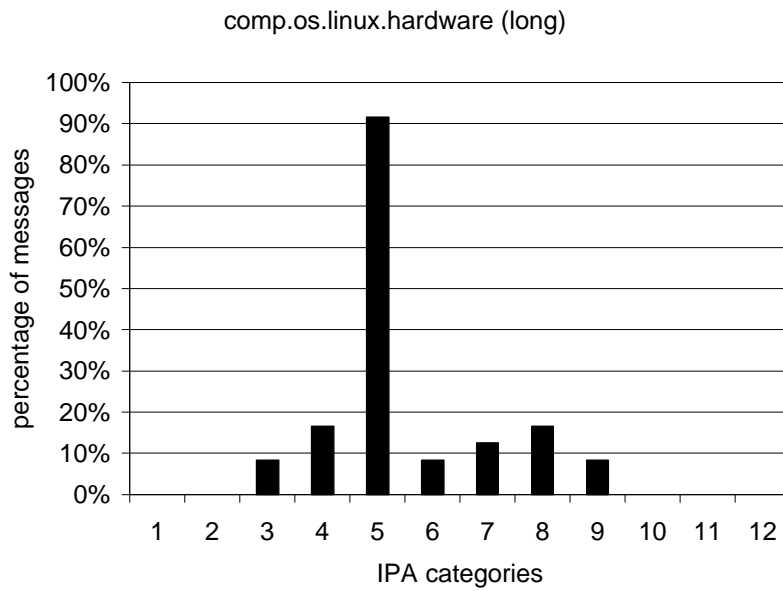
Figure 9-b. comp.os.linux.hardware "long" discussion

comp.os.linux.hardware (medium)
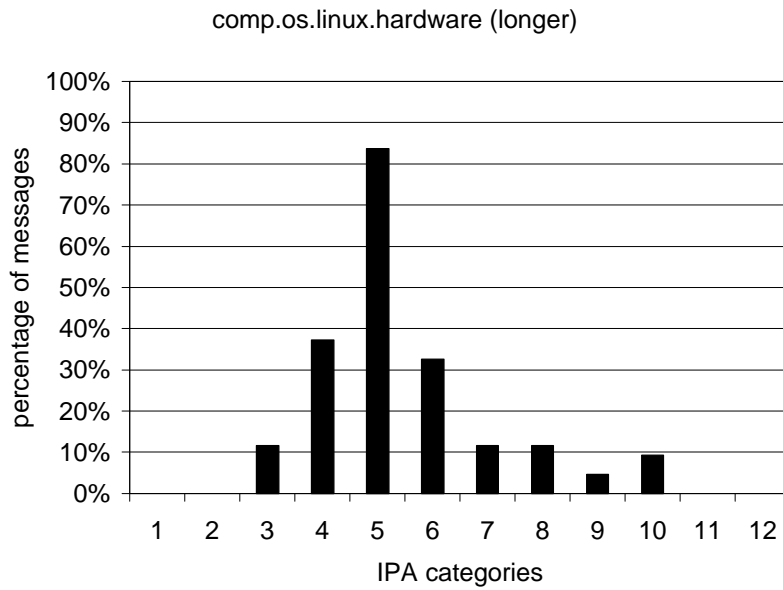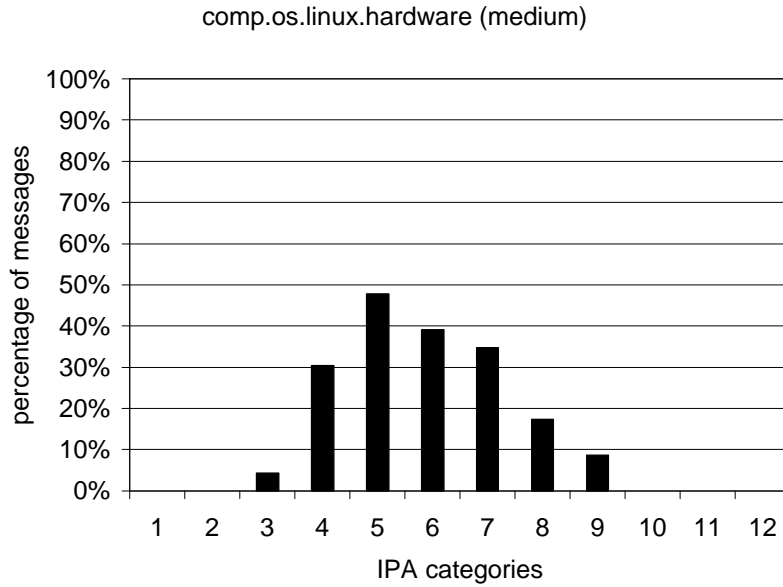

Figure 9-c. comp.os.linux.hardware "medium" discussion

comp.os.linux.hardware (short)
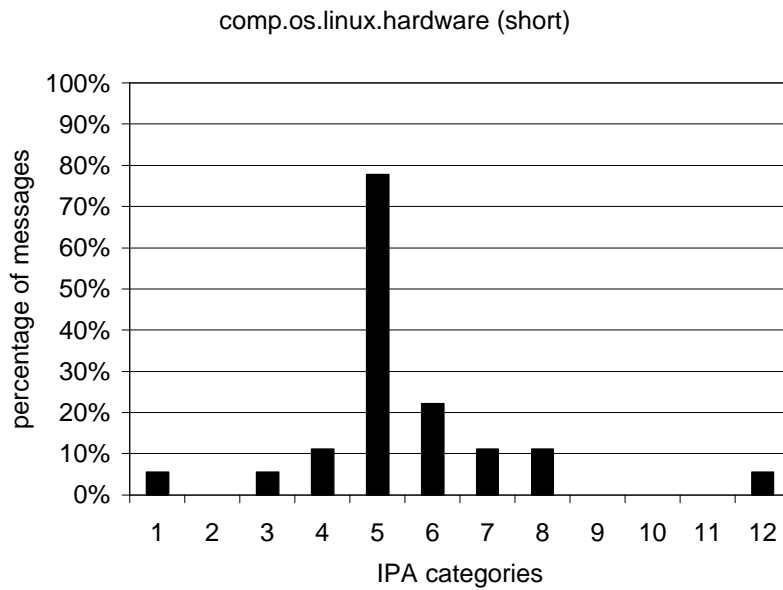
Figure 9-d. comp.os.linux.hardware "short" discussion

comp.os.linux.hardware (shorter)
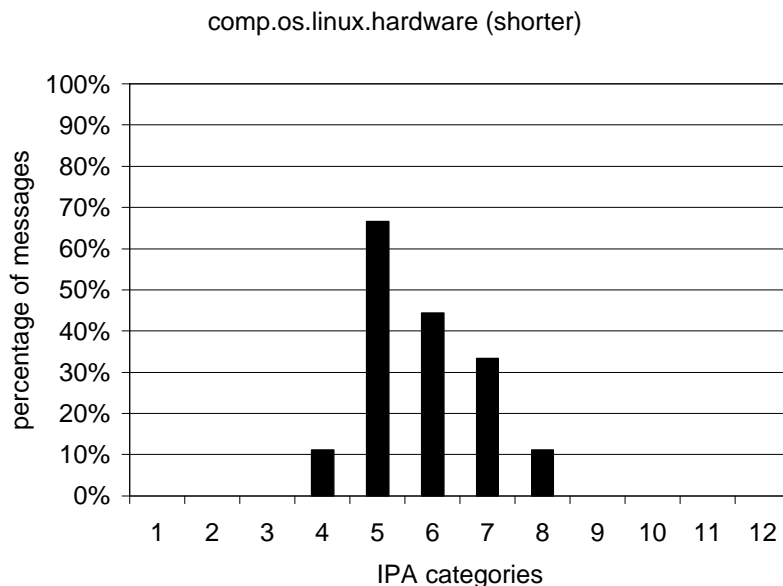
Figure 9-e. comp.os.linux.hardware "shorter" discussion

The "longer" discussion (Figure 9-a) consisted of 43 messages. The subject of this discussion was the relative merits of the Linux operating system. Category 5 ("Gives opinion") is again the behavior that occurs in the most messages. However, for the first time we see category 4 ("Gives suggestion") following rather than category 6 ("Gives orientation"). This is because the newsgroup member who began the discussion expressly asked what course of action they should take (9). It should also be noted that the extreme ends of the scale (1, 2, 11, and 12) were not represented in this discussion.

The "long" discussion (Figure 9-b) consisted of 24 messages. This discussion covered a technical matter. What is immediately noticeable about this discussion is the great number of occurrences of messages containing category 5 ("Gives opinion") behavior. Over 90% of the messages in this discussion offered opinions. The number of occurrences of messages displaying the other behaviors in the scale was dwarfed by messages containing opinions. There were no messages containing negative social-emotional behavior.

The "medium" discussion (Figure 9-c) consisted of 23 messages. The subject of this discussion was another technical matter. The messages revealed a fairly evenly distributed occurrence of behaviors in the task area, with category 5 ("Gives opinion") more in line with the other answers in the task area than usual. A small number of messages containing category 3 ("Agrees") behavior.

The "short" discussion (Figure 9-d) consisted of 18 messages. This discussion covered issues related to Y2K. Category 5 ("Gives opinion") was again the most prevalent behavior in the messages of the discussion. It is interesting to note the occurrences of extreme social-emotional behavior, which is uncommon in a discussion this short. One of the newsgroup members was very sarcastic (12) with another member in reaction to a statement the member made. As a result, a third newsgroup member came to the defense (1) of the member who was the subject of the sarcasm.

The "shorter" discussion (Figure 9-e) consisted of 9 messages. The subject of this discussion was a simple hardware issue. Because the discussion was so short, and the problem being discussed very straightforward, the behaviors exhibited in the messages were concentrated in the middle of the scale (the task area).

alt.os.linux Newsgroup
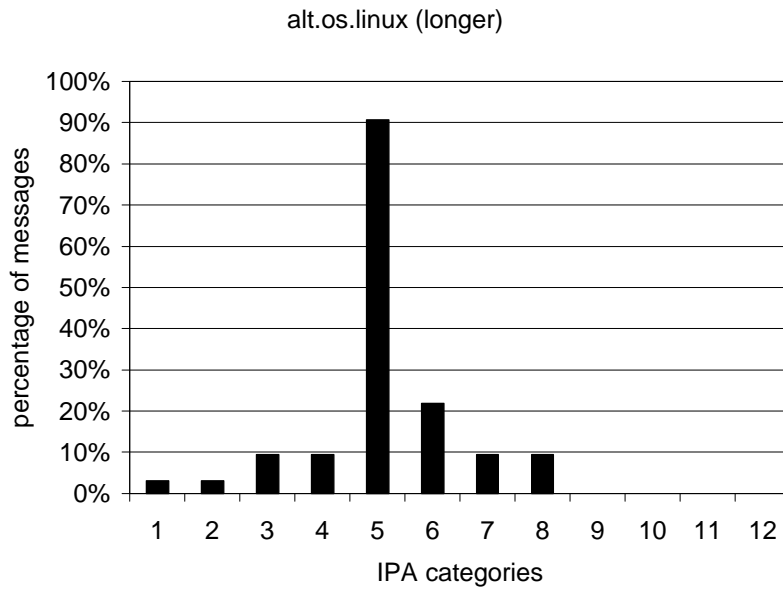
alt.os.linux (longer)

Figure 10-a. alt.os.linux "longer" discussion
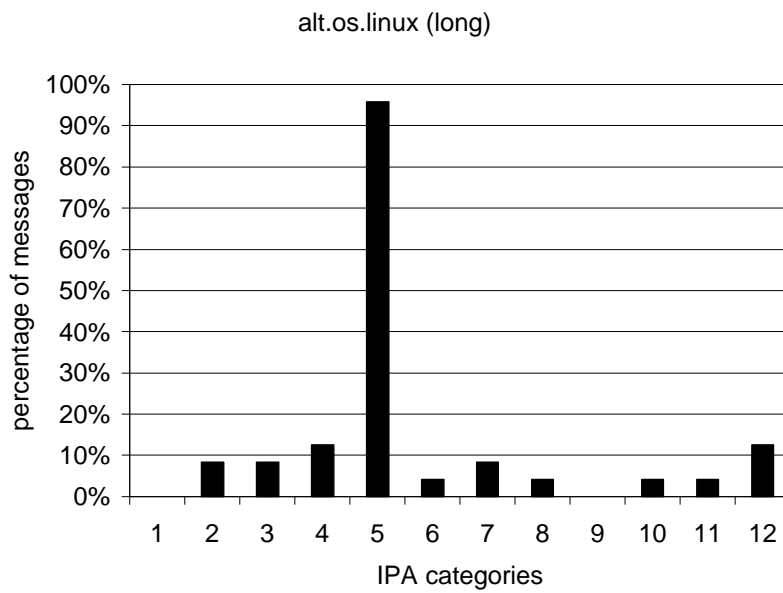
alt.os.linux (long)

Figure 10-b. alt.os.linux "long" discussion
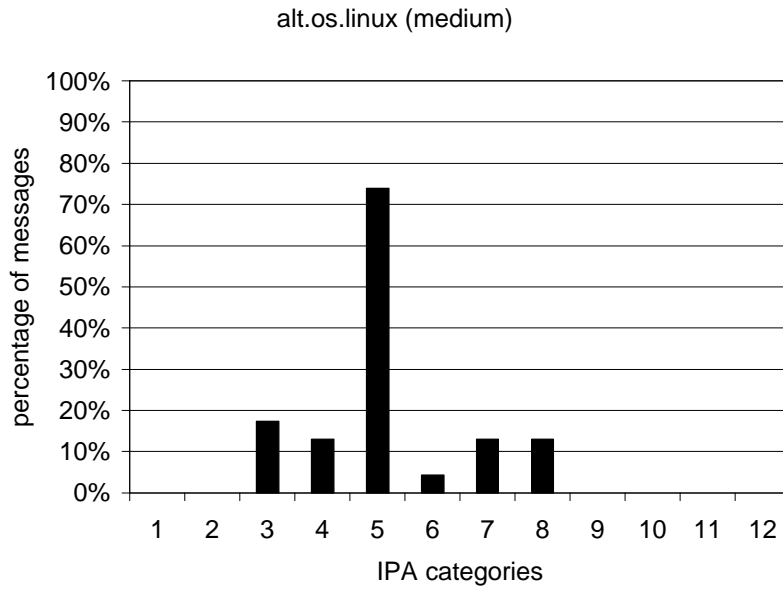
alt.os.linux (medium)

Figure 10-c. alt.os.linux "medium" discussion
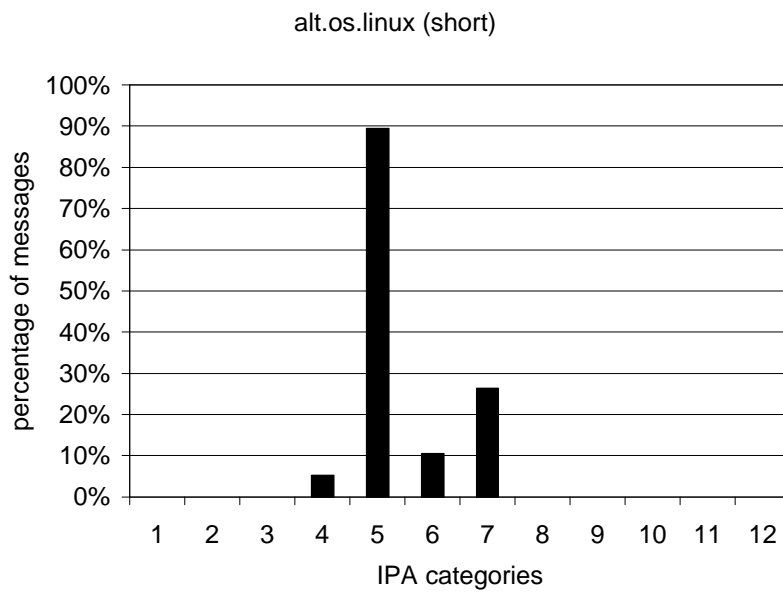
alt.os.linux (short)

Figure 10-d. alt.os.linux "short" discussion

alt.os.linux (shorter)
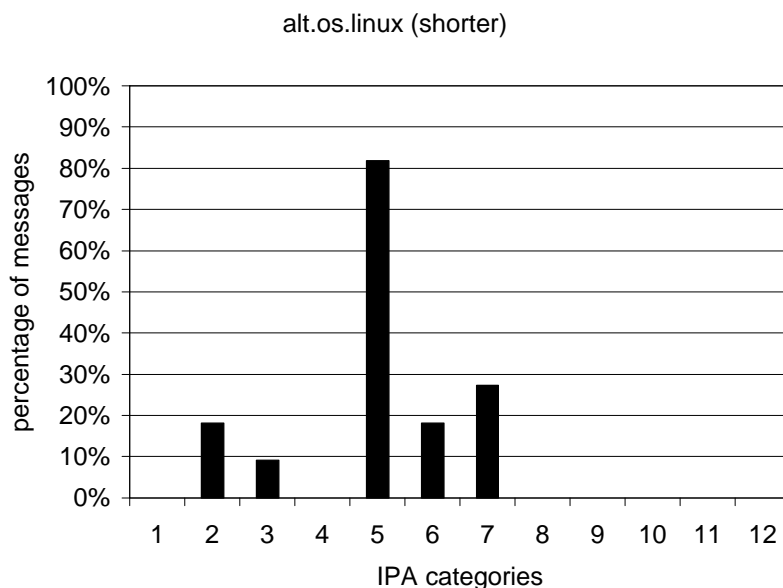
Figure 10-e. alt.os.linux "shorter" discussion

The "longer" discussion (Figure 10-a) consisted of 32 messages. The subject of this discussion was related to the relative merits of the Linux operating system. There were a high number of opinions (5) offered in this discussion, as this topic seems to draw them out. Not many questions were asked. The jokes (2) and agreement (3) on one side of the scale were balanced by the tension (11) and antagonism (12) at the other end.

The "long" discussion (Figure 10-b) consisted of 24 messages. This discussion covered the relative merits of the Linux operating system. What is immediately noticeable is the high number of messages containing opinions (5). This discussion contains the highest percentage (over 95%) of messages that express opinions of all the discussions in this study.

The "medium" discussion (Figure 10-c) consisted of 23 messages. The subject of this discussion was a technical matter. Messages were limited to members asking for information (7) and opinions (8), with other members offering information (6), opinions

(5), and suggestions (4). There were also some messages indicating compliance (3) with the suggestions that were given.

The "short" discussion (Figure 10-d) consisted of 19 messages. This discussion covered another technical matter, and thus the categories are limited to requests for information (7) along with the messages from members providing information (6), opinions (5), and suggestions (4). Messages containing opinions (5) were the most prevalent.

The "shorter" discussion (Figure 10-e) consisted of 11 messages. The subject of this discussion was a cartoon that was related to the Linux operating system. This explains the relatively high number of occurrences of category 2 ("jokes, laughs"). Category 5 is the prevalent behavior, but there are also a significant number of requests for orientation (7).

Overview of Newsgroups
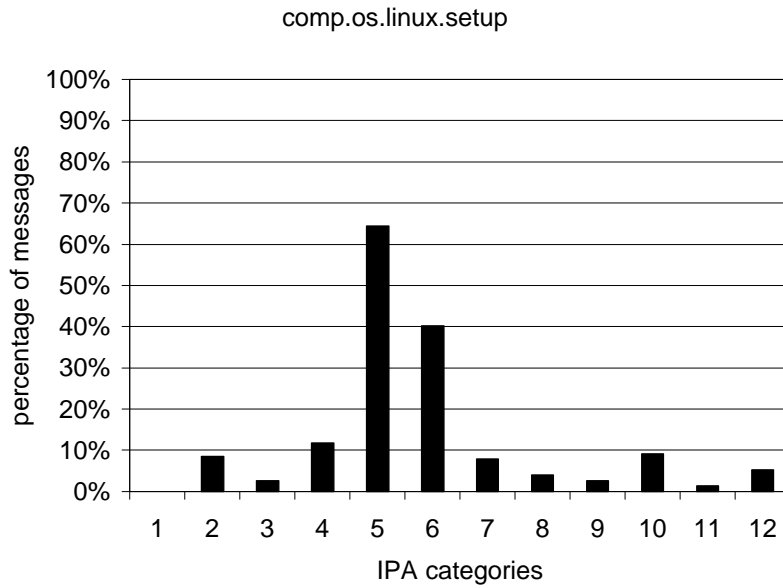
comp.os.linux.setup

Figure 11-a. All messages in comp.os.linux.setup

comp.os.linux.networking
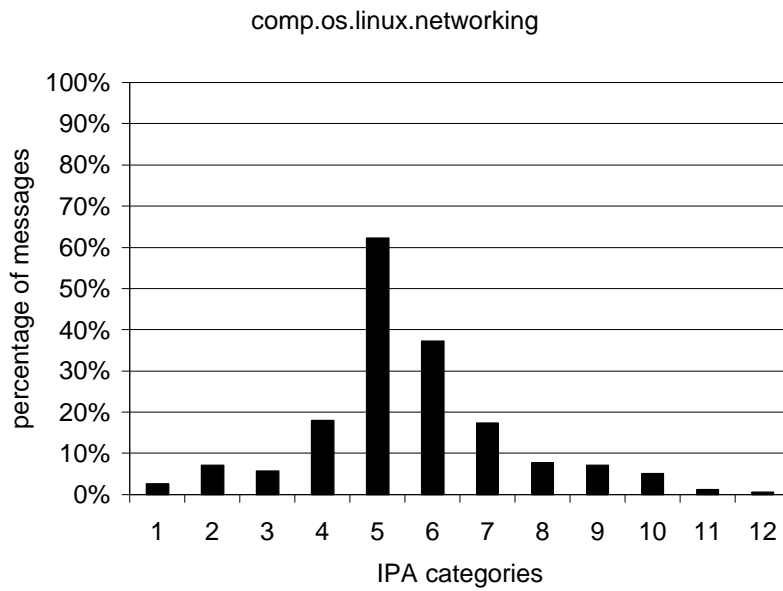
Figure 11-b. All messages in comp.os.linux.networking
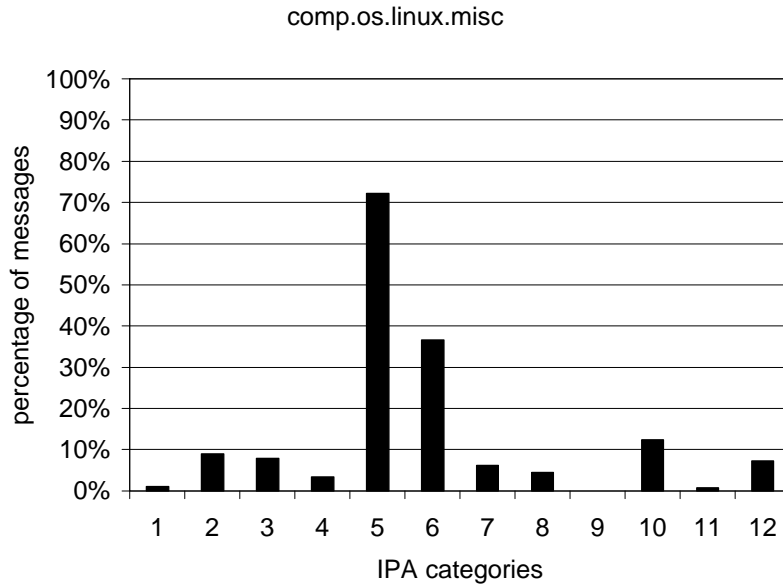
comp.os.linux.misc

Figure 11-c. All messages in comp.os.linux.misc

comp.os.linux.hardware
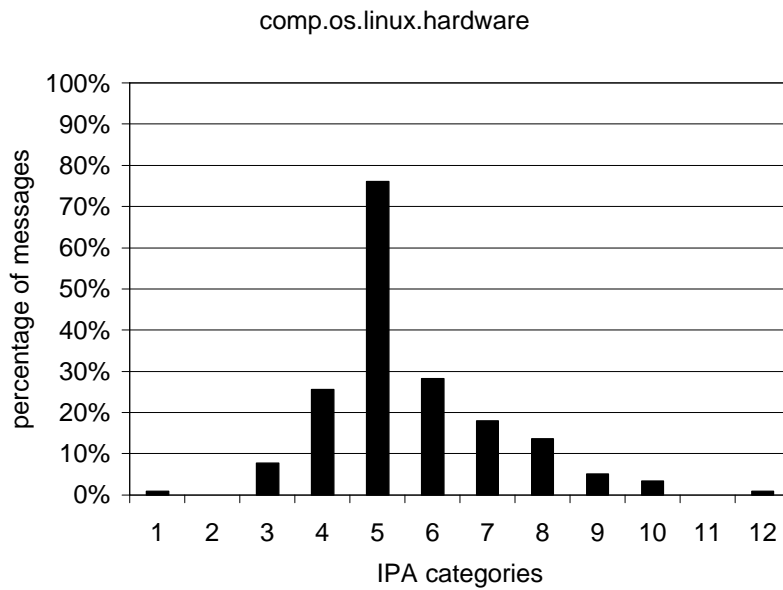
Figure 11-d. All messages in comp.os.linux.hardware
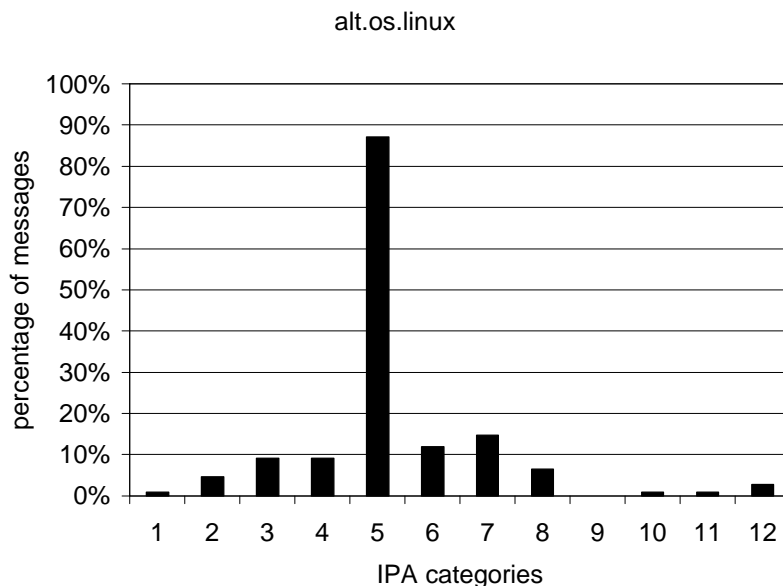
alt.os.linux

Figure 11-e. All messages in alt.os.linux

The five discussions in comp.os.linux.setup (Figure 11-a) consisted of 152

messages total. Categories 5 ("Gives opinion") and 6 ("Gives orientation") occur most

frequently in the messages, with the other categories dropping off significantly. There is

a complete lack of solidarity expression (1) in this newsgroup.

The five discussions in comp.os.linux.networking (Figure 11-b) consisted of 156

messages total. This graph shows a dramatic number of messages that contain category 5

behavior. The other categories do not even approach the number of messages that

category 5 occurs in. There are some members asking for orientation (7) and opinion (8),

but for the most part the group as a whole engages in the expression of opinions (5).

There is a fairly smooth distribution of categories in this newsgroup. Again, categories 5

and 6 have the highest occurrence, with categories 4 and 7 following at a roughly equal

level of occurrence.

The five discussions in comp.os.linux.misc (Figure 11-c) consisted of 292

messages total. This newsgroup is not as smoothly distributed. Categories 5 and 6 are

the most frequently occurring, with the other categories distributed fairly evenly, but not consistently. Of these, category 10 ("Disagrees") occurs in the most messages. This is an opinionated group, with a significant number of orientation answers (6) being given as well. There is a complete lack of questions seeking suggestions and direction (9).

The five discussions in comp.os.linux.hardware (Figure 11-d) consisted of 117 messages total. Again, this is a fairly smooth distribution of occurrences. This is an opinionated group, with a fair amount of orientation and suggestion being given as well. Category 5 ("Gives opinion") has by the highest number of occurrences, with categories 6 ("Gives orientation") and 4 ("Gives suggestion") sharing the second highest occurrence. There is a complete lack of tension release in this group (2), perhaps because of the complete lack of tension (11) and very low antagonism (12).

The five discussions in alt.os.linux (Figure 11-e) consisted of 109 messages total. This is the most opinionated group, with over 85% of the messages in the newsgroup containing opinions (5). The only other categories that occur in more than 10% of the messages are questions seeking orientation (7) and answers giving orientation (6).

Conclusions

After analyzing all of the discussions in all of the newsgroups, it is evident that the newsgroups in general exhibit a prevalence of category 5 behavior ("Gives opinion"), followed by category 6 behavior ("Gives orientation"). This is logical, as one question posted to the newsgroup will yield a number of answers from the other members of the newsgroup (and the largest newsgroup in this study has over 11,000 members).

There were some newsgroups that inspired and/or required extensive discussion, and others that were not as complicated and/or emotional. Shorter discussions and discussions about straightforward technical matters stay closer to center of the scale, with longer discussions and discussions about more controversial issues moving into the extremes (both positive and negative) of social-emotional behavior. Often the most controversial discussions were concerned with the merits of Linux versus the Microsoft Windows operating system.

Those discussions that contain a high number of messages with category 5 behavior ("Gives opinion") usually have associated with them more occurrences of negative social-emotional behaviors. It is possible that in some of the longer discussions, newsgroup members may find themselves becoming frustrated with the other opinions offered in the discussion, and seek to make themselves heard through stronger language. The shorter discussions do not seem to necessitate the expression of more extreme social-emotional behaviors from the newsgroup members.

There were 1350 total occurrences of behaviors within the 826 total messages analyzed. If we take this data and look at the four larger categories (positive social-emotional behaviors, attempted answers, questions, and negative social-emotional

behaviors) represented by A, B, C, and D on Bales' Interaction Process Analysis table, we find that most of the behaviors exhibited in the messages are *attempted answers* (70.96%), while the second most frequently occurring behaviors in the messages are *questions* (12.52%). Only 16.52% of the behaviors in the messages fall into the *social-emotional area* (8.82% are *positive reactions*, while 7.7% are *negative reactions*).

The lack of social-emotional interactions challenges standard definitions of groups and community. Many modern definitions assume social-emotional interaction among members. However, dictionary definitions of community include "a group of people with a common characteristic or interest living together within a larger society" (Mish, 1987, 267), "a body of persons of common and esp. professional interests scattered through a larger society" (ibid.) and "joint ownership or participation" (ibid.).

Based on our data these definitions apply to the Linux open source software community. It is *not* a community that in general shows solidarity, tension release (jokes, etc.), agreement, disagreement, or antagonism. It appears to be a community of initial and limited question asking followed by extensive answer giving.

The data suggest that the communication norms of the members of this collaborative community focus on communication concerning tasks rather than social-emotional behavior (83.48% of the behaviors exhibited in the messages were in the task areas, while, as mentioned above, only 16.52% of the behaviors in the messages fell into the social-emotional area). Should computer-mediated communication be used in any situation that requires the completion of a task, to promote efficiency and reduce social-emotional activity? Of course not. Face-to-face and social-emotional activity is an integral part of the workplace; it provides a sense of belonging and increased morale

among the members of the workplace community. However, within communities like the

Linux open source software community that are large, diverse, and widespread, on-line

communication through newsgroups is a necessity.

The data raise some other questions that should be investigated further. First,

how does the distribution of behavior in the task and social-emotional areas in face-to-

face groups compare to the distribution of behavior in the Linux on-line community?

The local Linux User Groups (LUGs), where Linux users meet face-to-face, would be a

good place to begin investigating this question. Second, is there any correlation between

the personalities of members of the Linux on-line community and their behavior in the

newsgroups (either in the task or social-emotional areas)? Third, if this forum fulfills the

task needs of the Linux community, and is not a forum for social-emotional behavior, is

there another forum where social-emotional needs are fulfilled? Again, the local Linux

User Groups would be worth investigating.

Appendix A

Bales' categories for Interaction Process Analysis

		1	*Shows solidarity*, raises other's status, gives help, reward
Social-emotional areas: Positive reactions	A	2	*Shows tension release*, jokes, laughs, shows satisfaction
		3	*Agrees*, shows passive acceptance, understands, concurs, complies
		4	*Gives suggestion*, direction implying autonomy for others
Task area: Attempted answers	B	5	*Gives opinion*, evaluation, analysis, expresses feeling, wish
		6	*Gives orientation*, information, repeats, clarifies, confirms
		7	*Asks for orientation*, information, repetition, confirmation
Task area: Questions	C	8	*Asks for opinion*, evaluation, analysis, expression of feeling
		9	*Asks for suggestion*, direction, possible ways of action
		10	*Disagrees*, shows passive rejection, formality, withholds help
Social-emotional area: Negative reactions	D	11	*Shows tension*, asks for help, withdraws out of field
		12	*Shows antagonism*, deflates other's status, defends or asserts self

Appendix B

Linux newsgroups and discussions used in this study (including number of messages)

	Disappointed with Linux and its performance (51)
	Linux Red Hat is JUNK (36)
comp.os.linux.setup	Swap partition: Solaris vs Linux (29)
	Can I use a 286 as an X-Terminal? (25)
	Small Fonts in Netscape (11)
	Computer games to play on a home network? (51)
	diff (if any) between hub & switch? (38)
comp.os.linux.networking	Slow network (29)
	3Com 3c905B and 3c90x driver troubles (24)
	Flaky BA PPPOE Linux behavior (14)
	Predicting the end of Java? (97)
	HTML Editor for Linux (88)
comp.os.linux.misc	New Internet DDoS Flood Attacks on Yahoo, etc. (58)
	Are you that person? (was: Do you hate vi?) (32)
	Netscape Blows Chunks! Tried Arena yet? (17)
	Help about Linux (43)
	VIA chipset (24)
comp.os.linux.hardware	Booting windows out of linux (23)
	Y2K problem (18)
	Athlon motherboard selection (9)
	Why Linux? (32)
	Alternative to Windows? (24)
alt.os.linux	Linux: How much memory (23)
	Partition Magic? (19)
	A Cartoon alt.os.linux might like (11)

Appendix C

Examples of Phrases used to Indicate categories in Bales' Interaction Process Analysis

		1	*Shows solidarity* "You're right," Thanks," "You've been very helpful"
Social-emotional areas: Positive reactions	A	2	*Shows tension release* use of smiling emoticons, friendly sarcasm, self-deprecation
		3	*Agrees* "I'll try it," I agree," "My mistake," "Good advice," "Valid point"
		4	*Gives suggestion* "My advice would be to," "I suggest," "You might want to"
Task area: Attempted answers	B	5	*Gives opinion* "I think," "I feel," "This is what's happening," "IMO," "IMHO"
		6	*Gives orientation* states facts, provides URL, name of book, name of software, name of hardware
		7	*Asks for orientation* "Does anyone know what," "Does anyone know where"
Task area: Questions	C	8	*Asks for opinion* "Does anyone know why," "What's going on?"
		9	*Asks for suggestion* "Does anyone know how," "What should I do?," "I'm desperate," "Help!"
		10	*Disagrees* "Rubbish," "Bull****," "You're wrong," "How can you say that?"
Social-emotional area: Negative reactions	D	11	*Shows tension* indirect sarcasm, passive aggressive behavior
		12	*Shows antagonism* direct sarcasm, "Why don't you go back to Windoze," name-calling, "RTFM"

Appendix D

Messages from Example of Coding Process

Jan 24, 04:47 PM
A
Message 1 of 9
Athlon motherboard selection
Hi, I'm trying to figure out which Athlon motherboard would be the best for my purposes. I'm looking at (in no particular order of preference):

Asus K7M
FIC SD-11
MSI 6167
Gigabyte GA-71X

I need something which will run fine with NT 4, Win98, and Linux. I run Debian on my other machines, but the only reason I need NT and 98 on this new box is because I need it for work (otherwise I would not bother).

However, the main concern here is that it be a good gaming box! ;) I intend to run Linux most of the time, and only use 98 for games that do not yet have linux ports.

Anyway they all look like great boards. I'm concerned however about the ATA-66 issue, I obviously need to be able to see my disks in Linux. If I can run the drives in ATA-33 mode that's fine, just as long as it works I don't really care.

The Asus is nice since it has 2 serial ports (some of the others have only 1 which bothers me) and on-board sound. However, I don't know if the mobo's audio would work in Linux, it says it's "AC97 compliant audio" and I have no idea what that means. Of course I could always put a regular sound card in there, but why pay for on-board audio if I'm not going to use it....

Also I've noticed that some (such as the Asus) use a VIA south bridge chipset. Is this better than the AMD south bridge?

Any performance/quality/compatibility differences between these boards? I've been using Asus Socket 7 boards for a long time now and I like them, but they are more expensive than the others and have that audio issue I mentioned.

TIA for any advice!!

Jan 25, 01:38 AM
B (in response to A)
Message 2 of 9
Re: Athlon motherboard selection
A wrote:
> *Asus K7M*

Thats mine at home and at work since two weeks.

> *I need something which will run fine with NT 4, Win98, and Linux. I run Debian on my other*
> *machines, but the only reason I need NT and 98 on this new box is because I need it for work*
> *(otherwise I would not bother).*

At work i´m using NT4.0 (sigh) and Linux (much better) and at home only natively Linux and Windows under Vmware and everything works fine.

My machines are using:

ASUS K7M
Athlon 600 (overclocked to ~640 MHz)
Voodoo3 2000, MGA200, Elsa Trio (S3…)
Adaptec 2940UW, or UDMA66 EIDE

> *However, the main concern here is that it be a good gaming box! ;) I intend to run Linux most of the time, and only use 98 for games that do not yet have linux ports.*

Nice performance but depends for sure on 3DNow or not : -)

> *Anyway they all look like great boards. I'm concerned however about the ATA-66 issue, I obviously need to be able to see my disks in Linux. If I can run the drives in ATA-33 mode that's fine, just as long as it works I don't really care.*

Since my one and only IDE-Drive wasn´t listed in the past tests i can´t tell you much about this issue, but compared with my SCSI-devices the performance is not bad.

> *The Asus is nice since it has 2 serial ports (some of the others have only 1 which bothers me) and on-board sound. However, I don't know if the mobo's audio would work in Linux, it says it's "AC97 compliant audio" and I have no idea what that means. Of course I could always put a regular sound card in there, but why pay for on-board audio if I'm not going to use it....*

The on-board-audio can emulate a SoudBlaster which works under Linux. I havn´t tested this myself but somebody else uses it that way.

cu

B

Jan 25, 07:13 AM
C (in response to B)
Message 3 of 9
Re: Athlon motherboard selection
Hi !

> *My machines are using:*
> *ASUS K7M*
> *Athlon 600 (overclocked to ~640 MHz)*
> *Voodoo3 2000, MGA200, Elsa Trio (S3...)*
> *Adaptec 2940UW, or UDMA66 EIDE*

How did you overclock you Athlon and, did you try higher speeds too ?

Regards,
C

Jan 26, 02:15 AM
B (in response to C)
Message 4 of 9
Re: Athlon motherboard selection
C wrote:

> *ASUS K7M*
> *Athlon 600 (overclocked to ~640 MHz)*

Voodoo3 2000, MGA200, Elsa Trio (S3...)
Adaptec 2940UW, or UDMA66 EIDE

How did you overclock you Athlon and, did you try higher speeds too ?

The ASUS K7M allows you to change the FSB-Clock from 100MHz to 150MHz. So if 100MHz-FSB-Clock leads to 600MHz processor-speed you can (in theoryfor sure) speedup to 900MHz. My Athlon at home works stable with 106MHz FSB which leads to 636MHz. 107MHz is stable enough for playing around with windows, but i need stability for LINUX.

I have tested five athlon-600 machines and i got tuning results from 102MHz to 112MHz. Another possibility is to use H.Oda´s WCPUA2. It changes the clock-ration of the processors internal cache and may tune up somewhat more. For this tool look at http://www.h-oda.com

cu

B

Jan 27, 04:24 PM
D (in response to A)
Message 5 of 9
Re: Athlon motherboard selection
A writes:
  Asus K7M

Tom's Hardware Guide likes this one. Mostly because it allows you to overclock the CPU. The only disadvantage I can think of is that it's not on AMD's list of "preferred" manufacturers.

  FIC SD-11
  MSI 6167
  Gigabyte GA-71X

These 3 are all on AMD's list, so they will be compatible.

Personally, I'd avoid the FIC board. Partly because their reputation for quality is not uniform (they have some good boards and some bad ones.) Also because they're too cheap to provide a second serial port - with a motherboard that costs over $150, you'd think they could throw in a lousy 50-cent connector.

I haven't personally used any of the four, though, so you may prefer to hear from others here.

Also read the Athlon articles at Tom's Hardware. Many benchmarks and stuff were run there.

  Anyway they all look like great boards. I'm concerned however about the ATA-66 issue, I
  obviously need to be able to see my disks in Linux. If I can run the drives in ATA-33 mode
  that's fine, just as long as it works I don't really care.

I would hope you can run the drives in ATA-33 mode. If you're not sure, all the manufacturers have web sites where you can download the manuals. They should tell you.

  The Asus is nice since it has 2 serial ports (some of the others have only 1 which bothers me)
  and on-board sound. However, I don't know if the mobo's audio would work in Linux, it says

it's "AC97 compliant audio" and I have no idea what that means. Of course I could always put a regular sound card in there, but why pay for on-board audio if I'm not going to use it....

I think the audio is SB-16 compatible. I don't know if any of the higher-powered features have Linux support, though.

Again, I recommend you download the manual. It should say.

-- D

Jan 31, 01:14 PM
E (in response to D)
Message 6 of 9
Re: Athlon motherboard selection
D writes:
> *A writes:*
>
>> *Anyway they all look like great boards. I'm concerned however about the ATA-66 issue, I obviously need to be able to see my disks in Linux. If I can run the drives in ATA-33 mode that's fine, just as long as it works I don't really care.*
>
> *I would hope you can run the drives in ATA-33 mode. If you're not sure, all the manufacturers have web sites where you can download the manuals. They should tell you.*

Stock linux 2.2 kernels (like the one on my K7M/600 machine, 2.2.13) will not recognize the hardware of the IDE controller that's on the asus k7m, but neither will it complain about it. It'll just quietly use a compatibility mode that is much slower than even UDMA/33, and since there's no errors, there's no way to know without checking hdparm manually. I get 4MB/second typically. Pitiful, ain't it?

So it's usable by linux. Just don't expect storage i/o performance.

There is a group (or maybe just one person) working on a patch to Linux that enables full udma/66 in theory on the VIA chipset that the k7m utilizes, but I've heard no stories about it, either success or failure, so I'm personally holding off on trying it out until it's incorporated into the linux 2.3 development tree. If you're daring and keep your data religiously backed up (I don't, hense why I'm not trying :) you can find it at:

ftp://ftp.kernel.org/pub/linux/kernel/people/hedrick/

-- E

Jan 31, 01:36 PM
F (in response to E)
Message 7 of 9
Re: Athlon motherboard selection

E wrote in message news:
> *Stock linux 2.2 kernels (like the one on my K7M/600 machine, 2.2.13) will not recognize the hardware of the IDE controller that's on the asus k7m, but neither will it complain about it. It'll just quietly use a compatibility mode that is much slower than even UDMA/33, and since there's no errors, there's no way to know without checking hdparm manually. I get 4MB/second typically. Pitiful, ain't it?*

I am using a 2.2.14 kernel configured to use the experimental VIA UDMA/66 driver and after putting an "hdparm -c1 -d1 -X66 /dev/hda" into my startup get a solid

~24MB/second according to "hdparm -t /dev/hda". Hardware is an Asus K7M and a fujitsu UDMA66 drive. Cheers,

F.

Jan 31, 09:59 PM
G (in response to D)
Message 8 of 9
Re: Athlon motherboard selection
What is Tom's hardware guide?

G

Feb 01, 08:41 AM
D (in response to G)
Message 9 of 9
Re: Athlon motherboard selection
G writes:
   What is Tom's hardware guide?

http://www.tomshardware.com

The site is full of reviews and tests of hardware components. The site is one of the more respected hardware-review sites.

-- D

References

Bales, R.F. (1950). A Set of Categories for the Analysis of Small Group Interaction. *American Sociology Review, 15*, 257-263.

Deja.com. [On-Line]. Available: http://www/deja.com/

Dempsey, B., Weiss, D., Jones, P., & Greenberg, J. (1999). A Quantitative Profile of a Community of Open Source Linux Developers. SILS Technical Report TR-1999-05.

Garton, L., Haythornthwaite, C., & Wellman, B. (1999). Studying On-Line Social Networks. In S. Jones (Ed.), *Doing Internet Research: Critical Issues and Methods for Examining the Net* (pp. 75-105). Thousand Oaks, CA: SAGE Publications, Inc.

Hiltz, S.R. & Turoff, M. (1993). The Network Nation: Human Communication via Computer. Cambridge, MA: The MIT Press.

Hiltz, S.R., Johnson, K., & Turoff, M. (1986). Experiments in Group Decision Making: Communication Process and Outcome in Face-to-Face Versus Computerized Conferences. *Human Communication Research, 13* (2), 225-252.

Leonard, Andrew (1998, June 26). User friendly? *salon.com*. Retrieved October 19, 1999 from the World Wide Web: http://www.salonmagazine.com/21st/reviews/1998/06/26reviews.html

Mish, Frederick C. (Ed.). (1987). *Webster's Ninth New Collegiate Dictionary*. Springfield, MA: Merriam-Webster, Inc.

Moody, Glyn (1997). The Greatest OS That (N)ever Was. *Wired*, *5.08*. Retrieved October 19, 1999 from the World Wide Web: http://www.wired.com/wired/5.08/linux_pr.html

Raymond, Eric (No date). The Cathedral and the Bazaar. Retrieved October 19, 1999 from the World Wide Web: http://www.tuxedo.org/~esr/writings/cathedral-bazaar/

RemarQ. [On-Line]. Available: http://www.remarq.com/

Rice, Ronald E. (1994). Network Analysis and Computer-Mediated Communication Systems. In S. Wasserman & J. Galaskiewicz (Eds.), *Advances in Social Network Analysis: Research in the Social and Behavioral Sciences* (pp. 167-203). Thousand Oaks, CA: SAGE Publications, Inc.