

An Examination of Remote Access Help Desk Cases

Aaron Steinfeld Ritika Sanghi Joseph Giampapa

Daniel Siewiorek Katia Sycara

September 2003
CMU-CS-03-190

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

This report also appears as
Human Computer Interaction Institute
Technical Report CMU-HCII-03-100

Abstract

As a precursor to explorations on future network interoperability problem resolution methods and tools, it is necessary to obtain an understanding of problems in the present day. The remote network access application area was chosen as a case study due to rich sources of information, frequent problems, and considerable detrimental impact on user efficiency. To this end, existing remote network access help desk data was acquired and analyzed. The data was used to characterize remote network access interoperability problems and identify key issues. For the data examined, the two largest problems specific to remote end users were obtaining modem phone numbers for their location and adequate user rights upon connection. Potential for better knowledge re-use and dissemination of solutions to common problems to the general population was also observed.

This work was completed under a grant from the Office of Naval Research, Interoperability of Future Information Systems through Context-and Model-based Adaptation (#N00014-02-1-0499). The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the Office of Naval Research or the U.S. Government unless so designated by other authorized documents.

Keywords: human computer interaction, help desk, system administration, problem resolution, remote access, mental models, network interoperability

Introduction

Resolving network interoperability problems is difficult and time consuming. Almost every user has experienced such a problem either directly, or as a by-product of a task they were attempting to complete. Problems may originate or be complicated by system heterogeneity, administrative policies, security practices, and end user errors or improper mental models.

Advances in network flexibility, self-repair, and reconfiguration will improve underlying performance (Meseguer, et al, 2003) but lead to increased complexity. Furthermore, it is likely that the user will be unable to rely on a consistent detailed mental model of network state and topology. As such, new human-computer interaction methods and tools will be required to enhance user awareness and problem resolution.

As a precursor to explorations on these methods and tools, it is necessary to obtain an understanding of network interoperability problem resolution in the present day. The remote network access application area was chosen for this project due to rich sources of information, frequent problems, and considerable detrimental impact on user efficiency.

To this end, existing remote network access help desk data was acquired and analyzed. The data was used to characterize remote network access interoperability problems and identify key issues.

Data Collection

The experimenters obtained remote access trouble ticket data from the School of Computer Science (SCS) Computing Facilities office. The data consisted of trouble ticket remote access case data from June 5th, 2000 through January 15th, 2003 (about 2.5 years). There were 528 cases in this sample. Due to the presence of certain special characters in certain database fields, some manual cleaning was required for certain cases prior to data coding and analysis. The analysis of incident reports is a common and accepted data collection methodology (Wickens, 1995; Salvendy & Carayon, 1997). It is especially useful when attempting to identify real-world problems across wide, diverse, populations.

Coding Method

The coding was done manually by multiple experimenters with a final pass by a single experimenter to ensure technique, syntax, and word choice conformity. Prior to full scale coding, a common batch of cases was coded by the different experimenters and examined for inconsistencies. This led to refinement of the coding process, syntax, and term usage.

Regular Code Entries

The bulk of the cases were summarized using a standard format and location description. This procedure utilized the following technique:

Problem (a; b; c)

(a) Which machine has problems;

[static, dynamic, single, network, entry, core]

Where:

Static	A machine that rarely changes configuration
Dynamic	A machine that changes frequently (e.g., laptop)
Single	Single event changes (e.g., relocating a desktop)
Network	A non-CMU problem (e.g., external ISP)
Entry	A problem with the SCS network entry point (e.g., error in modem pool)
Core	A problem with the core SCS network that has manifested as a remote problem (e.g., authentication server failure)

(b) Symptom;

(c) Root cause OR “unknown”

Solution (d)

(d) Summary of fix w/ tool name and new setting if possible OR “unknown”

Symptoms, root causes, and solutions were kept as terse as possible. Descriptions were made anonymous (e.g., omission of user name, IP number, etc) to prevent user identification. When needed, commas were used to separate details within fields. The pipe character (|) was used to separate multiple problems and solutions that occurred within one case. Every problem had a partner solution and pair order was synchronized for cases that had multiple pairs. These practices were put in place to provide greater ease and accuracy for subsequent keyword searching and summarization of subsets. For example, one case from the data was encoded as follows:

Case narrative

“User is having problems dialing up to this number: [local modem phone number]. He couldn't establish connection everytime [sic] he dials in. He has been having this problem 6-7 times since last week. One modem was hung yesterday and another modem was hung today. All modems are answering now.”

Problem

entry; no connections on local dial-up; modem down

Solution

reset modems in pool

Quick Code Entries

Certain frequent and easily resolved problems were classified using predefined codes in the problem field only. Similar codes were identified for non-problem cases. The codes used were:

int	Requests for international phone numbers, related financial questions
800	requests for outside Pittsburgh phone numbers, related financial questions

local	requests for the local Pittsburgh numbers
sales	sales pitches
suggest	suggestions that are not due to problems, messages sent out to users
junk	cases that appear to have nothing in them
help	coder does not understand content or wants another perspective
<case #>	case number for previous entry where this row is a follow-up

Operating System

During the manual encoding process, the experimenters examined the case description and the e-mail headers in an attempt to identify the operating system in question. Five bins were used:

Win	Any variant of Windows
Mac	Mac OS 9 or X
Li/unix (no Mac)	Any variant of *nix except Mac OS X
Unknown	Any indeterminable operating system
Mixed	Problems that involved multiple operating systems (e.g., dual boot machine, home LAN, etc)

No separation was made between Macintosh OS 9 and X due to difficulty in identifying which operating system was being used. Furthermore, it was felt that the bulk of the Mac OS X users would have experience and mental models of system behavior that matched Mac OS 9 rather than the Li/unix set of users (e.g., they would not be well versed in discussing and debugging low level system functions).

The Unknown field was used when the operating system was not reported or the mail client description in the e-mail headers was not sufficient for rapid identification. Besides Pine and mh (the user may have logged into a *nix server and run a mail client remotely), this typically occurred when the build description of Outlook or Eudora did not specify an operating system. Matching specific build numbers to platform was deemed an inefficient use of encoding time. However, the data were retained should a more detailed examination be needed.

Results

Overview

After extracting the non-help cases (i.e., sales, suggest, junk, and case extensions) there were 414 “help only” cases. Of these, 88 were for the Single classification and 137 were phone number requests. After an initial analysis, there was concern that the Static, Dynamic, Entry, and Core bins were too fragmented so a simpler classification scheme was developed where Static and Dynamic were merged into Leaf and Entry was merged into Core. This is illustrated in Figure 1.

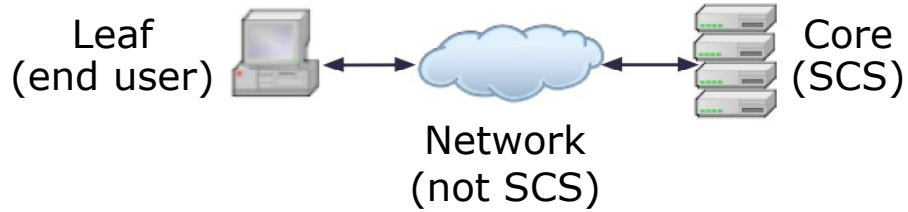


Figure 1. Refined location terminology

Case Load

During the period in question there was a phase out of Carnegie Mellon DSL service, an introduction of a VPN option, and a beta test of a licensed dial-up ISP for traveling users. The latter two may explain the slight rise for incoming cases near the end of the sampled period, while the DSL phase out may explain the small hitch in late 2000. However, it was interesting to note that, in general, the rate of incoming cases was remarkably linear (Figure 2). This suggests that little progress was made over this period in developing methods of reducing caseload.

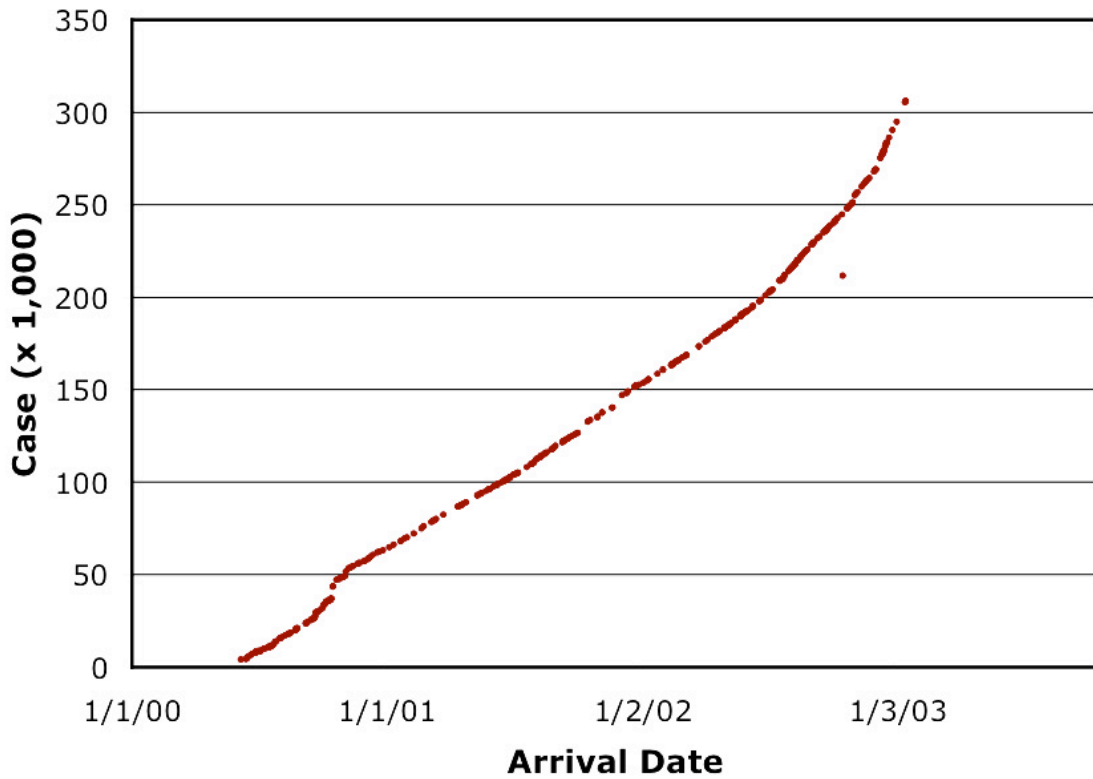


Figure 2. Case arrival rate

The handful of cases without recorded dates and the outlier near early 2003 do not have readily apparent explanations.

Analysis Set 1

The first pass of analyses involved limited filtering of the help only cases (414). Twelve cases with zero or null Hours to Resolve and 4 cases with more than 1,000 Hours to Resolve were removed. The elimination of cases over 1,000 hours was due to the fact that some cases were left in the queue as reminders. Zero or null hours cases were removed after inspection, as they appeared to be either incorrectly entered cases or glitches in the data collection process. **It is important to note that Hours to Resolve is not a good measure of staff time consumed. This is simply the time from the initial user query to the time the case was closed.** The subsequent 398 cases were analyzed on the characteristic of Problem Type.

Problem Type

The most salient observation upon looking at Problem Type was the high caseload and time sink resulting from phone number queries (Table 1). Also apparent was the high mean time to resolve problems stemming from third party networks. Problems due to single configuration change events were frequent (22%) and likely the result of shifting policies and network options (i.e., DSL phase-out, VPN roll out, and licensed dial-up ISP beta test).

Table 1. Problem type statistics

	N	Hours to Resolve		
		Mean	Std Dev	Sum
Core	69	58	111	4,013
Network	45	77	132	3,447
Leaf	66	60	133	3,957
Single	86	52	104	4,490
Phone Number	132	27	86	3,523
Overall	398	49	110	19,431

Over 25% of the cases and 3,500 hours were requests related to phone number requests. This may have been partly affected by a cost control policy to not widely publish international phone numbers. The high hours count was likely skewed by long turn around times on requests for phone numbers in foreign countries. Some of these were influenced by time zone effects (e.g., user is in Asia and staff is only fielding queries during the work day) and delays when requesting new numbers from the licensed ISP. This same factor may have increased the mean for Network as problems with ISP modem pools may have also arrived after hours. An additional delay may have also been introduced during communications between system administrators at SCS and the ISP.

A quick examination of a graph similar to Figure 2, but for only Phone Number cases, showed no apparent reduction in case frequency after introduction of a licensed dial-up ISP beta test. This was checked since the ISP client software included all available phone numbers. It is possible that frequency of Phone Number cases could drop should this ISP become the standard dial-up connection method when traveling.

Analysis Set 2

The second pass of analyses utilized the Analysis Set 1 (398) with the modem phone number requests (132) removed. The 266 cases were examined for Problem Type, Operating System, Connection Mode, Security Policies, and Knowledge Re-use.

Problem Type

It is readily apparent, and somewhat heartening, that a large number of cases were resolved within the first day (Figure 3). Inspection of the pace of problem resolution shows that Leaf cases linger on much longer than other types. This pattern, and the similar one for Network cases, is likely due to delays incurred during diagnosis due to iterations on communication of symptoms and state. Improper user mental models can lead to clarifying questions and requests for remote diagnosis (e.g., “What is entered in the username field?”). Problems for Core cases are diagnosed by staff without iterative communication with end users. The tail for Single cases may be due to waiting periods for new application versions (e.g., VPN client, ISP client, etc).

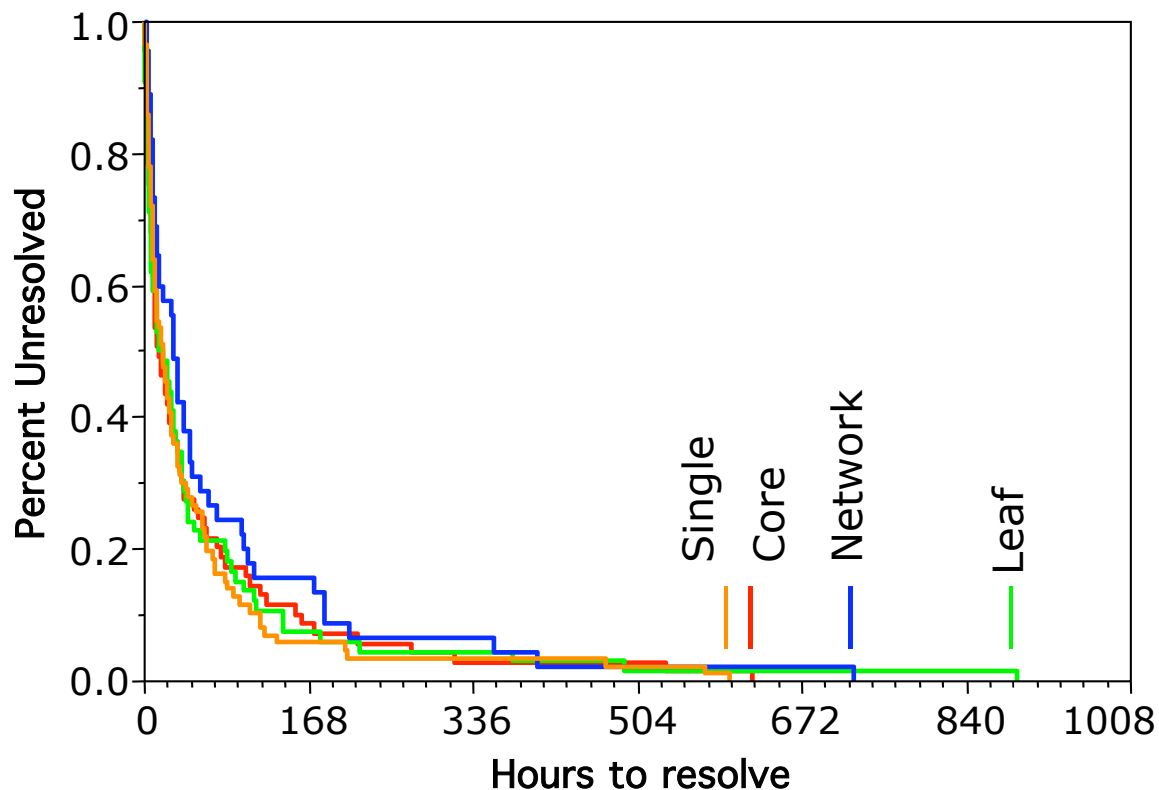


Figure 3. Resolution by type

Operating System

Comparison of case sample totals for the Win, Mac, and Li/unix categories (Table 2) to the population totals within the SCS user base is not possible as the latter statistics are not easily documented. While many computers are registered, SCS only documents computers owned by Carnegie Mellon and privately owned computers that utilize direct connections to the campus network (e.g., laptops brought to campus). Privately owned computers at the user’s residence are

not usually documented. Furthermore, SCS does not have a clear picture of which machines are utilized for remote access.

Table 2. Operating system statistics

	N	Hours to Resolve		
		Mean	Std Dev	Sum
Win	92	61	118	5,592
Mac	22	60	68	1,322
Li/unix (no Mac)	23	61	99	1,392
Unknown	121	54	108	6,493
Mixed	8	139	304	1,109
Overall	266	60	118	15,908

Note that Mixed category results should be viewed with caution due to the low sample size. As previously mentioned, some of the cases in the Unknown category could likely be shifted into the Win and Mac categories by examining individual mail client build numbers, but this was deemed to be an inefficient use of experimenter time. What is interesting is that the Unknown duration curve is very similar to the Win curve in Figure 4. These curves and the Li/unix curve exhibit the typical decay pattern expected in help desk operations.

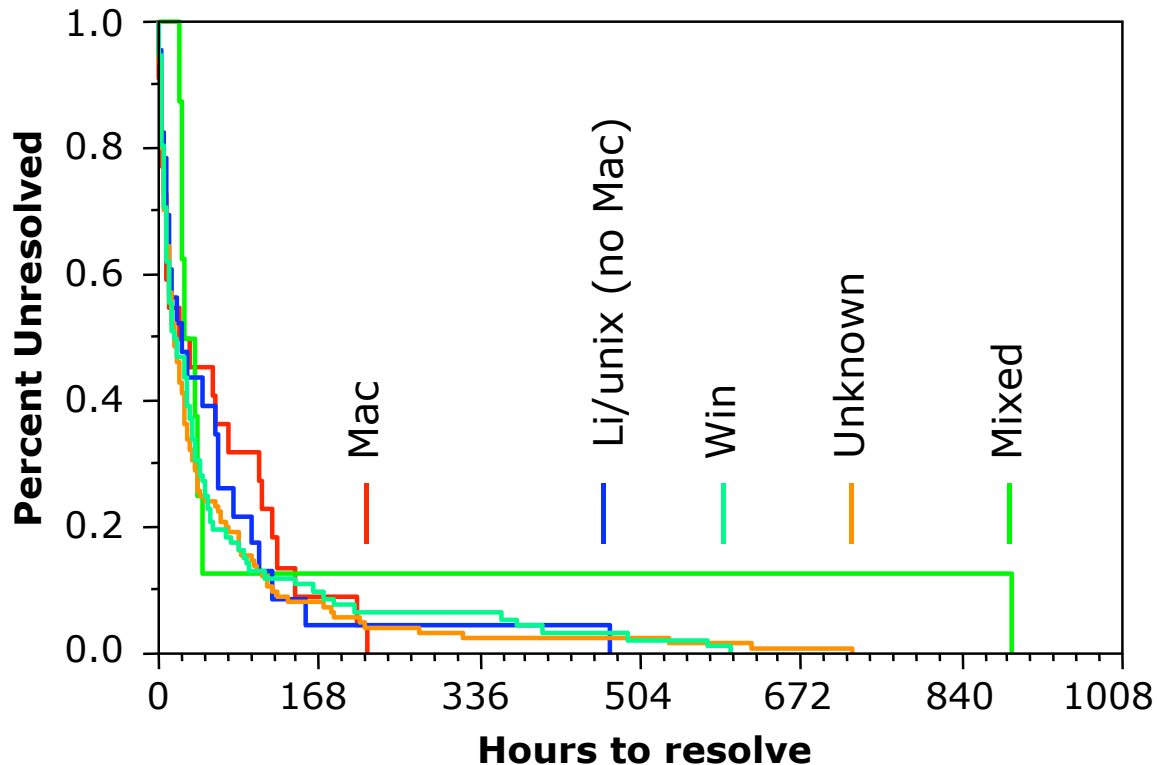


Figure 4. Resolution by operating system

Perhaps the most interesting observation from Figure 4 is the almost linear dive to full resolution by Mac users shortly after the first week. This manifested as considerably reduced variability for the Mac category when compared to the other categories. There is little explanation for this pattern and inquiries to staff offered no additional insights. Potential factors include greater ease of use of Mac systems, greater ease at identifying of system state, potentially greater use of peer support networks, and the possibility that Mac users abandon the help desk faster than their peers (e.g., lack of confidence in the help desk finding a solution, resignation that the problem may never be solved, etc). One observer joked “Mac users give up quicker” – which is a conceivable summary of the latter hypothesis.

The choppy curve depicted by the Mixed category is likely due to the considerably low number of samples. With more cases, the Mac, Li/unix, and Mixed categories may display smoother curves.

Connection Mode

Three types of remote connection mode were examined: Modem, Wireless, and DSL. These bins are mixed in that they document both third party ISP and Carnegie Mellon connection services. Carnegie Mellon provides options for the former two and phased out Carnegie Mellon DSL during the data period. One of the SCS licensed ISPs provides wireless nodes and two ISPs provide modem service. Users were directed to migrate to third party DSL services prior to the Carnegie Mellon phase out.

Approximately one quarter (26%) of the Analysis Set 2 cases logged were not associated with specific connection modes (Table 3). The bulk of these (43) were cases involving the VPN (e.g., installation, security policies requiring use).

The largest bin was problems associated with Modem connections (45%). Of these, 40% (48) were due to Core problems, the rest were spread across the other three types (Single, Network, and Leaf). It was not uncommon for multiple end users to notify the help desk when modem pool errors occurred.

Table 3. Connection mode statistics

	N	Hours to Resolve		
		Mean	Std Dev	Sum
Other	70	73	124	5,134
Modem	119	54	103	6,482
Wireless	26	50	121	1,312
DSL	43	44	82	1,890
DSL, Modem	5	29	47	147
DSL, Wireless	3	314	500	943
Overall	266	60	118	15,908

The smallest bin, and also the highest mean, was for cases that involved both DSL and Wireless connection methods. The bins with combined connection modes were usually requests to have

the same IP number in both modes (e.g., laptops). This was a more realistic request prior to phase out of Carnegie Mellon DSL.

Security

The high rate of problems associated with the VPN implies many cases were specifically due to problems originating from the use or application of security policies. Two security categories were examined: VPN and Realm. VPN cases included problems installing, configuring, and using the VPN. Cases associated with Realm included conflicts with security policies (e.g., mail relaying), authentication (e.g., password errors), and network card registration. Security problems were common – 41% of the cases and 47% of the time involved VPN and/or Realm (Table 4). VPN cases seemed to be particularly time consuming but this may be due to users waiting for new VPN client distributions. Windows deployment was more rapid than Mac and *nix.

Table 4. Security statistics

	N	Hours to Resolve		
		Mean	Std Dev	Sum
None	158	53	112	8,356
Realm	54	53	108	2,855
VPN	42	104	156	4,375
VPN, Realm	12	27	29	322
Overall	266	60	118	15,908

Almost 60% (94) of the None cases (those not involving VPN or Realm) were associated with Modem connections. An additional bias was observed when splitting Security between end users (Leaf, Single) and system administrators (Core, Network). Table 5 shows that the bulk of end user problems were due to insufficient user rights (55%) while system administrator problems were, for the most part, not related to security problems. Also interesting was that security problems for both classes of operators tended to consume more time than cases not involving security.

Table 5. Users and security

		N	Hours to Resolve		
			Mean	Std Dev	Sum
Leaf, Single	None	69	49	119	3,399
Leaf, Single	VPN, Realm	83	61	117	5,048
Core, Network	None	89	56	107	4,957
Core, Network	VPN, Realm	25	100	154	2,503
Overall		266	60	118	15,908

Knowledge Re-use

During the manual coding process it was observed that the Root Cause and Solution Summary fields in the database were rarely filled. As mentioned in the description of the coding process, experimenters documented Root Cause and Solution using whatever material was available

(usually the narrative). Through this process it became apparent that knowledge re-use was not possible for most problems since only 38% of the cases could be fully documented (Table 6). The lack of detailed solution documentation is consistent with anecdotal evidence reported elsewhere (Graham & Hart, 2000). The tag of “unknown” for Root Cause or Solution corresponded to lack of documentation and/or no cause or solution being found.

Table 6. Knowledge re-use statistics

	N	Hours to Resolve		
		Mean	Std Dev	Sum
Fully Documented	102	56	108	5,681
Unknown Solution	35	41	99	1,435
Unknown Root Cause	34	58	93	1,972
Both Unknown	95	72	141	6,820
Overall	266	60	118	15,908

Discussion

This effort was valuable in that certain classes of problems and nuances surfaced rapidly. A few key points were identified during this effort:

- The mean time to resolve all help related cases was about 2 days (49 hrs/case). This jumped to 60 hrs/case for the subset not including phone number requests.
- Over 25% of the cases were due to phone number requests.
- Problems with configuration changes resulted in 22% of the cases.
- Time to resolve cases with Mac OS (9 and X) was considerably less variable than other platforms. The longest Mac case was resolved after just over 10 days.
- Problems with modem connections were frequent and 40% of these cases were due to problems in the Core. Frequency for these cases may be biased by multiple cases for the same problem (e.g., multiple users report one authentication server malfunction).
- Security problems were frequent – especially for end users where over half of the problems were related to obtaining necessary user rights.
- Very little knowledge re-use was possible given the methods used to document cases. Details for this analysis were often extracted by hand and rarely in existing database Root Cause and Solution fields.

In general, it appears the two largest problems specific to remote end users were obtaining phone numbers for their location and obtaining adequate user rights upon connection. Shortly after the sampled data set, SCS Facilities officially rolled out a licensed ISP whose client included a full phone number database and a VPN. Both of these were present in this data as beta tests and early adoptions, but wider use among the population may lead to shifts in phone number and security related cases. Both clients are now part of the standard support package installed on end user machines. However, these also increase the number of required parameters and add to the overall complexity of problem events. In fact, the deployment of VPNs have already led to attempts to

alleviate problems with complexity and monitoring at the system administration end (Kuo & Burns, 2000).

There are still opportunities for better knowledge re-use and dissemination of solutions to common problems to the general population. A newly redesigned Facilities web site may improve the latter. Planned work under this project will include methods for addressing the latter.

The next steps of the project will be to formulate interaction models and develop network interoperability problem resolution methods and tools. It is hoped that these will leverage advances in intelligent networks and streamline interoperability problem resolution for remote access scenarios.

Acknowledgements

This work was completed under a grant from the Office of Naval Research, Interoperability of Future Information Systems through Context-and Model-based Adaptation (#N00014-02-1-0499). The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the Office of Naval Research or the U.S. Government unless so designated by other authorized documents. The Co-PI's for the grant are Dan Siewiorek and Katia Sycara. The authors would like to thank Bob Cosgrove and Mark Puskar of SCS Computing Facilities for their assistance and insight.

References

- Graham, J. & Hart, B. (2000). Knowledgebase integration with a 24-hour help desk. In *Proceedings of the 28th Annual ACM SIGUCCS Conference on User Services, October 2000* (pp. 92-95).
- Kuo, J. & Burns, C. M. (2000). A work domain analysis for virtual private networks. In *IEEE International Conference on Systems, Man, and Cybernetics, 2000, Vol. 3* (pp. 1972-1977).
- Meseguer, J., Agha, G., Gunter, C., Roman, G.-C., & Venkatasubramanian, N. (2003). *MURI Project CONTESSA on Adaptive System Interoperability*. <http://formal.cs.uiuc.edu/contessa/>.
- Salvendy, G. & Carayon, P. (1997). Data collection and evaluation of outcome measures. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics*. New York, NY: Wiley-Interscience.
- Wickens, C. D. (1995). Aerospace techniques. In J. Weimer (Ed.), *Research Techniques in Human Engineering*. Englewood Cliffs, NJ: Prentice Hall PTR.