SANDIA REPORT

SAND97-0614 • UC-900
Unlimited Release
Printed March 1997

Hand Geometry Field Application Data Analysis

Michael Ruehle, Janet Ahrens

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Sandia National Laboratories

SF2900Q(8-81)

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Prices available from (615) 576-8401, FTS 626-8401

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A04
Microfiche copy: A01

DISCLAIMER

SAND97-0614
Unlimited Release
Printed March 1997

Distribution
Category UC-900

Hand Geometry Field Application Data Analysis

Michael Ruehle
Janet Ahrens
Entry Control / Systems Engineering Department
P. O. Box 5848
Sandia National Laboratories
Albuquerque, NM 87185-0782

Abstract

Over the last fifteen years, Sandia National Laboratories Security Systems and Technology Center, Department 5800, has been involved in several laboratory tests of various biometric identification devices. These laboratory tests were conducted to verify the manufacturer's performance claims, to determine strengths and weaknesses of particular devices, and to evaluate which devices meet the U.S. Department of Energy's unique needs for high-security devices. However, during a recent field installation of one of these devices, significantly different performance was observed than had been predicted by these laboratory tests. This report documents the data analysis performed in the search for an explanation of these differences.

Contents

Figures and Tables

1 Introduction

Sandia National Laboratories Security Systems and Technology Center, 5800, has previously been involved in several laboratory tests of various biometric identification devices. These laboratory tests were conducted to verify the manufacturer's performance claims, to determine strengths and weaknesses of particular devices, and to evaluate which devices meet the U. S. Department of Energy's (DOE's) unique needs. Multiple reports have been published, as noted in Section 7. Much valuable information about the various devices was gained from these tests, and much was learned about the most efficient and effective way to perform testing. During the past years, the Recognition Systems, Inc. Hand Geometry Identification Unit, the ID3D HandKey model, has become the default biometric identification device for many DOE sites. This is due to a combination of reasonable performance, reasonable price, ease of installation, and a well established system of users and field support personnel.

Historically, Sandia's operational security organization, the Safeguards and Security Center, 7400, has not seen a need for installation of biometric identification equipment. However, in fiscal year 1993, 7400 expressed a desire to conduct a field test of the HandKey devices at a test building at Sandia. Department 7400's primary interest was in user acceptance and maintainability of the device.

The particular device selected by 7400 for installation in the field trial was the Recognition Systems, Inc. model ID3D HandKey biometric verifier. The decision to use this device was based on the fact that it is referenced in the DOE's *Guide for Implementation of the DOE Standard Badge* and because it appears to be the de facto standard biometric device for the DOE. During fiscal year 1993, the ID3D HandKey was tested in the laboratory by Department 5848. This test was not officially documented due to various funding and test configuration problems, but the HandKey had performed very well and had exhibited equal error rates of about 0.2 percent.

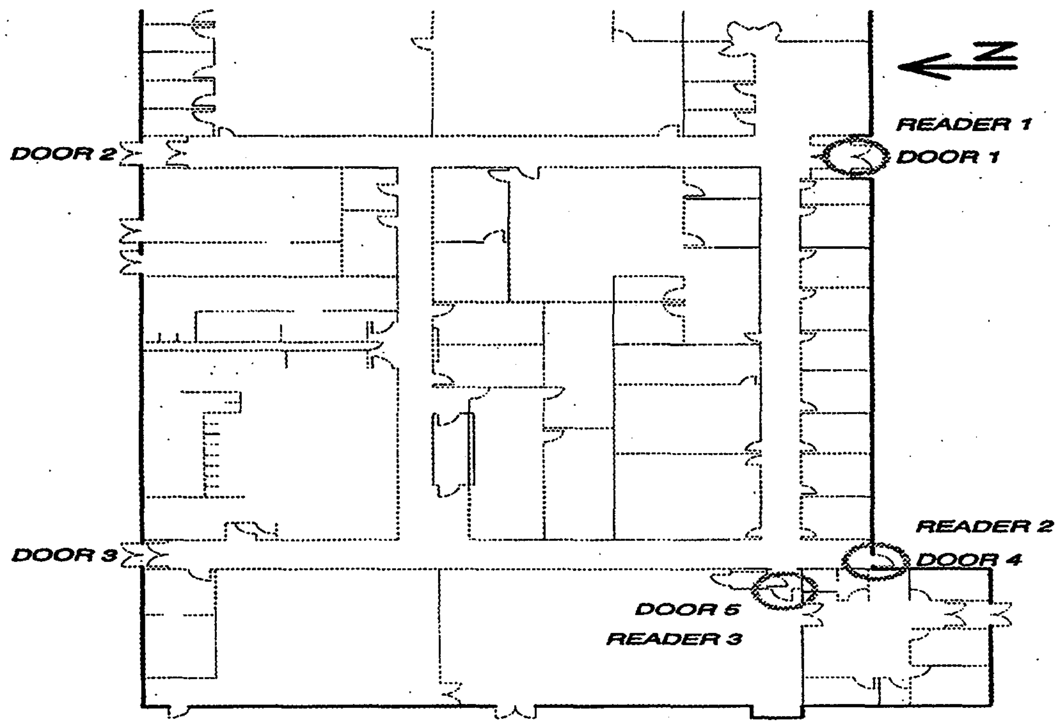
The two groups jointly designed a system and installed three HandKeys on the exterior of Building 956. These devices operated from September 1993 to early October 1995 and were eventually replaced by a card reader system. Data was collected during almost two years of operation. In the summer of 1994, data from the three readers was analyzed and a paper was presented at an Institute of Nuclear Materials Management (INMM) conference (see Reference 8). The system continued to run until early October 1995. During the summer of 1996, Department 5848 personnel analyzed this data to determine whether there was sufficient data to make any conclusions.

The primary conclusion is that the device operates significantly differently in the particular exterior, unattended field installation than in previous lab experiments. While our specific installation is not necessarily similar to the way the devices would be installed at any other DOE site, it does seem that there are some lessons to be learned from our experience. This report is an attempt to share that experience, document the data, and allow potential future users to draw their own conclusions and hopefully learn from our experience.

2.0 Project Background

Building 956 is located just south of Sandia's Technical Area I and houses Sandia's security force and employees of the technical security organization. When the building was originally put into operation, it was surrounded by a standard 8-foot high chain link fence. Originally, the fence line defined the Limited Area boundary. Access through the fence line was gained through a two-door entry control portal equipped with a video camera under the control of a security inspector at the Sandia Headquarters Control Center. This type of booth is common at Sandia and is typically referred to as a Mardix booth.

A floor plan of the building is shown in Figure 1. Once entrance to the area had been obtained, doors 1, 4, and 5 were equipped with Hirsch keypads and all building occupants were informed of the single access code needed to enter the building.



Building 956 Floor Plan

Figure 1

Doors 2 and 3 were locked from the outside and were intended to be used only as emergency exits. In addition to the Hirsch keypads, on-duty Security Police Officers (SPOs) possessed keys to all five doors, as they have bypass keys for all buildings at Sandia. The area west of door 5 is a workout room for the security force. All security personnel (whether on-duty or off-duty) have a key to the exterior door to the workout area. Since the key to door 5 was only in the possession of on-duty personnel, door 5 was considered to be the actual building 956 entrance. For purposes of data analysis no information was available regarding the use of keys or the opening of doors from inside the building.

During the summer of 1993, Sandia determined that DOE was willing to accept the definition of the Limited Area as the building boundary. At this time, Sandia was interested in exploring various access control options. More buildings were being constructed outside the traditional Technical Area I and these buildings were clearly going to require a re-definition of the Sandia Limited Area. Historically, all entrances to the Sandia Limited Areas required identification of the individual by an SPO, either in person or through a Mardix booth. Clearly the desire to have each of these new buildings (or even portions of the buildings) become Limited Areas would not be cost effective if an SPO was required for each entrance. Department 7401 decided to use Building 956 as a test bed to examine various access control alternatives. At that time, Sandia was not sure what would be acceptable to DOE. It was determined that, at the very least, the Hirsch system would require individual access codes rather than the single access code used previously. As long as a change was required, Department 7401 decided to first test what was believed to be a high-security option that would have no problem being accepted by DOE. So, during the summer of 1993, HandKey biometric identifiers were installed at doors 1, 4, and 5. The Hirsch keypads were removed from doors 1, 4, and 5. Also the Mardix booth was disabled and the chain link fence was opened.

In 1995, it was determined that DOE would accept a combined card reader and keypad as an entry control device to a Limited Area. The HandKey devices were removed and a stand-alone automated access control system with a card reader and keypad at doors 1, 4, and 5, was installed about October 1995. When Sandia converted to the DOE Standard Badge in the summer of 1996, this system was removed. At the present time, the entire building has been connected to what is becoming the standard Sandia/New Mexico entry control system. This consists of card readers installed at all five doors and connected to the database of card holders that is continually updated from the badge office.

3.0 Project Purpose

The purpose of installing the HandKey equipment at Building 956 was to gain experience with the devices. The primary concern of 7400 was whether the devices would be acceptable to the population of the building. It was also important to determine how well the devices would work in an exterior environment when subject to heat, cold, blowing dust, and sunlight. There was no attempt to provide routine maintenance for the devices.

4.0 Project Requirements

During the negotiations between 5800 and 7400 at the beginning of the project, the following requirements were defined:

- Enrollment was to be performed by a receptionist or administrative assistant.
- Enrollment must be relatively quick (less than five minutes).
- Enrollment should support individual selection of Personal Identification Number (PIN).
- Use of the system should not impose unreasonable requirements or delays on the staff.
- A system over-ride should exist in case of technical problems.
- Only entry to the building should be controlled, free exit should be allowed.
- Tailgating should be allowed. Any user should be able to open the door and admit as many others as desired. The users should be responsible for ensuring that everyone admitted had a valid badge.

As a result of these requirements, a HandNet, networked system of three HandKeys and a PC enrollment station was installed. This allowed the receptionist near door 1 to enroll users and the hand templates to be downloaded to all three HandKey devices. The receptionist was also provided with an override to unlock door 1 when she received a phone call from a phone located next to the HandKey at that door. This, combined with the fact that the SPOs working the night shift had keys to the building, allowed entrance in case of system failure or for visitors who were not enrolled in the system. A program was developed to assist the receptionist in enrolling users and allowing them to select a PIN.

The building was occupied by approximately 250 people. The occupants included Sandia technical and administrative staff members, as well as SPOs. Since the building housed the SPO's locker room and gun room, they required access 24 hours a day, including weekends and holidays. Most traffic through the building entrances occurred at shift changes.

5.0 Hand Geometry Installation Details

5.1 Hardware Details

The HandKey units were actually installed and placed into operation in September 1993. The device installed at door 5 was referred by data analysts as reader 3. This device was a standard ID3D HandKey unit and did not report actual scores which result from comparing the template of the claimed identity with the template obtained from the person requesting entry. This reader was installed on an interior wall in the workout room. This room is climate controlled but is usually slightly warmer and more humid than a regular office environment.

The devices installed at doors 1 and 4 are referred to as reader 1 and reader 2. These were also unmodified HandKey systems, but they included a custom program provided by Recognition Systems, Inc., which reports actual scores from each comparison. This allows more detailed analysis of the data by letting the analyst determine, mathematically, what the false reject rate would be for various acceptance thresholds. These two readers were installed on the exterior wall of the building. Reader 1 was in the entry alcove and on an exterior west facing wall. Reader 2 was on a south facing wall. Both readers had an aluminum faring extending over the top and sides to shade and protect them from direct rain and snow.

All devices were networked together and provided with identical templates after each enrollment. The log information was then saved on the computer near the receptionist.

5.2 Enrollment Details

Building occupants were given access authority based on their role within the building. Staff were allowed access only during normal working hours. SPOs were allowed 24-hour access daily. Maintenance personnel were allowed access only during normal working hours but were also given the authority to configure the enrollment station. Managers/Administrators and the receptionist were given 24-hour access daily and were also given the authority to configure the enrollment station.

The receptionist was responsible for enrolling new users. As people moved into the building or requested continuing access to the building, the receptionist enrolled them and generally provided them with 5-10 minutes of instruction before allowing them to use the system.

As employees transferred out of the department or were reassigned to another building, the receptionist was responsible for removing their template from the database, which automatically removed their template from all three readers and denied them access in the future. Visitors were admitted by phoning the receptionist or the person they were going to visit. They could also enter by tailgating with any authorized user.

6.0 Data Analysis

6.1 Previous Data Analysis

During the spring of 1994, the data from the three readers was copied for analysis by 5848. Jose Rodriguez and Dale Murray performed the analysis (documented in Reference 8) and several studies (summarized below).

They investigated the amount of variability between repeated trials of the same hand on a single reader. This was done by enrolling a mechanical "surrogate" hand and then making repeated attempts to verify the same mechanical hand. At least 200 measurements were taken from each of the three readers. The result of this analysis was that the difference in scores varied a maximum of 1 point (where a reading of less than 100 is usually considered an acceptable match for an entry control situation).

Next, they investigated the amount of variability between different readers when presented with the mechanical "surrogate" hand. The result of this analysis was that the difference in scores varied a maximum of 5 points.

They then analyzed the amount of variation in a real individual's scores. Considering only the data from readers 1 and 2 which reported scores, they looked at all scores for each PIN and calculated an average score and deviation. They found the average score to be 34 with a standard deviation of 20. So most of the variability seems to be due to an individual's human variability in the alignment of the hand on the device.

The final data analysis performed in 1994 was a calculation of the false reject rate. The results of this analysis showed a first try false reject rate of about 2 percent. Analysis of the data showed that those test subjects who failed on the first try and immediately made a second attempt experienced a second try reject rate of 43 percent. Similarly, subjects who failed twice and made a third attempt to use the device experienced a third try reject rate of 58 percent. In comparing this to previous laboratory data, it was clear that the system performs much better in the lab under controlled conditions. This is evidenced by two facts. The first is the simple single try false reject rate, which had been reported as about 0.2 percent in the lab versus 2 percent in the actual installation. Another item to note is that the second and third attempt numbers are much higher in the field than in the lab. Previously, it had been believed that during laboratory tests, people are usually more careful after they have been rejected and they tend to place their hand more correctly on the second and third tries. In the field application, we found that actually only half the people who were rejected on a first try even bothered to try a second time. We believe that SPOs may have used their key after the first failure. Office staff may have had a friend open the door for them. When they did attempt a second or third try, it may have been more likely that their first failure was due to the dirt or sun or the fact that they mistakenly entered someone else's PIN. All of these factors are unlikely to allow them to improve their performance for the second or third try simply by being more careful.

In addition to analysis of the data from the readers, Reference 8 describes the development of a tool to allow the security system operator to optimize individual thresholds. This allows "fine-tuning" of the system to possibly increase security without completely rejecting those individuals who consistently are unable to be verified by the system. For more detail on the results of the optimization see Reference 8.

6.2 1996 Analysis of HandKey Field Data

The subsections below include various ways of looking at the data from the HandKey devices. Beginning with Section 6.3.1, each subsection begins with a graph or plot of data. Each graph is followed by an attempt to discuss what conclusions might possibly be made from that particular method of looking at the data. This work was performed by Mike Ruehle during the summer of 1996. No attempt has been made to generalize or draw too many conclusions, and the amount of effort spent in performing the analysis was whatever Mike was able to perform along with his other assignments for the summer.

6.3 General Analysis

This section includes background information about the data. Basic information regarding the data analyzed in this report is summarized in Figure 2. The "HandNet" access control system was running in Building 956 as early as September of 1993. However, before mid-April 1994, the score data was not available due to previous analysis, and in other ways the log-file data was difficult to use. Therefore, this analysis considered only the transactions from April 14, 1994, until the final system shutdown on October 2, 1995—transactions from about a year and a half.

Hand Geometry Reader Use Statistics

Total Transactions	117,213
Total 1st tries	112,757
Total 2nd tries	3,489
Total 3rd tries	967
Reader 1 1st tries	67,398
Reader 2 1st tries	8,285
Reader 3 1st tries	37,074
Number of users	316
Start date of data	4/14/94
End date of data	10/2/95
Days of data	537

Figure 2

Three hand geometry units were installed as part of the HandNet system. Reader 1 stood at the main (southwest) entrance and logged the majority of the transactions. Reader 2 was at a secondary (southeast) building entrance and was used much less frequently. Reader 3 secured an inside door between the gym and the secure portion of the building and was used with moderate frequency. The location of these readers is also shown on Figure 1. Overall, the number of hand scans taken during the data period was in excess of **one hundred thousand**, generating far more data than any previous controlled laboratory biometrics test at Sandia.

The first major step of processing the HandNet log file data was to divide the hand scan transactions into first, second, and third tries. This seems reasonable because the HandNet system itself is set up to allow a user three tries to gain access. Since the HandNet software did not record try numbers, this was accomplished by a "C" program which decided that a given transaction was secondary to a previous one if:

- the transactions occurred at the same reader,
- the transactions were made by the same user (same PIN),
- the later transaction came within two minutes of the first and on the same date,
- the former transaction was a reject ("ID REJECTED"), but not a third and final reject ("ID REFUSED"), and
- the former transaction was assigned a try number of 1 or 2.

This program output a formatted text file with each line holding scores and accept/reject decisions for a set of one, two, or three tries, along with time and date, PIN, and reader number (1, 2, or 3). This file was imported into a Microsoft FoxPro database, from which the various results presented in this report were extracted. (Microsoft Excel was used to graph these results.)

As can be seen from Figure 2, only about 3 percent of first tries led to a second try, and 28 percent of these led to a third try. As discussed above, we believe that, in many cases, the person who was rejected probably waited for another employee and tailgated through the door. This was verified through discussions with some users. In the remainder of this report, except where the three tries are considered separately, only first try data is generally considered, so that second and third tries resulting from low-scoring transactions do not skew the results.

Since the data period started about half a year after the installation of the HandNet system, many of the 316 users with transactions logged were enrolled before the start of the data period, but some others were enrolled at various times throughout the system lifetime. The HandNet system did not log any information on the quality of enrollments or whether an enrollment was of a new or previously enrolled user, so not much can be determined regarding these issues or any possible learning curve.

The number of first-try transactions represents an average of 4.65 first tries per user per week. Not all users used the system actively during the entire data period, so the actual average frequency of use may be somewhat higher than this. Additionally, it should be noted that "tailgating" (a user accepted by a hand geometry unit allowing another cleared individual to follow him immediately through the door without scanning) was explicitly allowed in the requirements for this access control system. From the statements of users, tailgating was often practiced, so the frequency with which users entered building 956 is probably significantly higher than 4.65 times per user per week.

6.3.1 Reader Use by Month

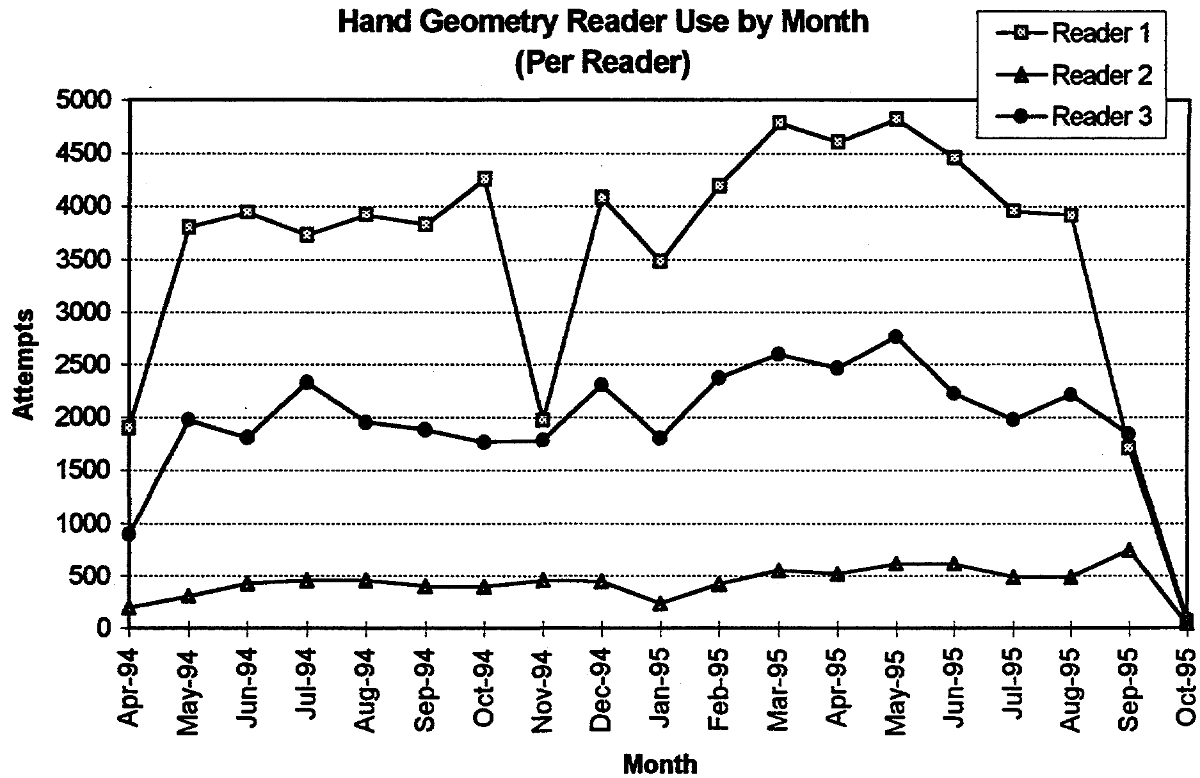


Figure 3

Figure 3 shows the reader use on a monthly basis. When seen on a scale of months, the three readers were used in fairly constant proportions throughout the data period. Overall use was up slightly around spring of 1995. The low amount of use in April of 1994 results from our considering only data after the 14th of that month. Similarly, the almost complete drop off in October 1995 can be explained by the data period ending on the 2nd of that month. However, reader 1's usage dropped by more than half during September 1995, the month before the end of the data period. This is because the new card reader/keypad device was initially installed only on door 1, so users were free to use either the card reader or the HandKey during part of September at door 1.

The three readers experience a common dip in use during January 1995. This may be taken to reflect vacations during the holiday period.

Reader 1 also shows a major drop in usage during November 1994. Neither of the other two readers has anomalous usage at this time.

6.3.2 Reader Use by Week

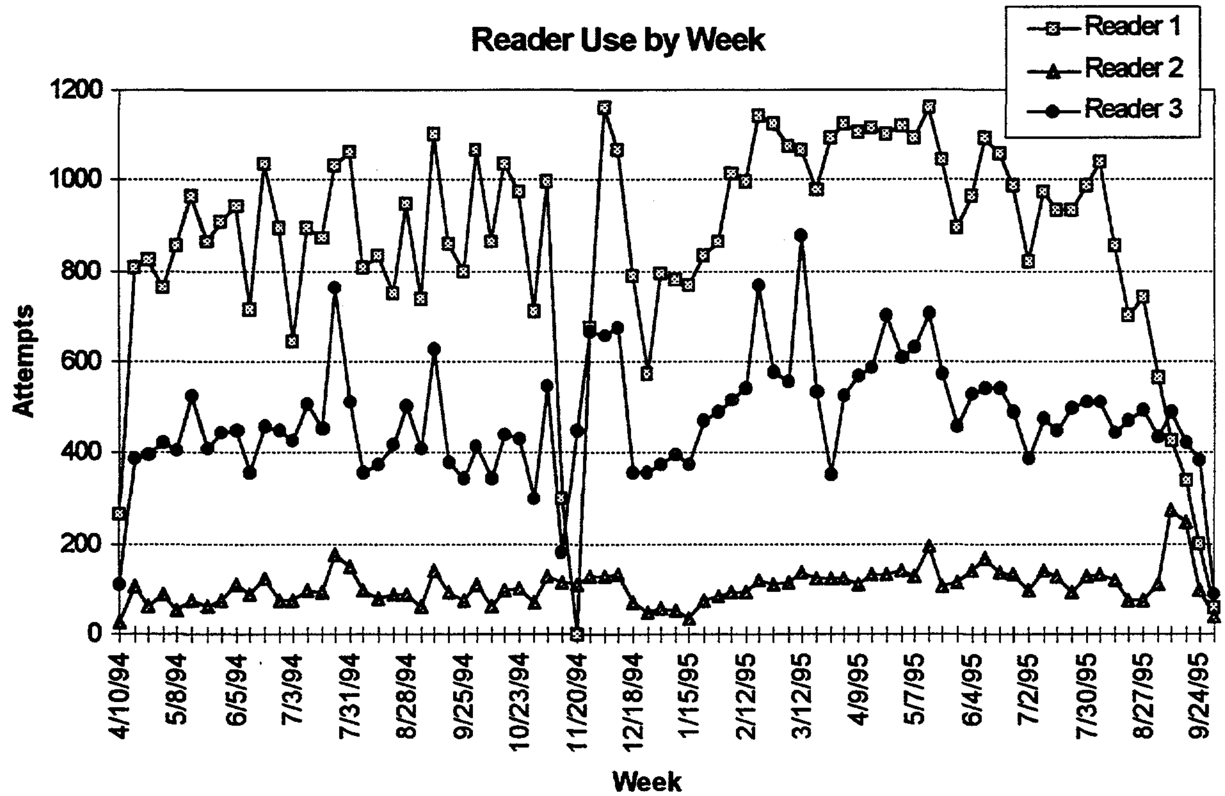


Figure 4

When the reader use is plotted by week as in Figure 4, the major drop in the use of Reader 1 during November is seems to be even sharper than is apparent from Figure 3.

On a weekly scale, there is considerably more noise in the amount of use the system as a whole received. However, there is still much correlation among the three readers' usage.

It is clear from this chart that Reader 1 had no usage logged for the entire week of November 20, 1994. A look at the data reveals that no transactions were logged for reader 1 from 1:55 PM on November 15, 1994, until 7:20 am on November 30, 1994. It is possible that there were, indeed, transactions, but these transactions were not logged; or perhaps reader 1 was completely out of service for the two-week period. We were unable to discover the cause of this anomaly.

A final interesting note here is that while readers 2 and 3 dropped off rather abruptly at the end of the data period, reader 1 shows a gradual decline in use for the last eight weeks of the data period. Since users were allowed to use either system after their badges were encoded for the new card reader, and the new reader was installed first at reader 1, it may be that some people actually preferred the HandKey system.

6.3.3 Reader Use by Day of the Week

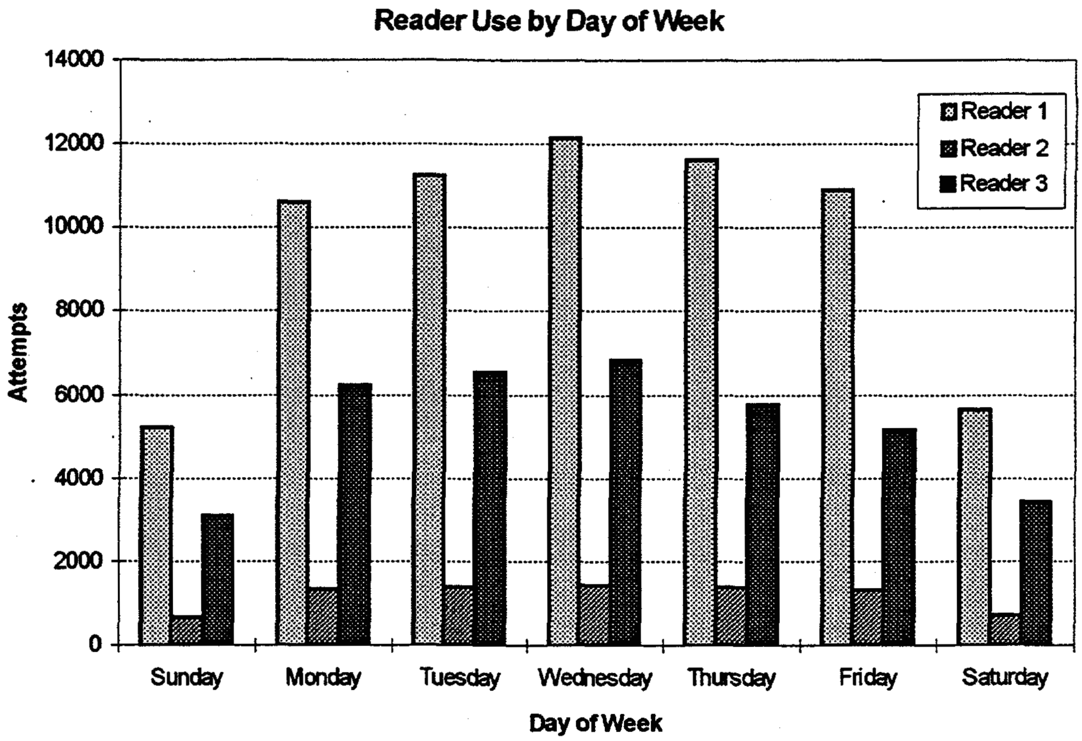


Figure 5

Figure 5 shows reader usage by day of the week. There is not much to say about Figure 5, except to note that fewer people entered building 956 on Fridays and Mondays than in the middle of the week, with use on weekends about half of that during the week.

6.3.4 Reader Use by Day of the Week

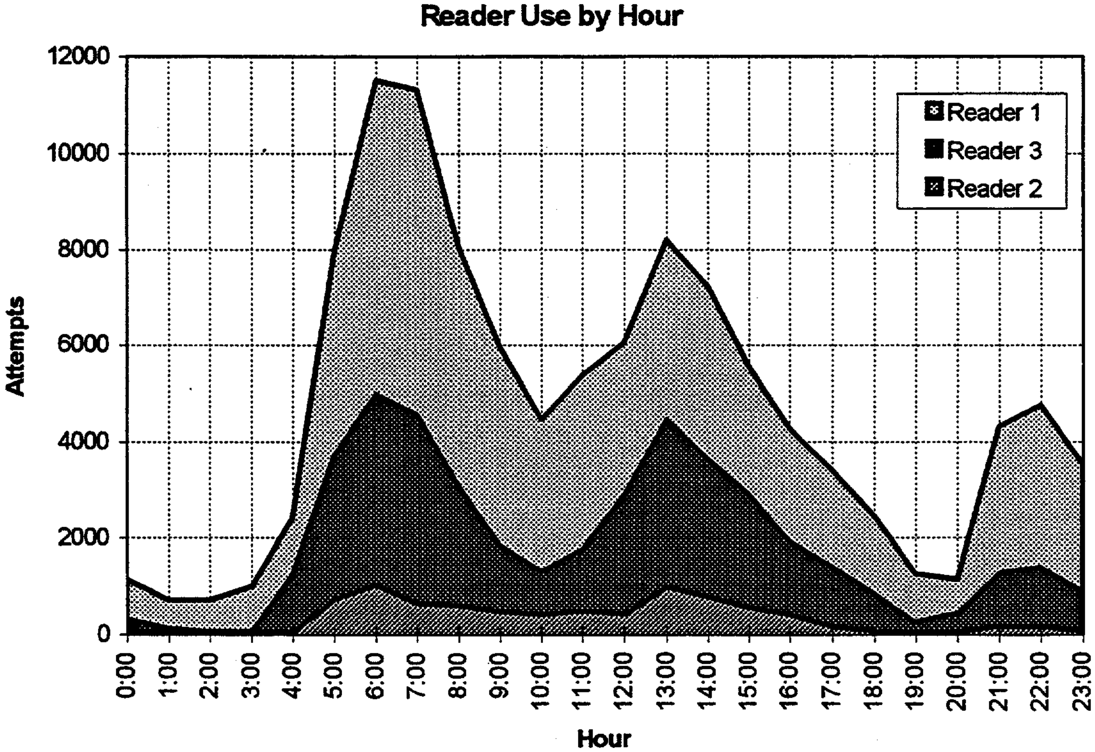


Figure 6

Continuing to analyze reader usage, Figure 6 shows usage by hour of the day. Figure 6 shows the expected peaks in usage around the times of the three guard shift changes, at 7:00 AM, 3:00 PM, and 11:00 PM. The afternoon peak actually comes somewhat early, but this may be because of lunch time for other employees.

The other note to be made is that the three readers stayed in fairly constant proportions of use, except that reader 3, at the door into the gym, logged less than its share of transactions at night, which is not surprising since it is further from the parking lot and less well lighted.

6.3.5 Reader Use Number of Attempts per User

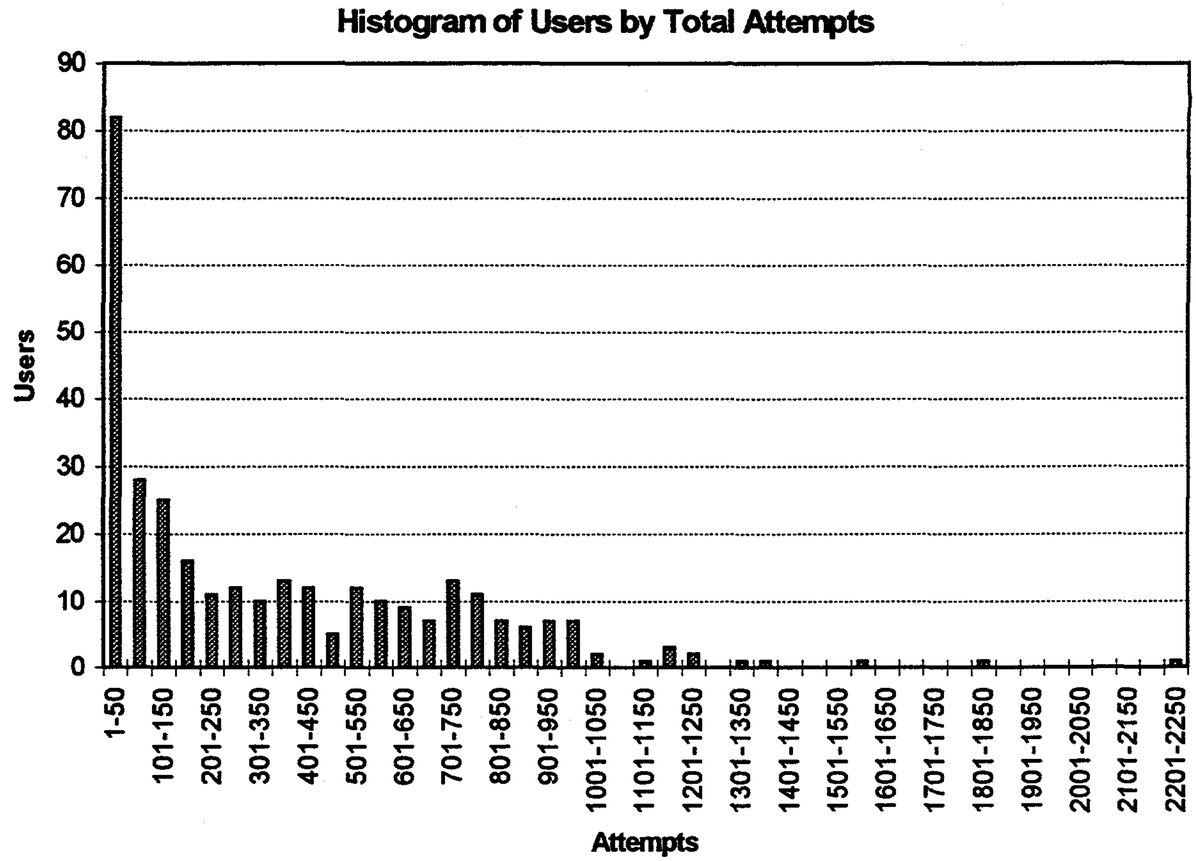


Figure 7

Figure 7 shows the number of attempts made by various users of the system. The most interesting feature of Figure 7 is the bias toward low amounts of use per individual. Over a quarter of the users had their hands scanned no more than fifty times over the year and a half data period. This may be due to a dislike of the system and a tendency to tailgate. Since there are over 300 users, it also appears that some people who were really building visitors were enrolled in the system even though their office was in another building. This is known to be the case for several managers and DOE employees.

Also, the tail of the histogram extends to the right quite a long way. The over 2200 transactions logged for the most frequent user works out to over 28 transactions per week on average, and this with the possibility of tailgating. The building is a no-smoking building, with smoking areas designated outside doors 2, 3, and 4. It may well be that the more frequent users are smokers who happen to frequent the area outside of door 4.

6.3.6 Score Statistics

Score Statistics

Total Scores	79,146	
Total 1st try scores	75,680	
Total 2nd try scores	2,737	
Total 3rd try scores	729	
Reader 1 1st try scores	67,395	
Reader 2 1st try scores	8,285	
Reader 3 1st try scores	0	
Total 2222's	4,215	(5.3%)
1st try 2222's	3,363	(4.4%)
2nd try 2222's	635	(23.2%)
3rd try 2222's	217	(29.8%)
Reader 1 1st try 2222's	2,390	(3.5%)
Reader 2 1st try 2222's	973	(11.7%)
Reader 3 1st try 2222's	0	

Scores	1st Try	2nd Try	3rd Try
Mean	138	634	785
Std. Dev.	457	888	948
Mean Log	3.33	5.13	5.54
Std. Dev. Log	1.31	1.77	1.69
Median	22	129	214

Figure 8

This section is an introduction to various ways of looking at the score information for the devices. Figure 8 summarizes by providing some counts for score data. Reader 3 did not have the custom software required to report scores, so only 75,680 first-try transactions (67 percent of first-try the transactions logged) included score data.

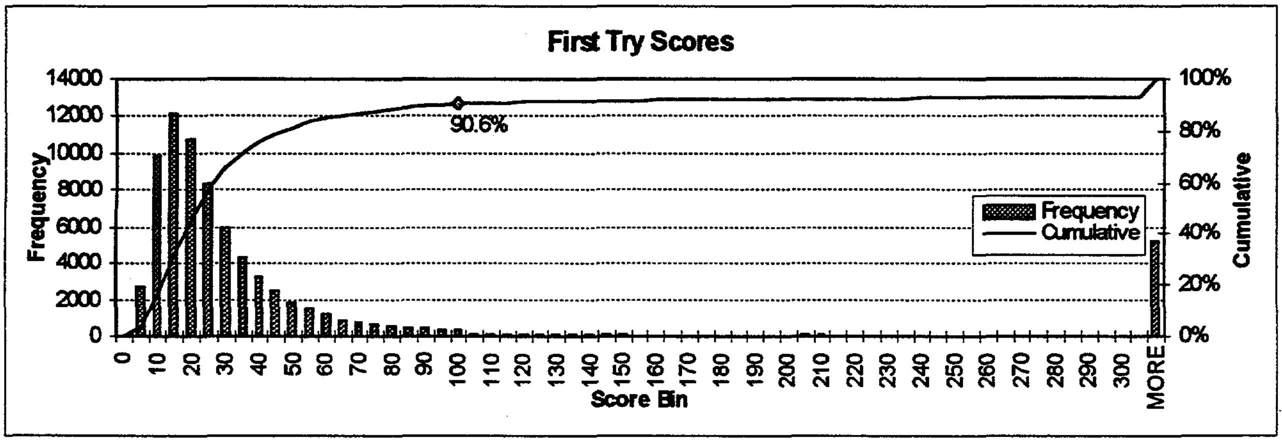
Scores ranged from 0 to 2222, with 0 being the best possible score, and anything under a threshold value (100 by default) considered an accept. There are a large number of 2222's present among the scores; we do not know whether a score of 2222 has some special significance other than being a definite reject, or why no scores greater than 2222 are produced. However, this is consistent with laboratory tests where we have observed that a score of 2222 is given by Recognition Systems, Inc.'s programming when the hand presented is in some way clearly not the hand of the actual enrolled user, or when the hand is placed extremely improperly. Since we were not absolutely certain of the meaning of this score, we generally did not treat this data any differently or remove it from the data set.

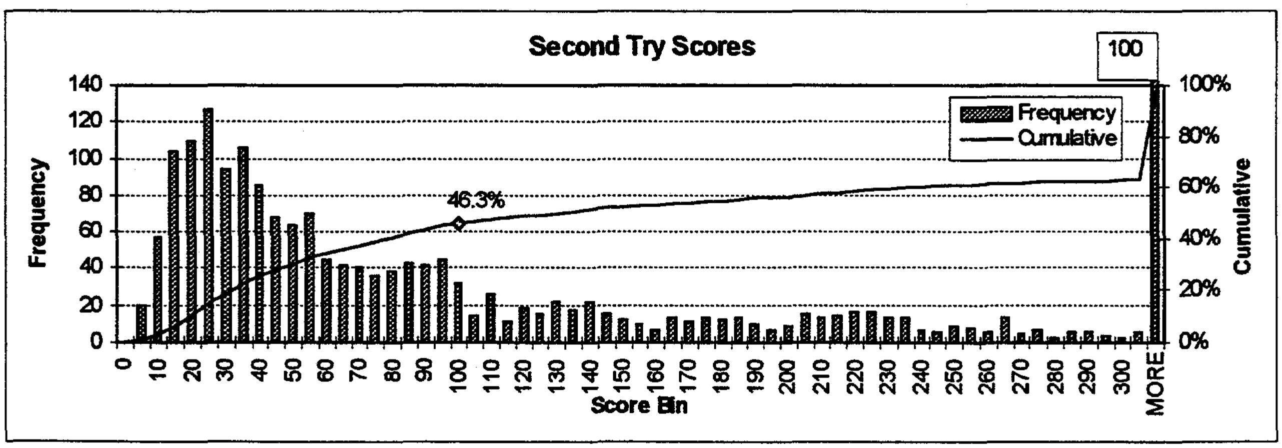
The proportion of 2222's did increase for second and third tries, but so did scores in general, as shown in the other statistics in Figure 8, and as score histograms in Figure 9 will demonstrate.

Perhaps more significantly, reader 2 has a much higher relative frequency of 2222's than does reader 1 (reader 3 does not report scores). Figures 14, 15, and 26 will show that reader 2 rejected users with much greater frequency than did either reader 1 or reader 3. A few former users of the building 956 HandNet system were interviewed and agreed (separately) that reader 2 behaved strangely, often not recognizing that a hand was in the reader. We are unable to determine any technical reason for this difference but have not been able to test the equipment in the lab since this discrepancy was discovered.

The large proportions of 2222's substantially bias the mean scores presented above. For this reason, we have from this point on taken the median as the best single measure of score data.

6.3.7 Score Graphs





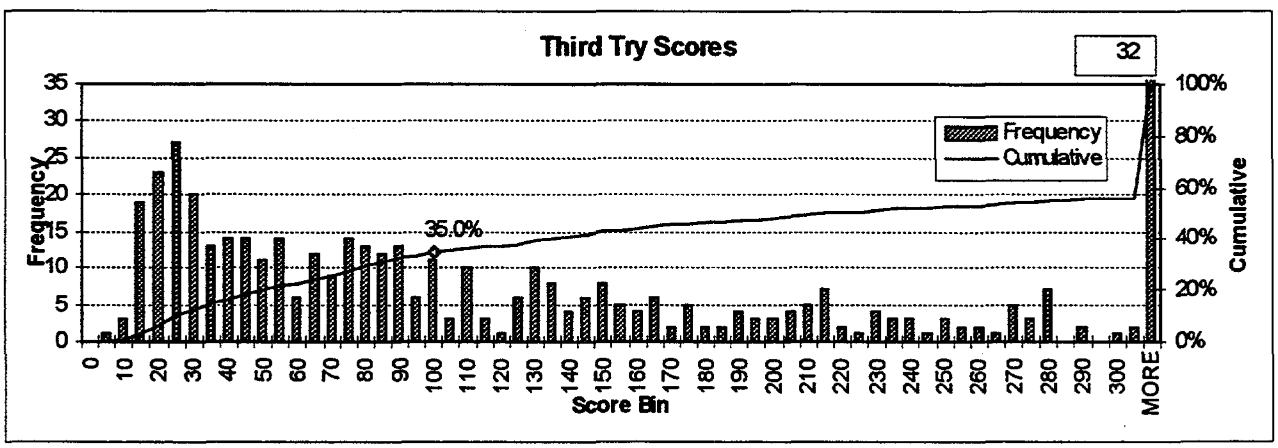


Figure 9

The mean logs and standard deviations of logs of score data sets are given in Figure 9. The distributions of scores—divided into first, second, and third tries—appear to generally exhibit the shape expected. It has been suggested that these graphs may be fitted to log-normal distributions; but in fact, the log-normal distributions obtained from the means and standard deviations of logarithms of scores do not fit well at all.

It is easy to see that the mass of scores moves to greater values in second and third tries. This is a common effect with biometrics: those users who give a second try by definition failed on the first, and so are likely the users who perform worse overall. Beyond this, though, the scores tend especially in second and third tries to accumulate in extremely high values; for third tries, nearly half of all scores are over 300. In fact, comparison with the statistics in Figure 8 shows that perhaps two-thirds of all of these high (300+) scores are actually 2222's, in all three tries.

The increased noise in the second and third try histograms is due to the smaller number of transactions represented.

The cumulative distribution as a fraction of the whole (the lines in the histograms above) can be taken as probabilities that first, second, or third try scores will be at or under the threshold in question. In particular, this suggests a 90.6 percent probability for an accept on the first try at the default threshold of 100. The values of 46.3 percent and 35.0 percent for second and third try accepts at this threshold are more theoretical, since not all users were actually operating at the default threshold of 100. In any case, not everyone who failed to gain access on the first try made a second attempt, and likewise for third attempts.

6.3.8 Histogram of Users With Score 2222

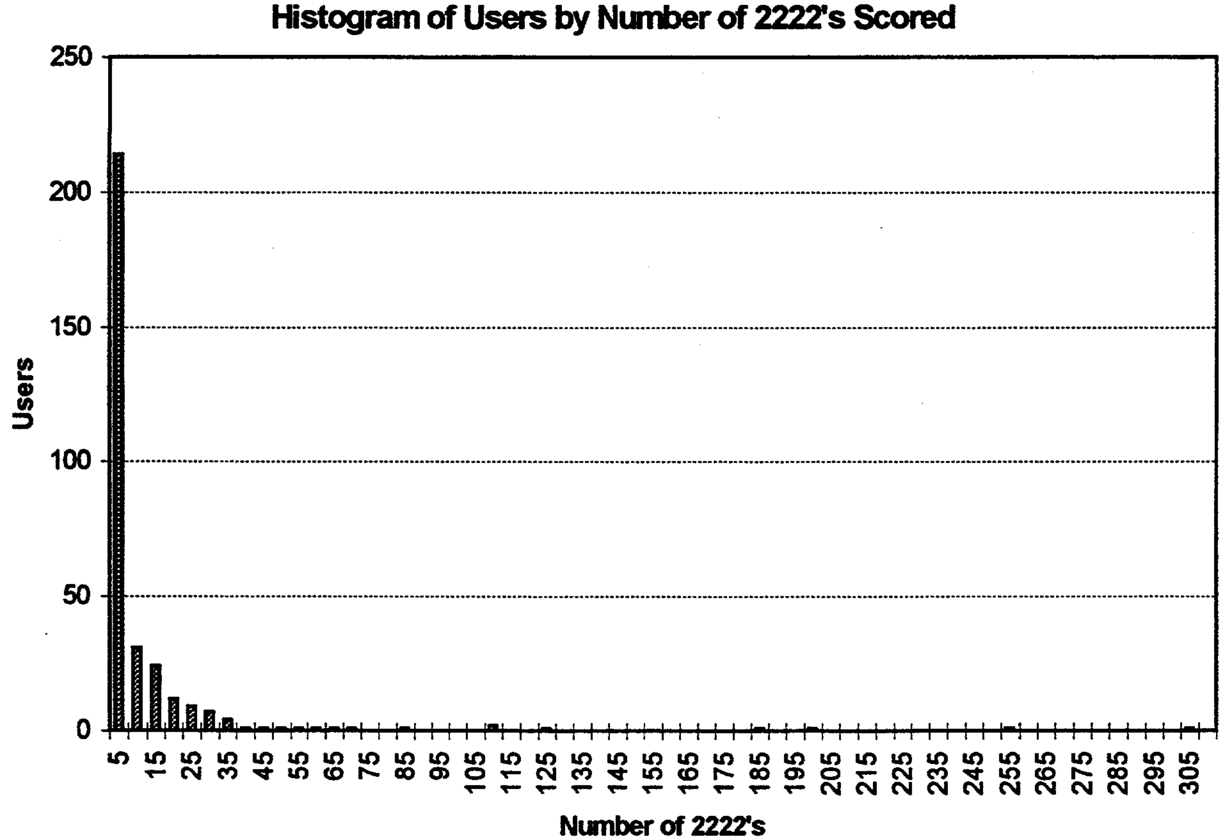


Figure 10

As Figure 10 shows, the majority of users scored less than five 2222's. A significant minority of users, though, scored on the order of a few tens of 2222's; and a handful of users single-handedly generated one or more hundreds of 2222's. Since there was also a wide distribution in the total number of attempts made by individual users, the usefulness of this chart basically ends there.

6.3.9 Analysis of Percent of 2222's Scored

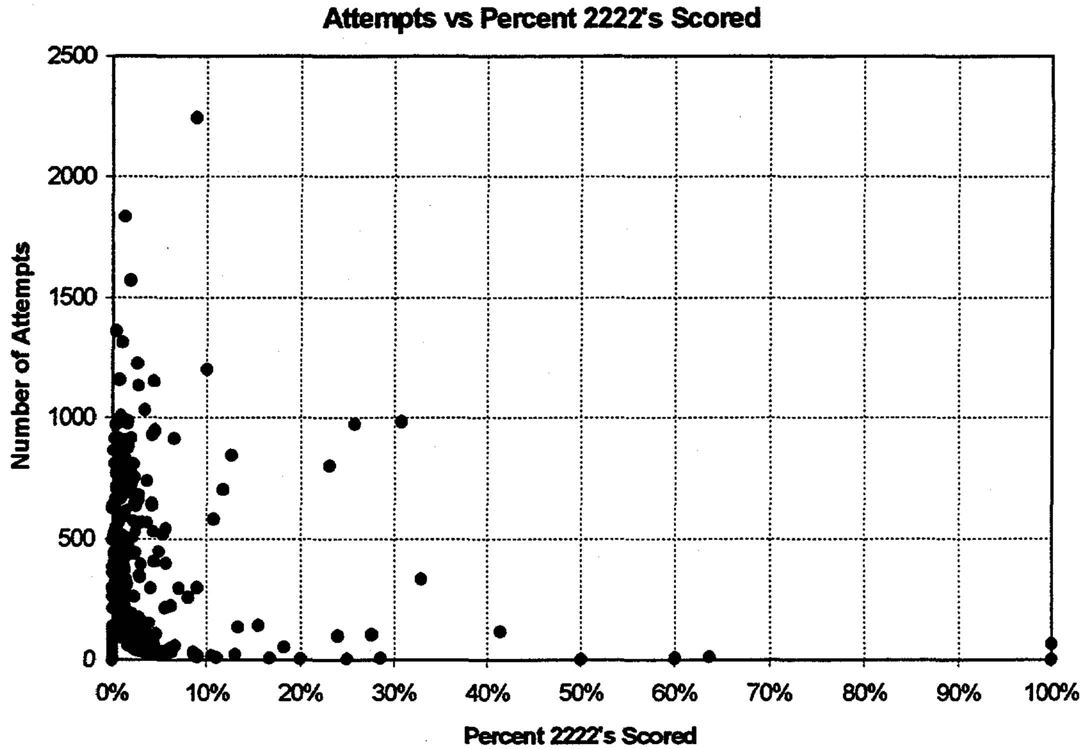


Figure 11

Given that the score of 2222 appears to mean that the hand presented was not acceptable for some reason, it is not surprising to see any given user generating this score a small percentage of the time; nor is it especially surprising to see a user with very few total attempts having a large portion of scores being 2222's. What is interesting, at least, is to see a user making 1000 attempts, 30 percent of which are 2222's. Several users have points in Figure 11 plotted away from the lower and left edges of the above graph, indicating that a significant fraction of their attempts resulted in a score of 2222. Since these users could not get into the building, it is unclear to us why they did not complain and contact us about the poor performance of the system. Perhaps they were intentionally testing the system by using their friend's PIN. Perhaps they forgot their chosen PIN and were not willing to admit the problem in order to be re-enrolled.

6.3.10 Median Scores of Users

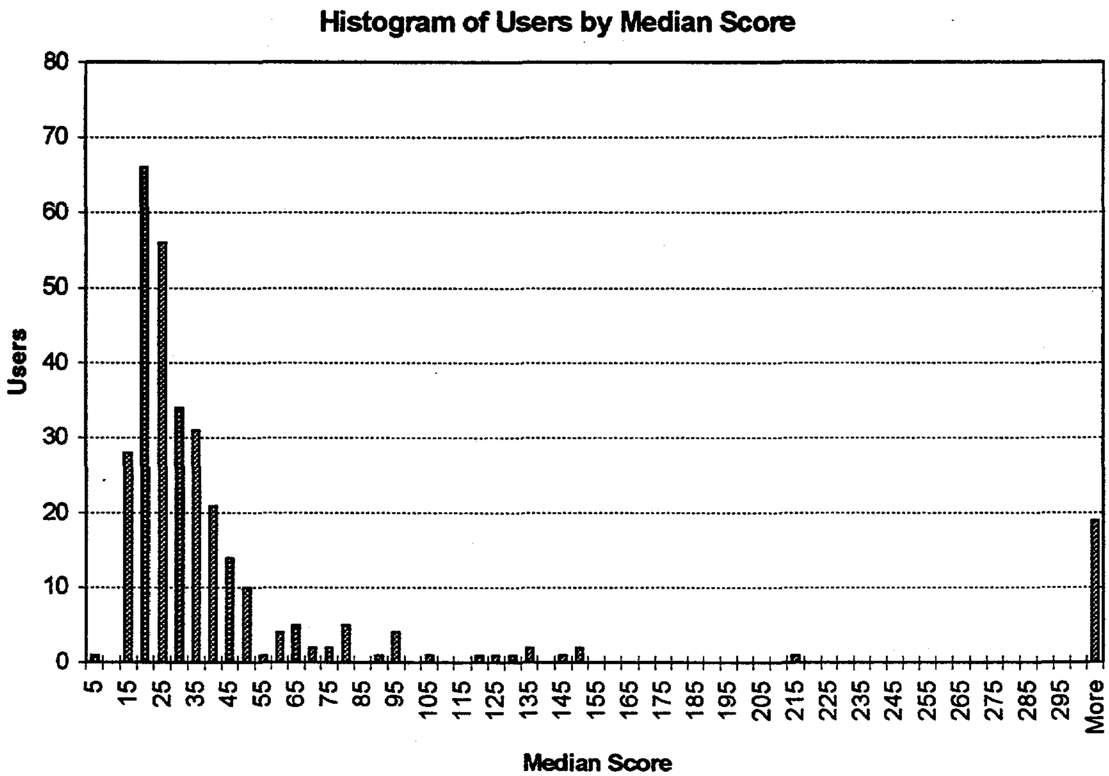


Figure 12

As was stated above, the disproportionate number of high scores, especially 2222, has led us to take the median as more appropriate a measure than the mean of a set of scores from the HandNet system. Figure 12 shows the vast majority of users with median scores well under the default threshold of 100. There are also, however, a significant number of users with median scores near or above 100, including almost 20 individuals with median scores off this chart (300+). Note that since the measure we are using is the median, an individual with a median score of 100 has by definition scored under 100 half the time, and over 100 the other half (neglecting scores equal to 100.)

6.3.11 Median Scores with 2222's Excluded

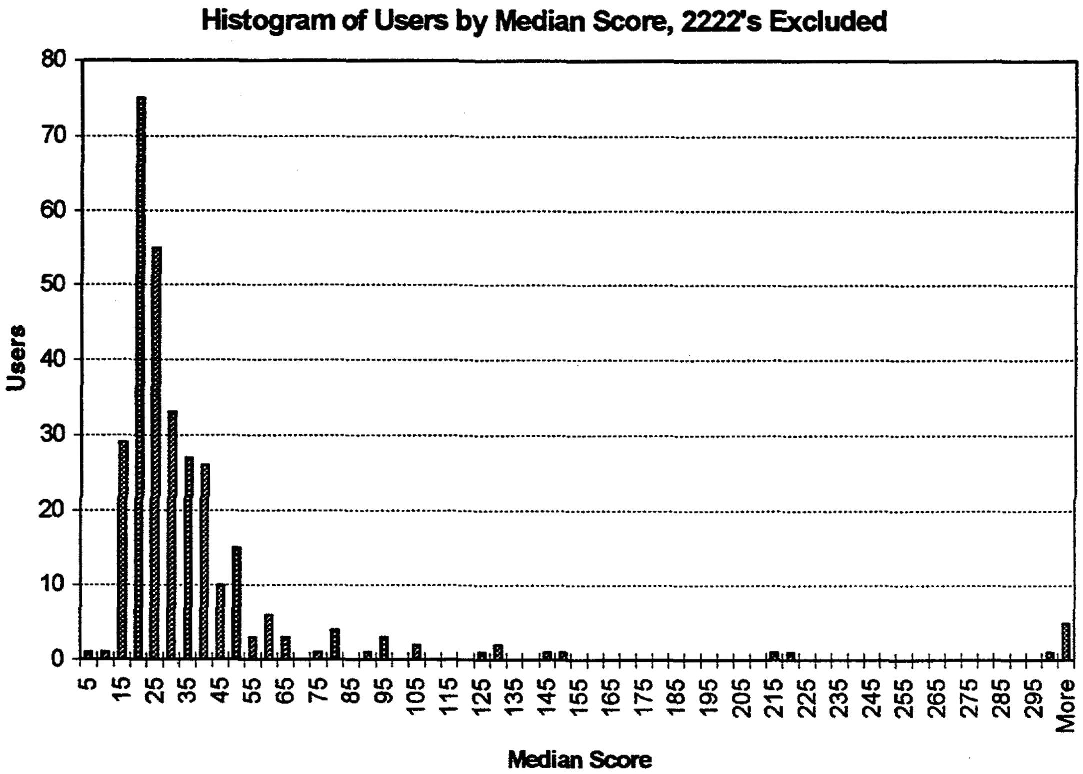


Figure 13

If we ignore all first-try transactions with scores of 2222, the histogram in Figure 12 changes to that in Figure 13, above. Most of the users with median scores over 300 disappear, and the remainder of the histogram improves (shifts toward lower scores) very slightly. The portion of users with median scores over the default threshold of 100 shrinks substantially by this process. Overall, though, the histogram is largely unchanged by removing the 2222's.

6.3.12 Reject Rate Analysis, All Transactions

Table of Reject Rates

Attempts by 6 "goats" removed				
Rdr 1 attempts during glare removed				
Number of attempts removed	-	6,272	13,219	18,804
Size of resulting data set	112,757	106,485	99,538	93,953
Reject Rates by "A" / "R"				
1st try reject rate	7.20%	5.72%	6.12%	4.67%
2nd try reject rate	53.48%	49.71%	57.82%	53.18%
3rd try reject rate	66.49%	65.08%	69.83%	64.65%
Actual 3-try reject rate	5.48%	3.95%	5.01%	3.54%
Theoretical 2-try reject rate	3.85%	2.84%	3.54%	2.48%
Theoretical 3-try reject rate	2.56%	1.85%	2.47%	1.61%
Reader 1: 1st try reject rate	7.69%	5.85%	5.82%	3.93%
Reader 2: 1st try reject rate	19.28%	14.92%	19.28%	14.92%
Reader 3: 1st try reject rate	3.62%	3.51%	3.62%	3.51%

Table 1

Here we begin analysis of "reject rates," which are the fraction of the time that users were rejected by the system under various circumstances. There is a tendency to think of these values as "false reject rates," meaning they represent users rejected who should have been accepted, but here we do not have enough information about the actual circumstances of each transaction to make such a determination. They may have been rejected because they were not who they claimed to be.

The method used to calculate the reject rates in Table 1 is just counting the transactions marked as "Rejects" in a subset of our data and dividing that by the total size of that subset. Another method, involving score data, will be introduced in Table 2.

Table 1 is read as follows: the first two rows show the optional removal of two subsets of data from the original data set; columns which have their intersections with one of these rows shaded contain results reflecting the removal of that row's subset. The third row shows the net number of attempts removed by this process, and the fourth shows the size of the remaining data set. Note that the fourth column comes from removing both of the data subsets, as it is shaded in both of its first two rows, but the number of transactions removed for that column is somewhat less than the sum of the numbers of transactions removed in the previous two columns because the two data subsets are not disjoint. The 6th through 14th rows of the table contain various reject rates calculated for the data set of each column.

The first (numerical) column of this table is instructive by itself. The first, second, and third try reject rates given here (7.20, 53.48, and 66.49 percent) are somewhat better than the accept rates (90.6, 46.3, and 35.0 percent) found by examining the cumulative score histograms in Figure 9 at threshold scores of 100. (Subtracting these latter three numbers from 100 percent to get reject rates yields 9.4, 53.7, and 65.0 percent for first, second, and third tries.) This can be explained in two possible ways: either reader 3, which did not report scores, accepted users with greater frequency than did readers 1 and 2, or else users must have been accepted with scores greater than 100. In fact, both of these explanations are true. Reader 3 evidences a 3.62 percent reject rate, compared to the 7.69 percent reject rate of reader 1 and the 19.28 percent reject rate of reader 2. Additionally, a look at the data reveals that there were quite a few instances of users accepted with scores greater than 100, as well as a smaller but still significant number of times when a user was rejected with a score under 100. We do know that an attempt was made to tighten security by "optimizing" users' individual thresholds, primarily setting them to values *under* 100 (determined by a process involving their individual score distributions). We also suspect that many users were enrolled with thresholds *over* the default of 100, or had their thresholds adjusted to such values in order to make verification easier. What will be shown in Table 2 is that the latter adjustment had a greater net affect. Overall, security was looser in this system than it would have been with consistent user thresholds of 100.

A few other rows of the first column deserve explanation. The actual three-try reject rate (5.48 percent) is the portion of the time that a user did not, in the one, two, or three consecutive attempts he or she made, get accepted. Thus, fifty-five times out of 1000, a user had to call for someone already inside the building to open the door or tailgate through behind a more successful user, or whatever. If a user tried once, failed, and quit, that transaction is counted as a failure here, just as is a set of three consecutive failures. The theoretical two-try reject rate (7.20 percent) is just the product of the first-try reject rate and the second-try reject rate; it is, in theory, the odds of a users being rejected by the system if whenever they fail on the first attempt, they always try again. Similarly, the theoretical three-try reject rate (2.56 percent) is the product of first, second, and third try reject rates, and represents the chances of users being accepted who keep trying up to three times to gain access. This is significantly lower than the actual three-try reject rate (5.48 percent) because users often did not try again when they were rejected.

The second column of this table shows how all of these values change when the transactions by six users with unusually bad performance ("goats") were removed from the data set. (The method used to select these users is shown in Figure 17.) Although these six users represent less than 2 percent of the population, and their transactions some 6 percent of the data set, their removal substantially lowers the various reject rates.

The third column has reject rates for a data set which does not include transactions made on reader 1 between 1:30 PM and 6:15 PM. The reason for this step is that the angle of sunlight at these times dramatically hurt the performance of reader 1; one definite recommendation for future installations of a hand geometry access control system is to properly shield the readers from the sun. Figures 27 and 28 further demonstrate this point, but it is noteworthy here that removing these transactions lowers reader 1's reject rate from 7.69 percent to 5.82 percent.

The fourth column gives the results of removing both the transactions due to the six "goats" and the transactions during the "glare" period for reader 1; the reject rates here are naturally lower than in either of the two preceding columns.

6.3.13 Reject Rate Analysis for Transactions With Scores

Table of Reject Rates, Transactions Without Scores Removed

Attempts w/ scores = 2222 removed		■		■		■		■
Attempts by 6 "goats" removed			■	■			■	■
Rdr 1 attempts during glare removed					■	■	■	■
Attempts w/o scores removed	■	■	■	■	■	■	■	■
Number of attempts removed	37,074	40,437	42,310	44,547	50,293	52,624	54,842	56,270
Size of resulting data set	75,683	72,320	70,447	68,210	62,464	60,133	57,915	56,487
Reject Rates by "A" / "R"								
1st try reject rate	8.96%	4.72%	6.85%	3.79%	7.61%	4.03%	5.40%	3.01%
2nd try reject rate	54.73%	41.06%	49.89%	37.80%	62.41%	49.86%	55.78%	45.37%
3rd try reject rate	68.45%	55.08%	66.72%	54.05%	76.14%	66.27%	73.27%	64.32%
Actual 3-try reject rate	7.02%	3.31%	4.85%	2.38%	6.60%	3.14%	4.37%	2.14%
Theoretical 2-try reject rate	4.90%	1.94%	3.42%	1.43%	4.75%	2.01%	3.01%	1.37%
Theoretical 3-try reject rate	3.36%	1.07%	2.28%	0.77%	3.62%	1.33%	2.21%	0.88%
Reader 1: 1st try reject rate	7.69%	4.29%	5.85%	3.43%	5.82%	3.40%	3.93%	2.45%
Reader 2: 1st try reject rate	19.28%	8.53%	14.92%	6.89%	19.28%	8.53%	14.92%	6.89%
Reject Rates by Thresh = 100								
1st try reject rate	9.43%	5.22%	7.37%	4.33%	8.08%	4.51%	5.92%	3.54%
2nd try reject rate	53.71%	39.72%	48.76%	36.39%	60.41%	47.20%	53.28%	42.28%
3rd try reject rate	65.02%	50.20%	62.95%	48.86%	69.32%	56.63%	65.35%	53.74%
Actual 3-try reject rate	7.64%	3.95%	5.51%	3.06%	7.18%	3.75%	5.01%	2.79%
Theoretical 2-try reject rate	5.06%	2.07%	3.59%	1.58%	4.88%	2.13%	3.15%	1.50%
Theoretical 3-try reject rate	3.29%	1.04%	2.26%	0.77%	3.38%	1.21%	2.06%	0.80%
Reader 1: 1st try reject rate	8.17%	4.79%	6.38%	3.98%	6.30%	3.89%	4.46%	2.99%
Reader 2: 1st try reject rate	19.70%	9.01%	15.40%	7.42%	19.70%	9.01%	15.40%	7.42%

Table 2

The reject rates in Table 2 are all calculated from a subset of the data which does not include any transactions without scores logged (i.e. transactions by reader 3). These transactions are removed so that we can do some analysis which depends on score data, and be able to compare the various results sensibly.

This table is read in the same general manner as Table 1: the first four rows show the optional removal of four subsets of data from the original data set; columns which have their intersections with one of these rows shaded in contain results reflecting the removal of that row's subset. Note that the fourth row is completely shaded to point out the fact that, as was just mentioned, this table is made entirely without transactions missing score data. The fifth row shows the net number of attempts removed by this process, and the sixth shows the size of the remaining data set. As in the last table, several of the columns reflect removal of more than one subset of data. The rows are arranged in an intuitive manner, with data subsets being toggled in and out in the familiar pattern of binary counting. The 8th through 15th rows of the table contain various reject rates calculated for the data set of each column calculated in one manner, while the 17th through 24th rows contain those same reject rates calculated in a second manner.

As in Table 1, the first column of numbers contains the most basic results, from the data set which contains the data sets of all the other columns. In the top block of percentages, reject rates are calculated just as they were for Table 1: "Reject" transactions are counted, and this number is divided by the total number of transactions. Here, the first, second, and third try reject rates are still lower than what was obtained from the histogram, demonstrating that whatever was done to vary individual users' acceptance thresholds from the system default of 100 did in fact end up accepting more users. The entries for reader 1 and reader 2 (reader 3 is, of course, absent), and for actual 3-try, theoretical 2-try, and theoretical 3-try reject rates mean the same thing as they did in Table 1.

The lower block of percentages uses a second method of calculating reject rates: count the number of scores greater than 100 in the data subset, and divide this by the total size of the data subset. For the first, second, and third try reject rates, this produces precisely 100 percent minus the values of the cumulative score histograms (Figure 9) at a threshold score of 100. The other rows in this block are similarly analogous to the corresponding rows in the upper block. It is worth noting that the percentages in this block are generally higher than those in the upper block, because users' acceptance thresholds in many cases were set at other than the default of 100, and this resulted in significantly more accepts at scores greater than 100 than rejects at scores under 100.

The second column of the table shows the results of eliminating all transactions with scores of 2222 from the base data set. Note that we could not legitimately do this in Table 1, since the data set used there included transactions without score data. The result of removing the 2222's is a dramatic decrease in reject rates across the board. We cannot really justify removing these transactions from the data set; but transactions with scores of 2222 are at least questionable, and it is certainly interesting to see such results.

The third column removes from the base data set the six "goats," individuals with particularly bad performance histories, described under Table 1 and Figure 17. The results of this are similar to what we observed in Table 1: a significant drop in the first-try reject rate, along with a moderate drop in the second-try rate, and a slight drop in the third-try rate; the multiple-try reject rates also drop by appropriate amounts. Both readers' first-try reject rates drop by about the same relative amount.

The fifth column (skipping the fourth column for now) removes reader 1's transactions during the "glare" period (1:30 PM - 6:15 PM) described under Table 1 and in Figures 27 and 28. The result is, as in the last table, a significant drop in reader 1's reject rate, which feeds more dramatically into the overall reject rates because this table does not consider data from reader 3.

The other columns remove various combinations of the three types of transactions named above from the base data set. The fourth data column removes both scores of 2222 and the six "goats;" the sixth data column removes 2222's and reader 1 transactions during the "glare" period; the seventh removes the six "goats" and the reader 1 "glare" period; and the eighth data column removes all three: 2222's, "goats," and "glare." In this final column, the first-try reject rate (by "Accept" / "Reject") reaches 3.01 percent, with the theoretical 3-try reject rate dropping under one percent (0.88 percent by "Accept" / "Reject," and interestingly, even lower at 0.80 percent by score threshold of 100); but we have had to remove fully 50 percent of our original data set, or 25 percent of this table's base set of transactions with scores, to achieve this.

6.3.14 Actual vs. Theoretical Reject Rates For All Transactions

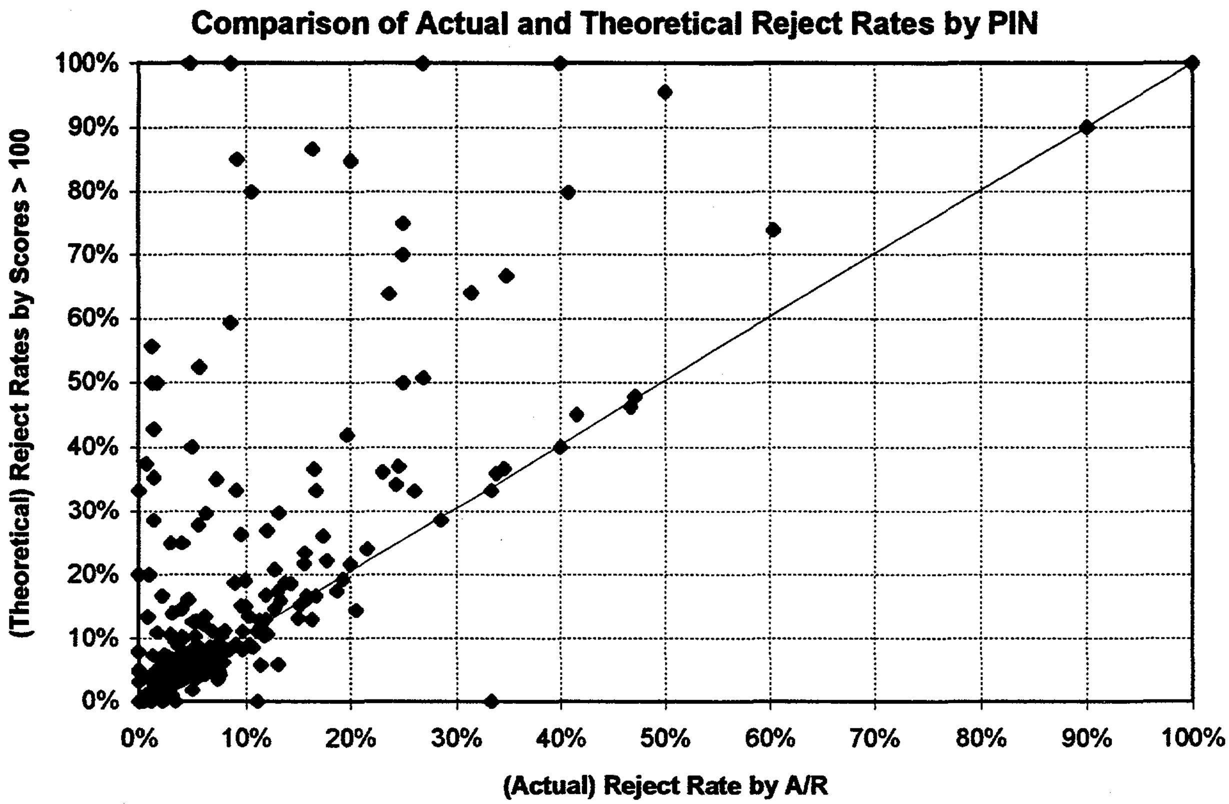


Figure 14

In Table 2, one important thing to notice was that the theoretical reject rates found by counting scores over 100 were generally higher than the actual reject rates found by counting transactions flagged by HandNet as "Rejects." Figures 14 and 15 take a closer look at that phenomenon. Each of the 316 users has one spot marked on the scatter chart, with its horizontal location corresponding to that user's individual "actual" average reject rate (by "Accept" / "Reject"), and its vertical position corresponding to that user's theoretical reject rate (by a score threshold of 100). A line of slope 1 is drawn across the chart, so that any mark falling on this line represents a user with equal actual and theoretical reject rates. So, marks falling to the left of this line represent users with *lower* actual reject rates than theoretical reject rates; such users had, on the whole, easier times getting accepted by the system than they would have if they were always set at the default acceptance threshold of 100. Marks falling to the right of this line represent users with *higher* actual reject rates than theoretical reject rates; such users had, on the whole, more difficult times getting accepted by the system than they would have if they were always set at the default acceptance threshold of 100. It is useful to consider that a user's performance determines the vertical position of that user's mark; while the acceptance threshold(s) chosen for that user determine the horizontal position, with higher (easier) thresholds sliding the mark left of the diagonal line, and lower (more difficult) thresholds sliding the mark right of the line.

Note that this chart includes data from reader 3, which did not report scores, so the vertical position of each mark is calculated from a different data set than is the horizontal position. Figure 15 amends this situation; but this version is included first, just because here the "actual" reject rate axis is taken from the actual, entire data set.

What stands out most in Figure 14 is that the majority of users are left of the diagonal line, and many are far to the left of it, indicating easier acceptance thresholds than the default. The smaller group of users right of the line is gathered mainly close to the horizontal axis, where reject rates are low. This indicates that the users who were kept at individual acceptance thresholds lower (tighter) than the default were the ones who were good at using the HandNet system, which only makes sense. Users further away from the horizontal axis are almost universally left of the line, having assumedly been assigned higher (looser) acceptance thresholds to offset their greater difficulty in using the system. What is more disturbing is that there are quite a few users marked to the left of the line but very close to the horizontal axis; these users have apparently been given easy acceptance thresholds in spite of generally good performance that should not require any such assistance.

6.3.15 Actual vs. Theoretical Reject Rates For Transactions With Scores

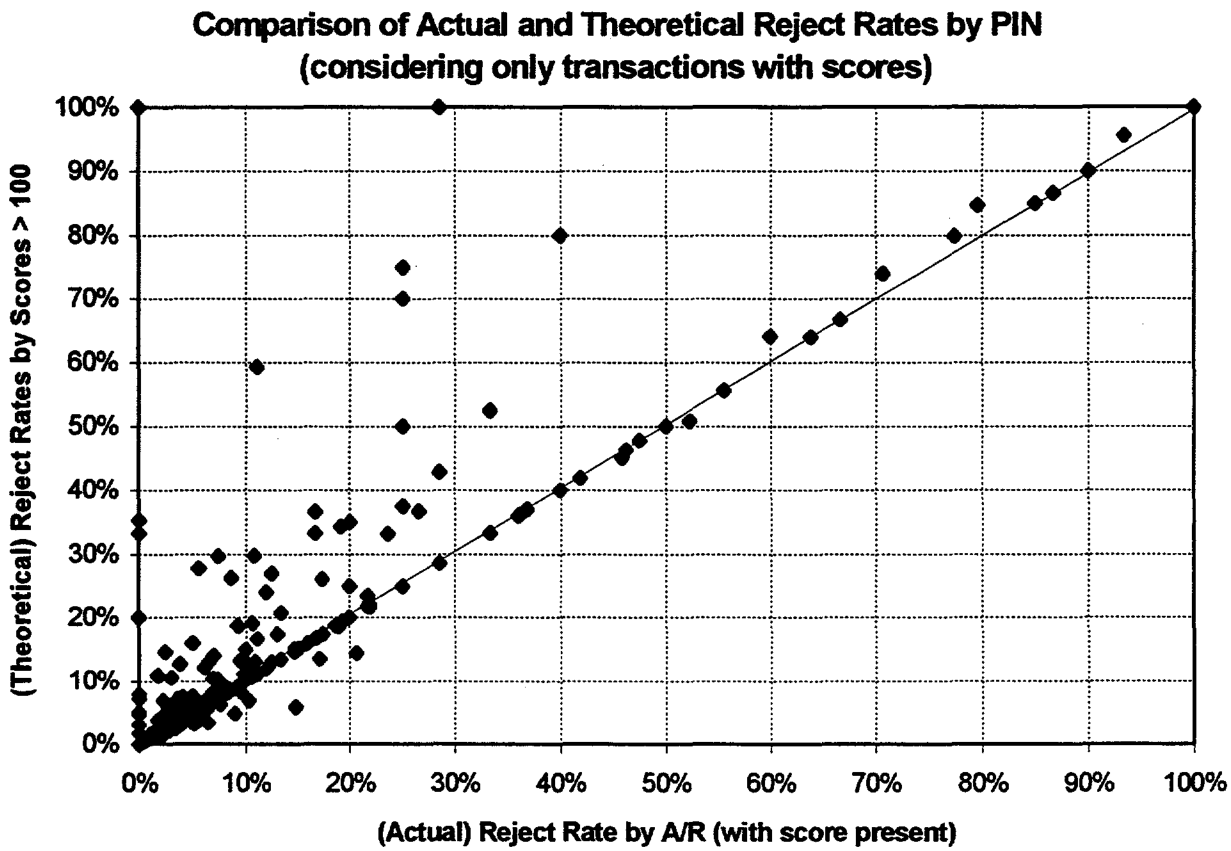


Figure 15

Figure 15 is identical to Figure 14, except that it is generated from data coming only from readers 1 and 2, so that all transactions have score data, and both actual and theoretical reject rates can be calculated from the same data set.

The most noticeable difference in Figure 15 is that substantially more of the marks fall precisely on the diagonal line, indicating users who were kept at the default acceptance threshold of 100. The spread of points is otherwise somewhat tighter about the diagonal line. Still, though, we see a large number of points left of the diagonal line, including quite a few points close to the horizontal axis; and still there is a much smaller set of points to the right of the diagonal line, and mostly close to the horizontal axis. (Refer to the text after Figure 14 for more discussion.)

6.3.16 Histogram of Users by Reject Rate

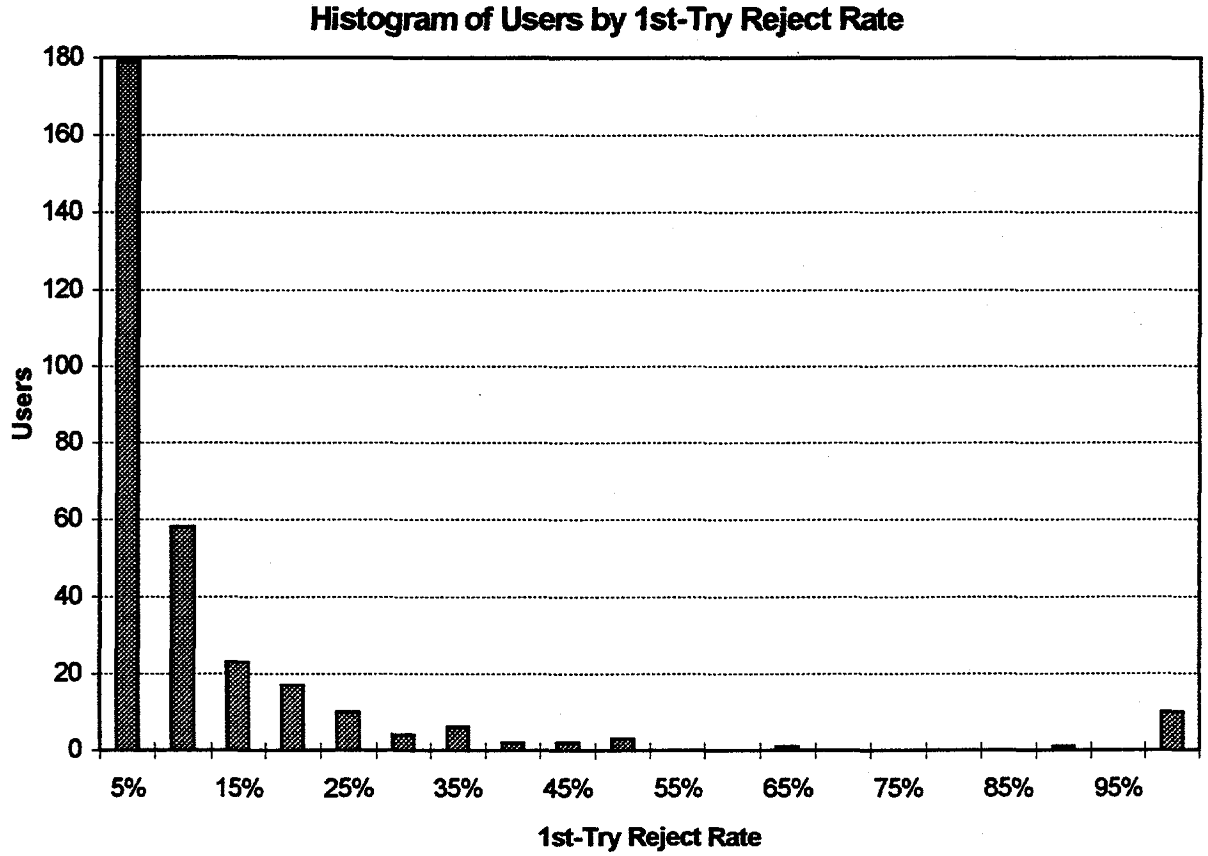


Figure 16

Figure 16 groups users into bins of first-try reject rate (calculated by "Accepts" / "Rejects") five percent wide. Most of the users are, as they should be, in the first bin, and the histogram drops off quickly from there, if not as quickly as it could. The only other interesting feature is the last bin, which rises above the previous ones with ten or so users at 95 percent - 100 percent reject rates. However, since this chart contains no information on how many transactions those users made, not too much can be concluded.

6.3.17 Reject Rate vs. Number of Attempts

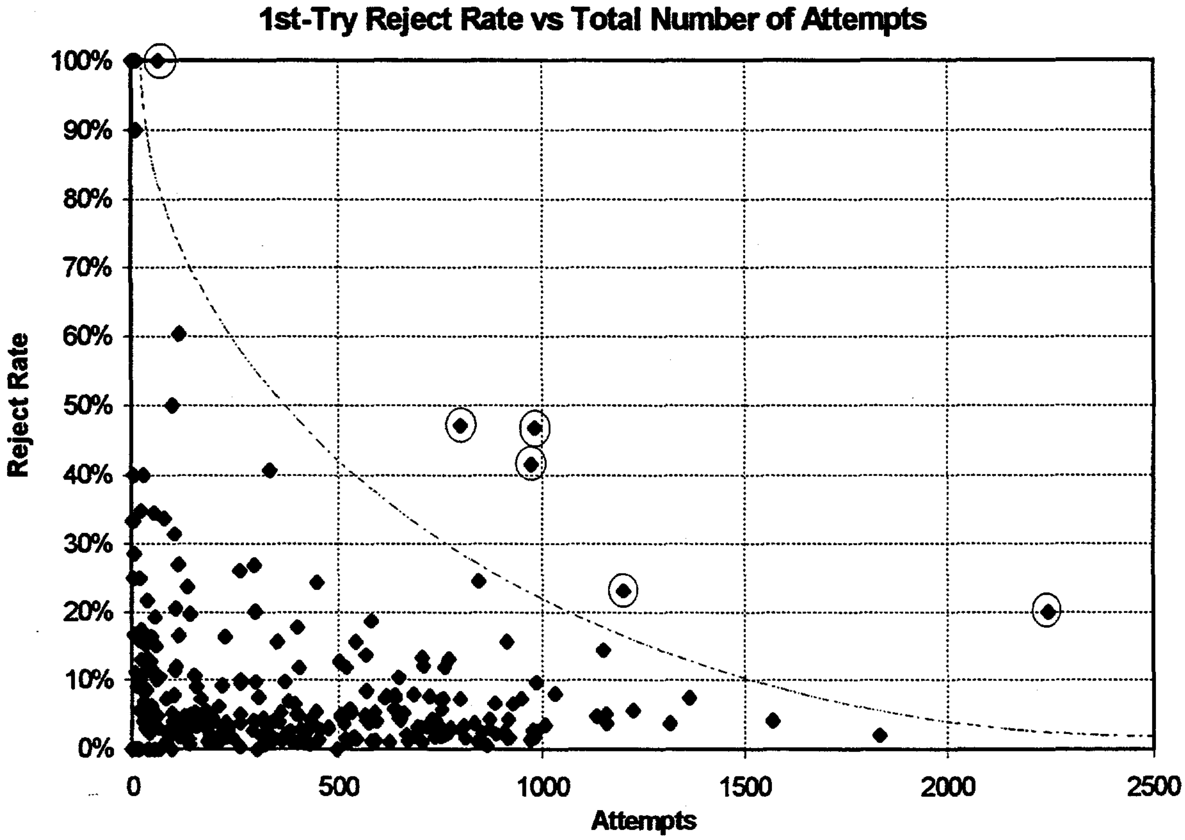


Figure 17

In Figure 17, individual users' first-try reject rates are plotted against the total number of attempts they made to gauge how relevant their reject rates actually are. As we may have expected, all the users that fell into the last bin of the previous graph (i.e., 95 percent - 100 percent first-try reject rates) have made very few attempts, except possibly for one user who consistently failed every one of 66 first-tries.

As with the 2222's vs. attempts scatter chart (Figure 11), it is no surprise to find points plotted far out along either axis, as long as they are actually close to that axis. This includes both users who failed their first attempts a large portion of the time but made few attempts, as well as users who made large numbers of attempts and had a low first-try reject rate. It is, however, surprising to see points plotted toward the interior of the chart. Six points have been circled in Figure 17, selected for their relatively extreme placement, away from the hyperbolic region that seems to define most of the chart. (We have drawn a rather arbitrary ellipse quadrant after the fact as a visual aid.) It is these six users who were selected as the "goats," to test the results from their elimination from the data set, as shown Tables 1 and 2. Thus, the six "goats" are not simply the users with the highest reject rates, but were selected due to various combinations of high reject rates and large numbers of attempts.

6.3.18 Average Reject Rate by Attempt Number

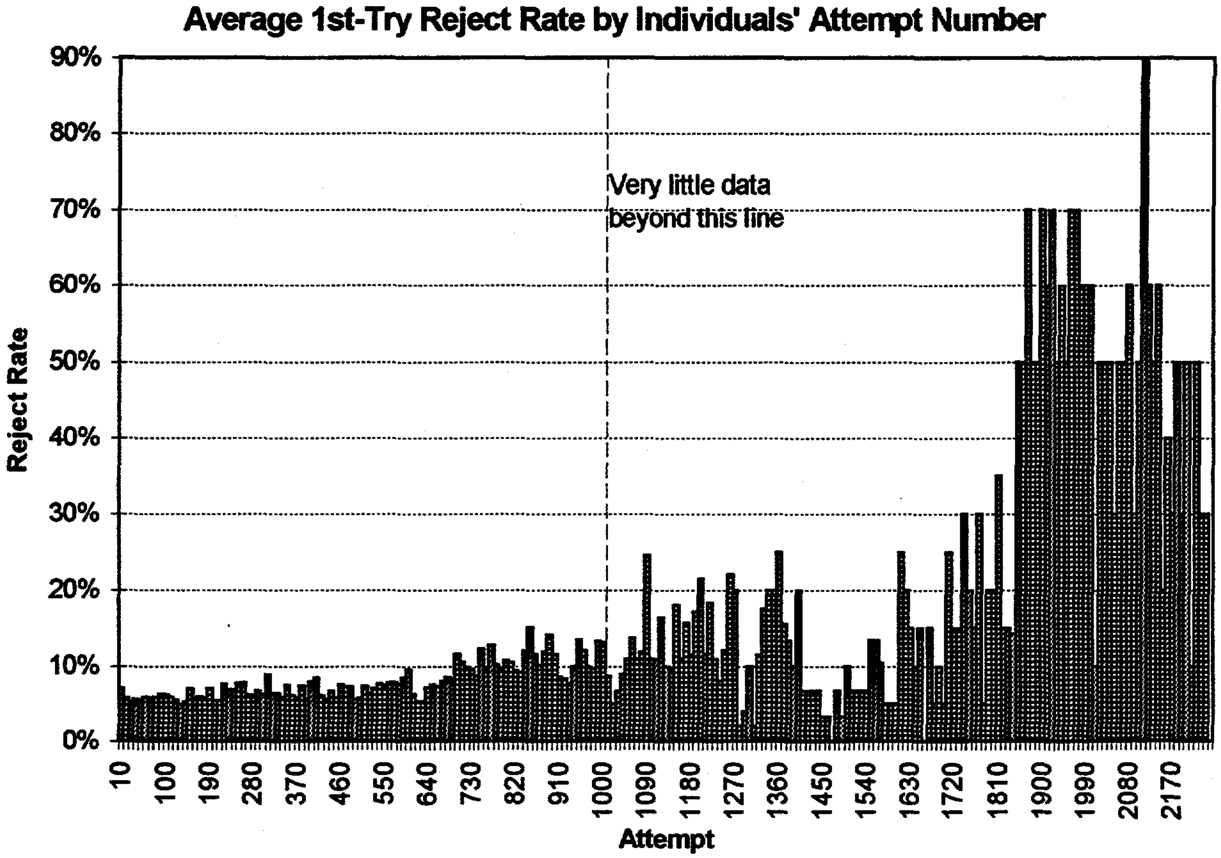


Figure 18

Figure 18 *does not* classify users by their total number of attempts. Rather, it was generated by considering all of the transactions that were, say, numbers 1 through 10 for their respective users, followed by all the transactions that were numbers 11 through 20 for their respective users, and so on. What we had hoped to see here was a "learning curve;" we expected that users would generally do better on their 500th attempts to use the readers than on their 10th such attempt. Now, this analysis does have one major flaw: since our chosen data set starts several months after the initial setup of the HandNet system, transactions which are considered here as, say, 100th attempts, may include the 312th attempt for one user, and the 115th attempt for another, and the 100th attempt of a user enrolled after our data set starts. In spite of this problem, though, one would expect some sort of downward trend in reject rates as users become more experienced attempt by attempt.

Such a downward trend is not what we see here. In fact, even in the left portion of the graph where we have a considerable number of users still participating, there seems to be a slight *upward* trend in reject rates. So, we are failing to detect any kind of positive learning curve by this method. The upward trend could be explained in various ways: the users are actually getting worse attempt by attempt, the readers are degrading in performance perhaps due to lack of maintenance, or the composition of users who make more than, say, 800 attempts is different than that of the users who never made more than, say, 200 (during our data period).

The vertical line in this chart marks a point beyond which there is too little data for the results plotted here to be meaningful. A glance at Figure 7 will reveal that only 13 users made more than 1000 attempts. Arguably, this line could be placed even further left than where we selected to locate it.

6.3.19 Predicted Reject Rates by Threshold

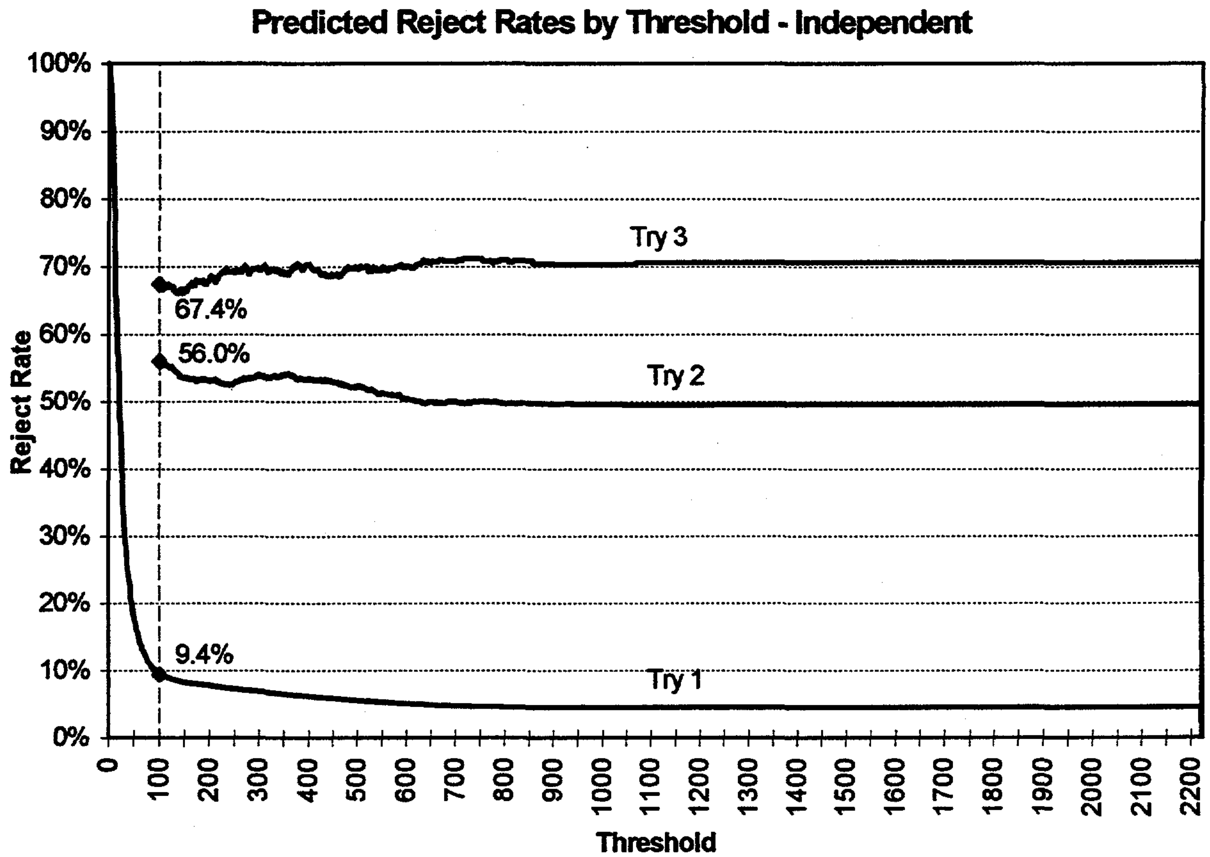


Figure 19

Figure 19 represents "what-if" analysis of the possible threshold settings. In fact, the acceptance threshold was set at many different values for many different users, but it could have been set at a uniform value of 100, or anything else. (The HandNet system may only support thresholds up to 200, but in theory the threshold can be anything.)

One disclaimer must be made here. For purposes of generating second and third try reject rates at various thresholds, we have effectively assumed something which is absolutely false: that every user's threshold was actually set at 100 throughout the test. Without this assumption, no second or third try reject rates can be predicted by this method. A small amount of testing of our data has been done which suggests that the assumption does not produce results too far off the mark; but nevertheless the second and third try results here (and in Figures 20-22) need to be taken with an extra grain of salt.

At each threshold value on the horizontal axis of this chart, the question is asked, "What would the first, second, and third-try reject rates have been had all thresholds been at this value?" (The chart is labeled "independent" because these three reject rates are considered separately.) For a threshold of 100, that question has already been answered in Figure 9 and Table 1; these values (9.4 percent, 56.0 percent, and 67.4 percent) have been labeled on Figure 19.

For a threshold under 100, the predicted first-try error rate is found by counting the first-try scores over that threshold, and dividing by the total number of first-try scores. But for second and third tries, we have a problem: many of the users who would have been rejected on their first try at this threshold, and then perhaps made a second and third try, actually got accepted and went in the door without another try, because their thresholds were actually set at 100 (by our assumption). So, the only data we have is from users who scored badly enough in the first place to be rejected at a threshold of 100 — and this is a very bad subset of data to be using to calculate reject rates. For this reason, no values are plotted on the chart for second and third try reject rates at thresholds under 100.

For a threshold over 100, the predicted first-try error rate is found by the usual method of counting first-try scores over that threshold and dividing by the total number of first-try scores. For the second-try error rate, we have to first drop from consideration all transactions where the user was rejected by the system on his or her first try, but actually scored well enough to be accepted on the first try at this threshold. The remaining second-try scores come from users who would actually have been rejected on their first try at this threshold and then tried again; so we count those of these scores which are over our threshold, and divide by the total number of these scores to obtain the predicted second-try reject rate. For the third-try reject rate, we must similarly discard any transaction sets which would have been accepted on the first or second try at this threshold, and then count third-try scores over this threshold and divide by their total number. Note that because of this process of discarding subsets of data to calculate second and third try reject rates at thresholds over 100, the second and third try graphs are not logically bound to be decreasing (which is good, since they are not.)

As for features of Figure 19, there is very little to note. The first-try graph starts (nearly by definition) at 100 percent at a threshold of 0, passes through 9.4 percent at the default threshold of 100, and asymptotes to 4.4 percent (the percentage of first-try 2222's) until it drops abruptly to zero at a threshold of 2222. The second-try graph starts at 56.0 percent at a threshold of 100, and meanders slightly before approaching a limit reject rate of 49.5 percent until it likewise drops abruptly to zero at a threshold of 2222. The third-try graph starts at 67.4 percent at a threshold of 100, and also meanders slightly before approaching a limit reject rate of 70.6 percent, dropping abruptly to zero at a threshold of 2222.

6.3.20 Detail of Predicted Reject Rates

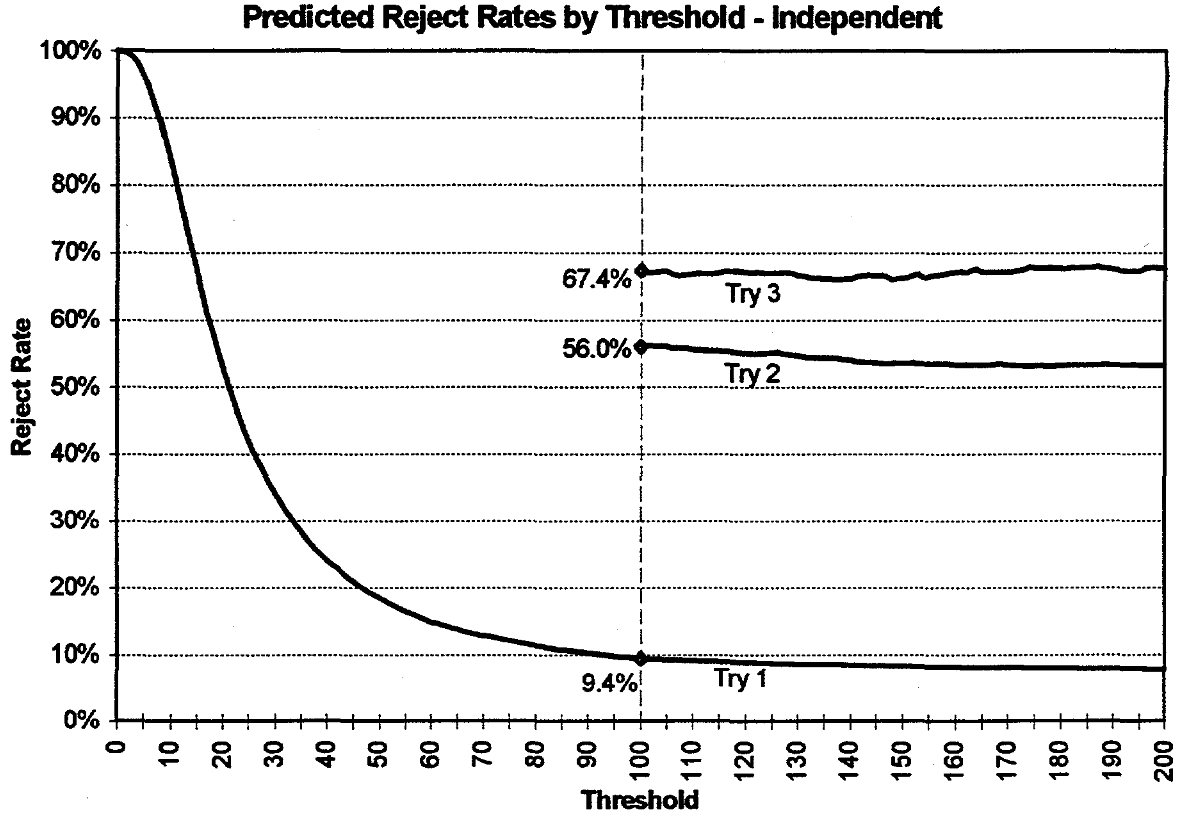


Figure 20

Figure 20 is a simple blow-up of Figure 19. The first-try curve reaches a 7.8 percent reject rate at the threshold of 200 which bounds this chart.

6.3.21 Cumulative Predicted Reject Rates

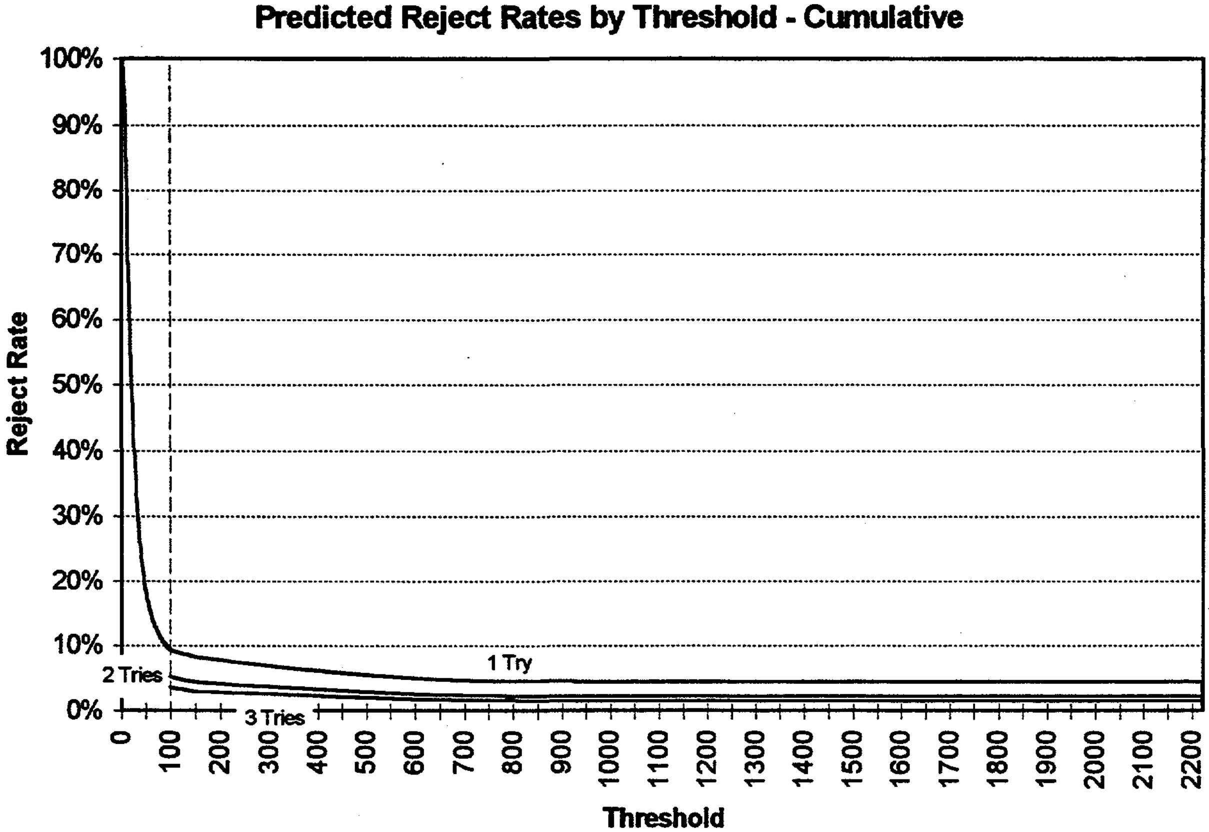


Figure 21

Figure 21 uses the same reject rates generated for the last two "independent" predicted reject rates charts (Figures 19 and 20). The 1-Try curve is, in fact, identical to the 1st-Try curve of Figure 19. The difference is that the 2-Tries curve on this chart represents the predicted odds of being rejected twice in a set of two tries, rather than just the second-try reject rate. It is calculated as the product of the first and second try reject rate curves from Figure 19. Likewise, the 3-Tries curve on this chart represents the predicted odds of being rejected all three times out of a three-try set, and is equal to the product of the first, second, and third try reject rate curves from Figure 19.

The one-try graph starts at 100 percent at a threshold of 0, passes through 9.4 percent at the default threshold of 100, and asymptotes to 4.4 percent until it drops abruptly to zero at a threshold of 2222. The two-try graph starts at 5.3 percent at a threshold of 100, approaches 2.2 percent, and drops to zero at 2222. The three-try graph starts at 3.6 percent at a threshold of 100, approaches 1.6 percent, and drops to zero at 2222.

6.3.22 Detail of Cumulative Predicted Reject Rates

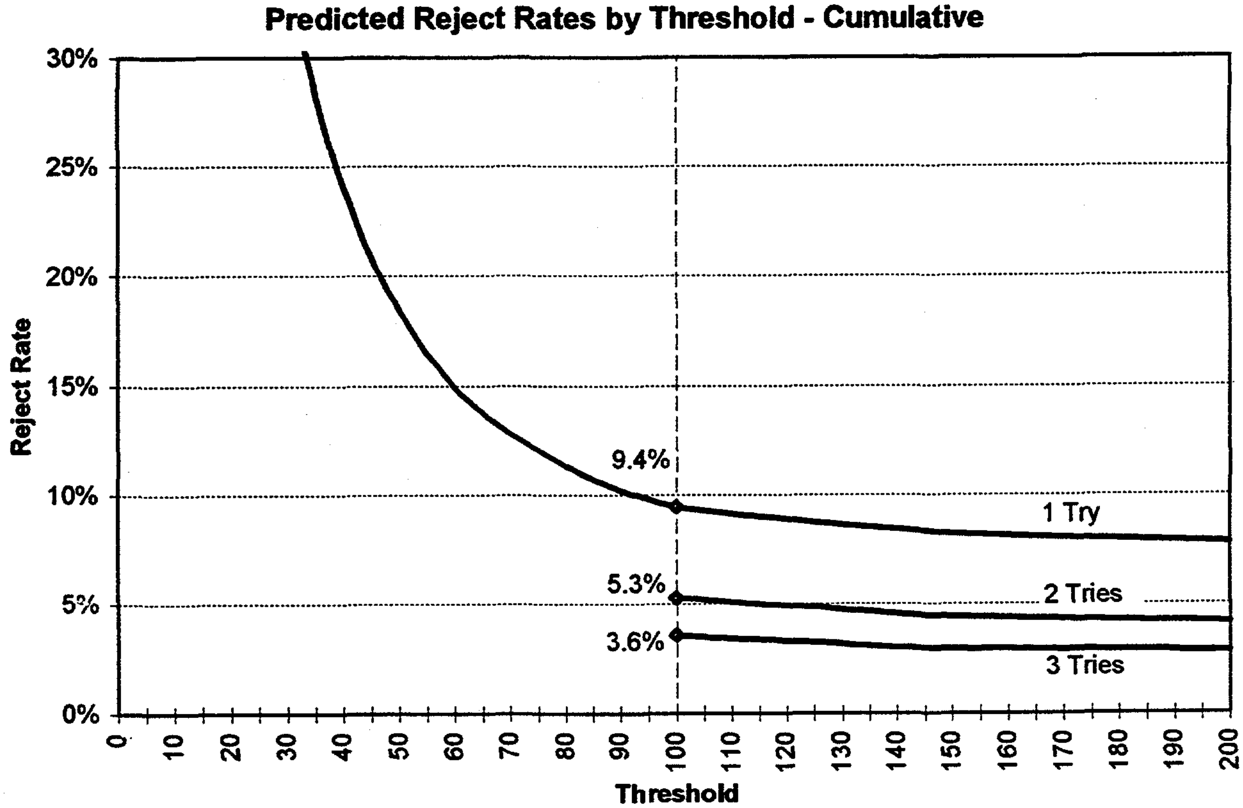


Figure 22

Figure 22 is a simple blow-up of Figure 21. The one, two, and three try curves pass through 7.8 percent, 4.1 percent, and 2.8 percent, respectively, at the right edge of the chart (a threshold of 200.)

6.3.23 First Try Reject Rate by Month

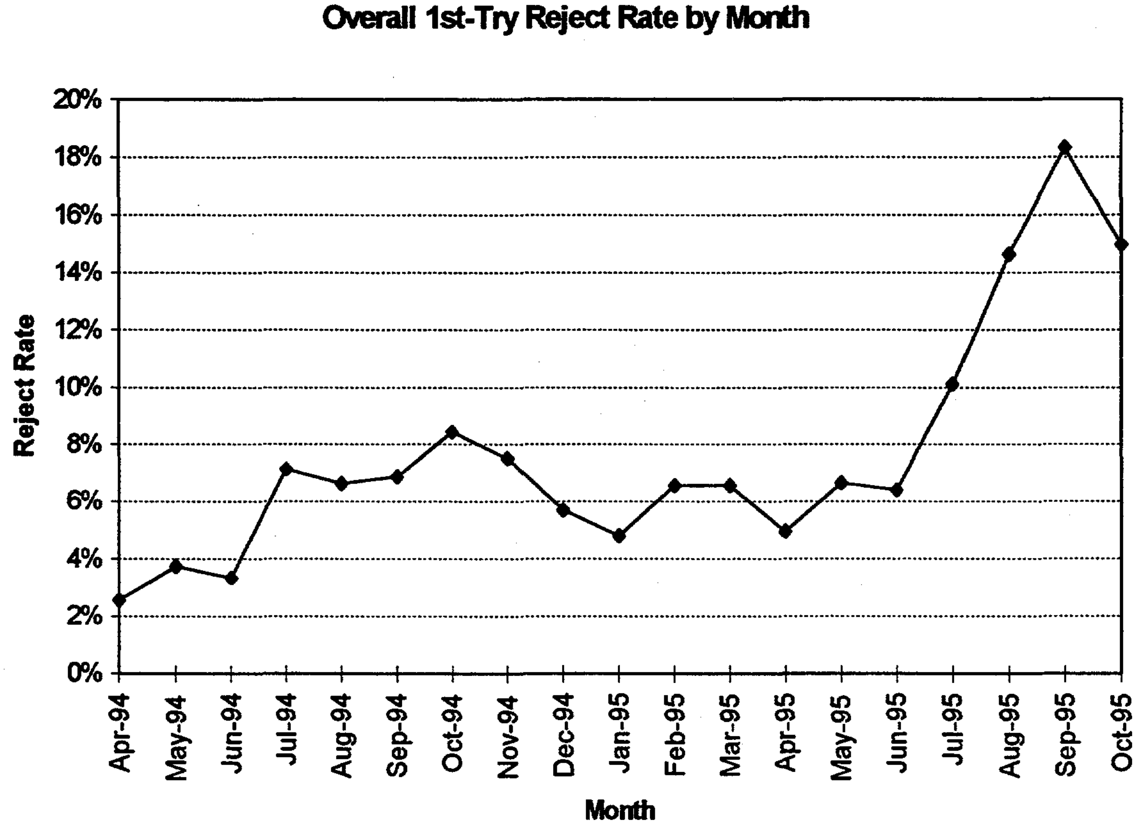


Figure 23

Figure 23 shows, on a scale of months, how the average reject rate evolved over the course of the data period. What must be pointed out here is the obvious: it increases. Each month from May of 1994 through August of 1995 represents on the order of 6,000 transactions (see Figure 3), so this is not random noise; the climb from 3.7 percent in 5/94 to 14.6 percent in 8/95 really happened. In September 1995, the system was starting to be shut down, and reader 1 in particular dropped by about half in transactions logged for the month (see Figure 3), but still there were over 4000 total transactions to lend credibility to the peak of over an 18.3 percent first-try reject rate. Attempts at explanation will be held until after Figure 24.

6.3.24 Reject Rates by Month, by Reader

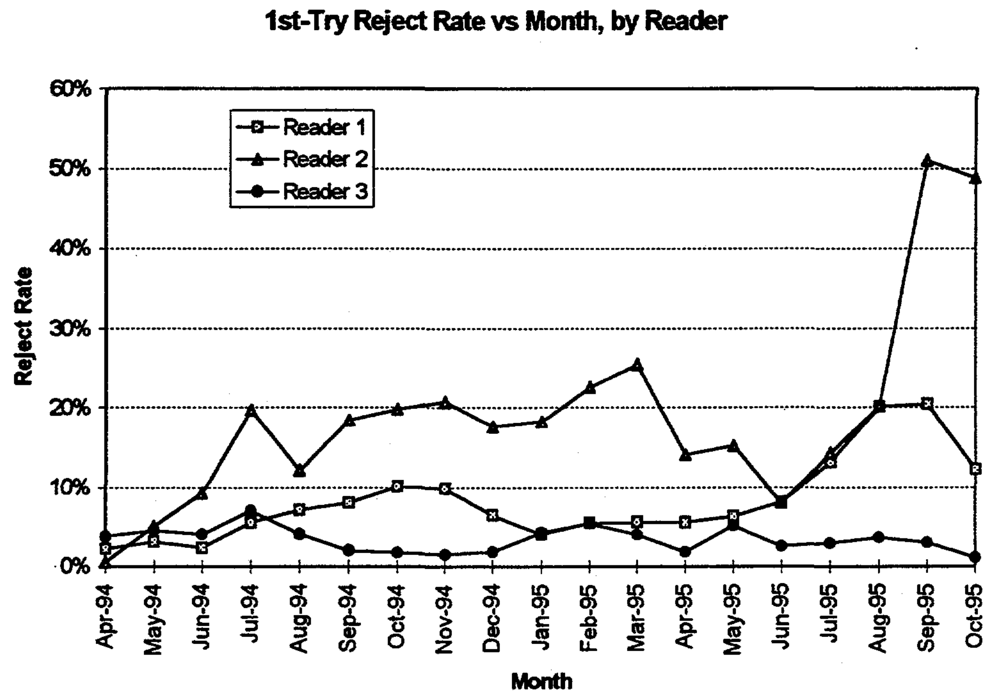


Figure 24

In Figure 24, the monthly first-try reject rates of Figure 23 have been split to show the individual three readers. It initially appears that reader 2 played a major role in pulling reject rates upward, especially in September of 1995. While this must certainly be true, the overall reject rate is not greatly affected by this, since reader 2 contributed only 7 percent of all the transactions (see Figures 1 and 2). Rather, the overall first-try reject rate curve of the previous chart seems to follow closely the behavior of reader 1, which accounted for 60 percent of all transactions (see Figures 1 and 2). Reader 1 begins to develop a high reject rate in July of 1994, easing off a bit around December of 1994, and then starting a climb to alarmingly high reject rates in July of 1995. Use of reader 1 dropped by half in September of 1995 (see Figure 3), but at this point the 51 percent reject rate of reader 2 helps to pull the overall reject rate up to the high of 18.3 percent observed in Figure 23.

It seems clear that there were definite problems with both readers 1 and 2. Interviews with past users of the HandNet system revealed common opinions that reader 1 functioned badly in the direct sunlight it received in the afternoon hours, which had to be blocked by the body in a very awkward position; and reader 2 worked strangely all the time, often having trouble noting that a user's hand was inserted. Nothing else is known about the behavior of reader 2, and perhaps it was simply a faulty reader which should have been repaired or replaced. As for reader 1, Figures 27 and 28 will demonstrate that the problem was, indeed, glare from the sun.

6.3.25 Overall Reject Rate by Week

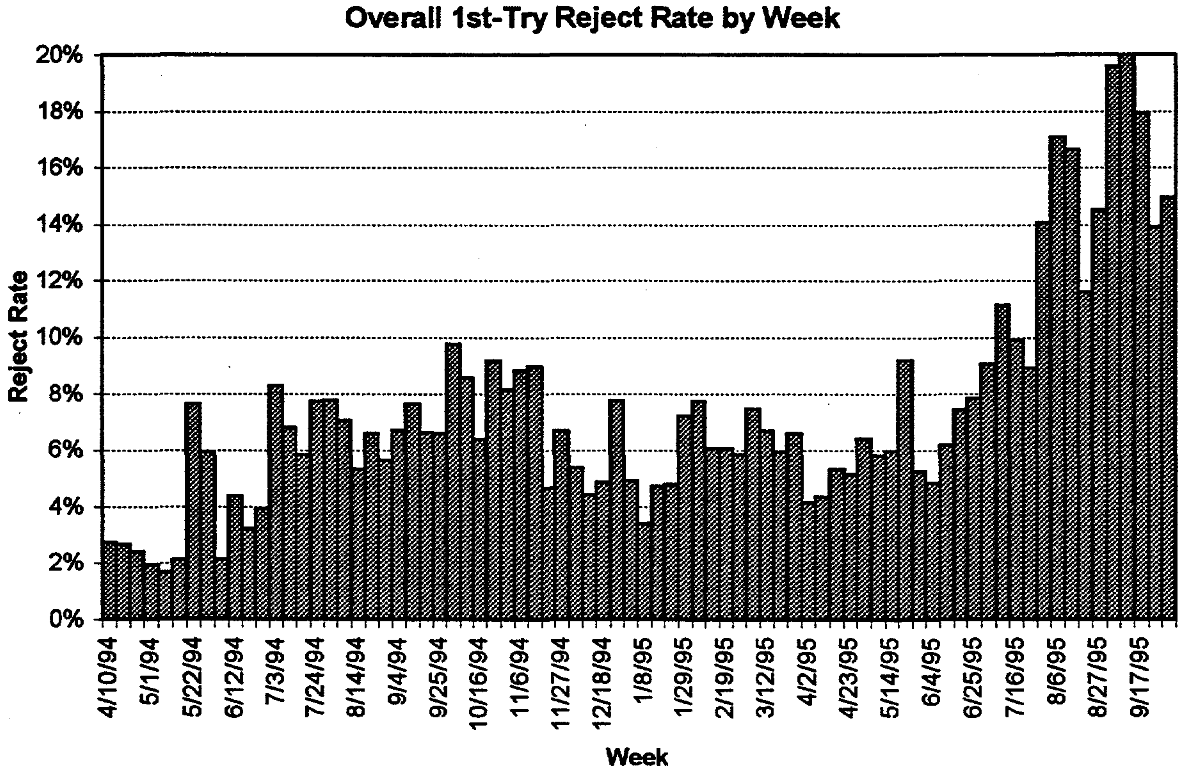


Figure 25

Figure 25 shows the overall first-try reject rate data on a scale of weeks. There is not much further to observe here, except perhaps to notice the present but not too great week-to-week variation as some measure of random noise in the monthly charts (Figures 23 and 24.)

6.3.26 Overall Reject Rate by Date

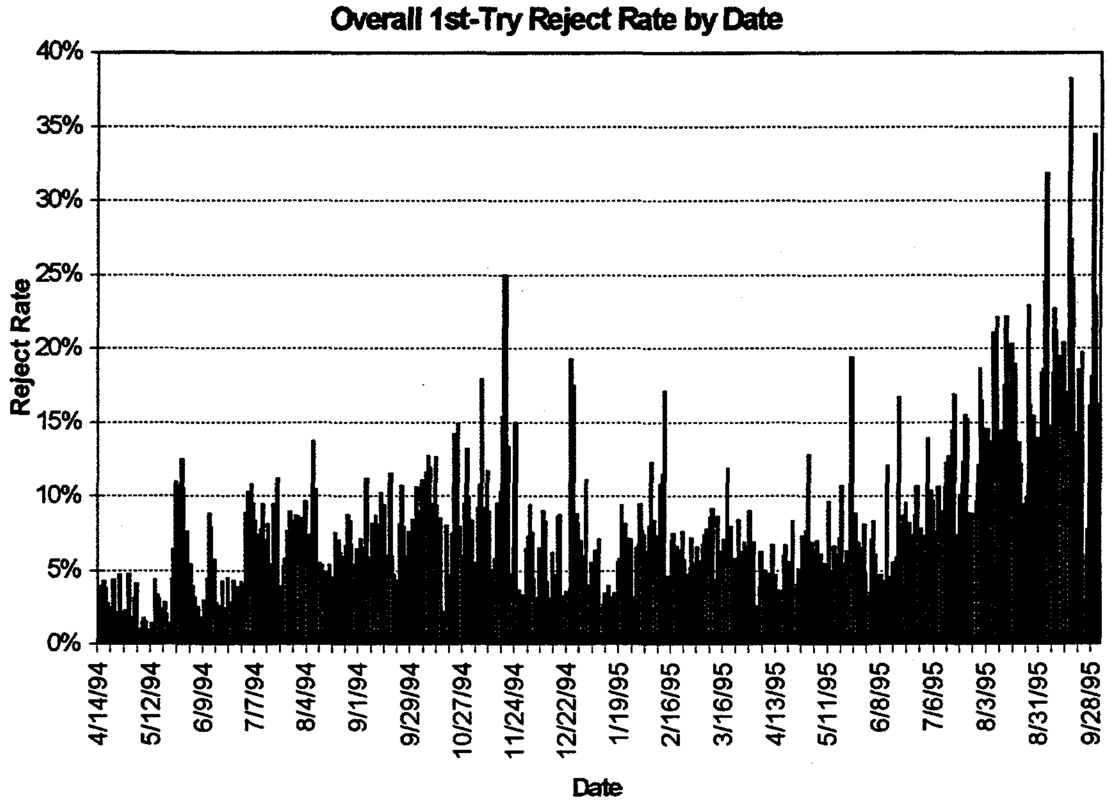


Figure 26

Figure 26 exists only to demonstrate the substantial degree of day-to-day variation in reject rates.

6.3.27 Reader 1 Reject Rates by Time of Day

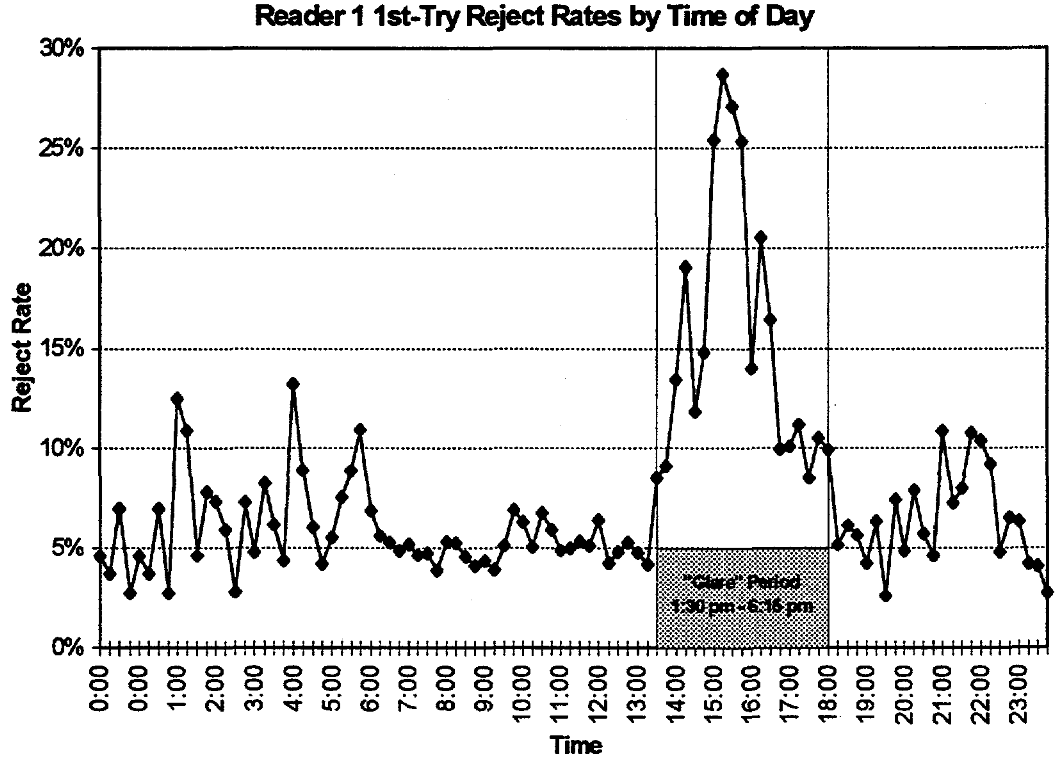


Figure 27

Figure 27 was created after other data and user observations had indicated a problem with direct sunlight on reader 1 during the afternoon hours. The purpose of the analysis is to determine the actual time period during which the reader was affected. There is a pronounced spike in first-try reject rates that extends clearly over the baseline level from the 1:30 PM mark to the 6:00 PM mark, and which reaches as high as 28.7 percent in the interval starting at 3:15 PM. As intervals for this chart are 15 minutes long, the reader 1 "glare" period was taken to be from 1:30 PM to 6:15 PM. The results of removing this data are shown in Tables 1 and 2.

The variations at the start of this graph may be attributed to small amounts of data (see Figure 6.) Interviewees were of the opinion that the reader worked fine when it was dark; and while there are interesting variations here in the late evening hours, the apparent average reject rate in the region does not contradict this claim.

6.3.28 Reader 1 Reject Rates by Month and Hour

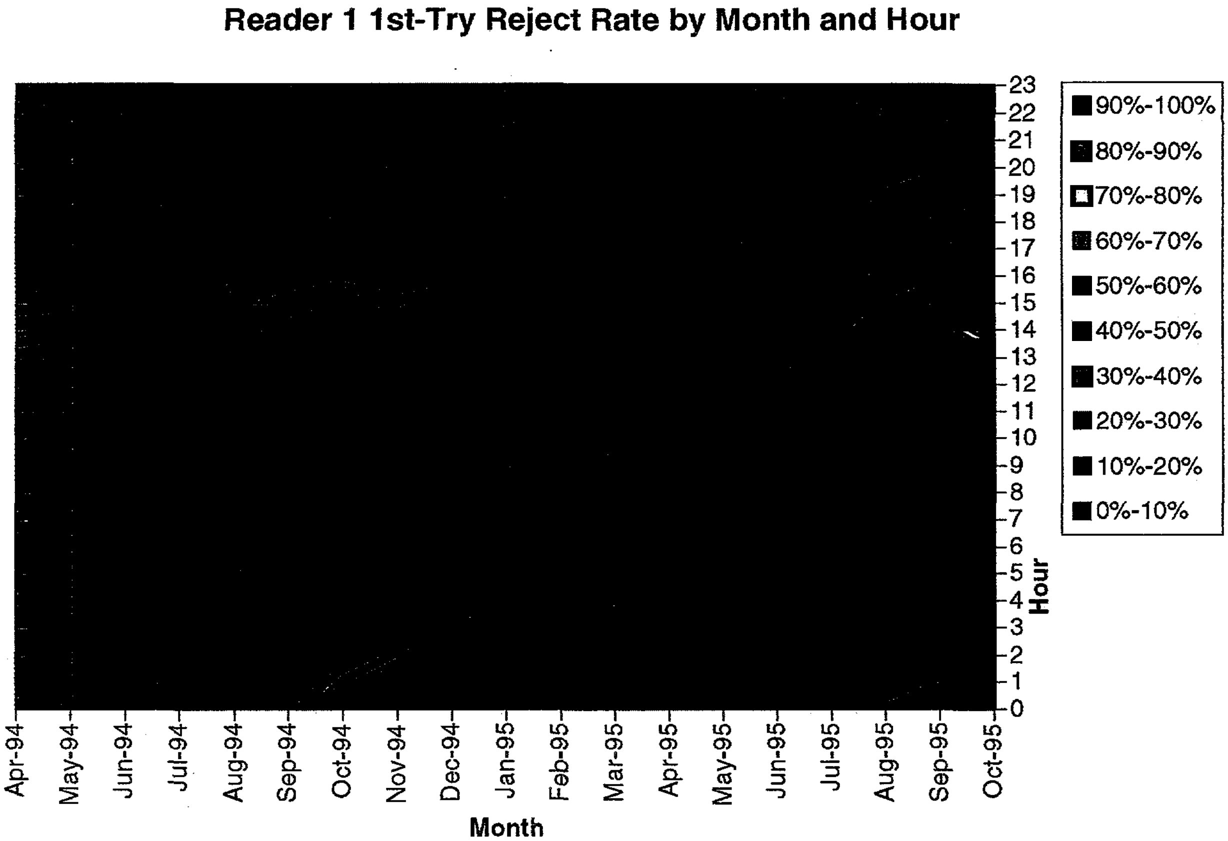


Figure 28

Angle of sunlight is not only a factor of the time of day, but also of the time of year; and certainly other environmental conditions, including temperature, are a function of the season. Thus, Figure 28, a plot of reader 1's average first-try reject rates by month and hour.

Some of what is in Figure 28 is as likely noise as anything else. The feature in October of 1994 at 1 am and November at 2 am, for example, is backed up by only 10 transactions for the first data point and 5 transactions for the second; so basically a couple of guards each had a bad night at the reader. During daylight hours, though, there are enough transactions to make this chart meaningful (see Figure 6.)

There is a very plain strip of high reject rates in the afternoon hours, just as the Figure 27 would predict. What is more interesting, though, is that this strip gets quite a bit worse around October of 1994, and seems ready to do the same thing around October of 1995. (There is actually almost no data to rely on in October of 1995, but the previous month sets up the trend well enough.) So, we may guess, at least, that the sun was in the worst position in the fall. You can even pick out a lesser pair of peaks around April of 1994 and 1995, when the sun, in theory, should have gone through the same angles.

We do not know whether the HandNet host computer was set to and from daylight-savings time, so the apparent waving of this horizontal strip up and down could be due to the change in the sun's angle month to month, or due to reset clocks, or both, or neither. The building 956 receptionist is no longer available to determine how the computer's clock was set.

There is also a major vertical strip of high reject rates in July through September of 1995. Figure 24 already displayed a peak in reader 1's reject rates during this period; but Figure 28 additionally shows that the problem was all day, not just in the afternoon "glare" period. This begs the question of whether reader 1 was starting to degrade in some major way toward the end of the test.

6.3.28 Reader 2 Reject Rates by Month and Hour

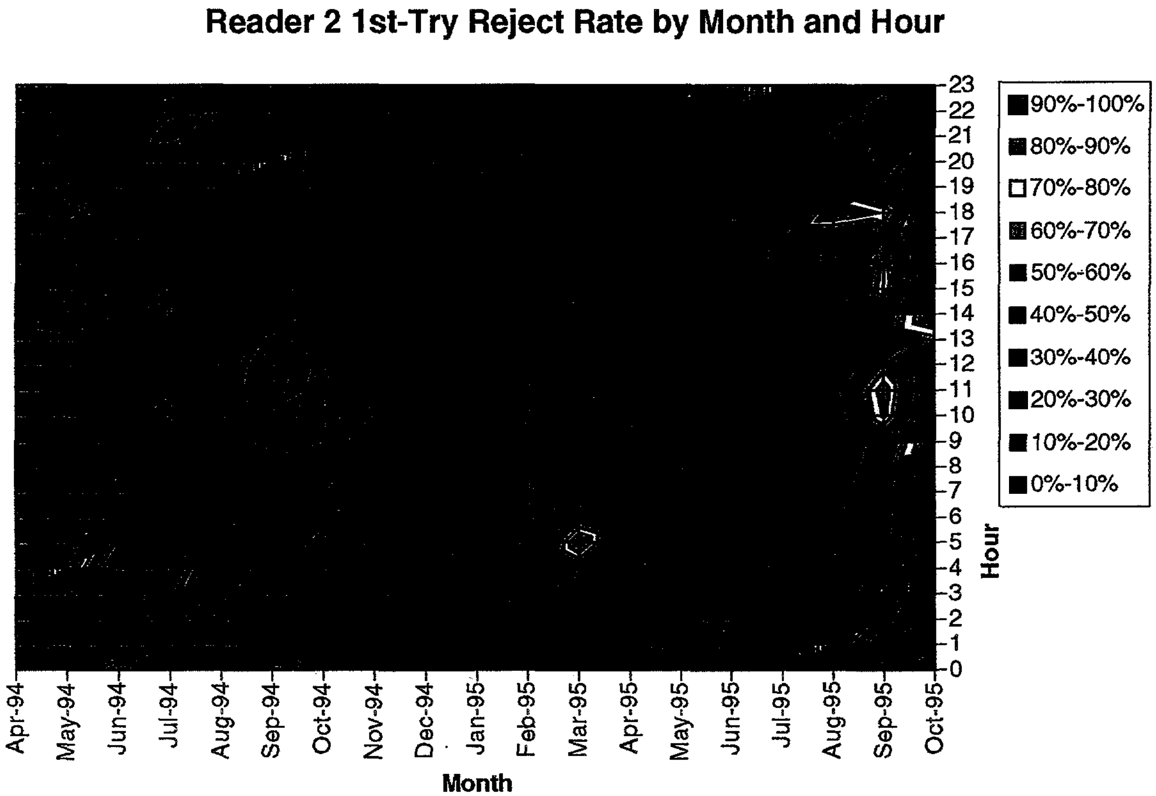


Figure 29

Figure 29 is the same style of chart as Figure 28, only for reader 2. Reader 2, however, did not receive enough transactions to make this view very meaningful. (Large regions of what appear to be 0 -10 percent reject rates are actually due to absolutely no transactions.) Basically we just see a lot of noise overlaid on a generally bad spread of reject rates. We can see the huge leap in reject rates in September of 1995 that we observed in Figure 24, but that is about it.

6.3.29 Reader 3 Reject Rates by Month and Hour

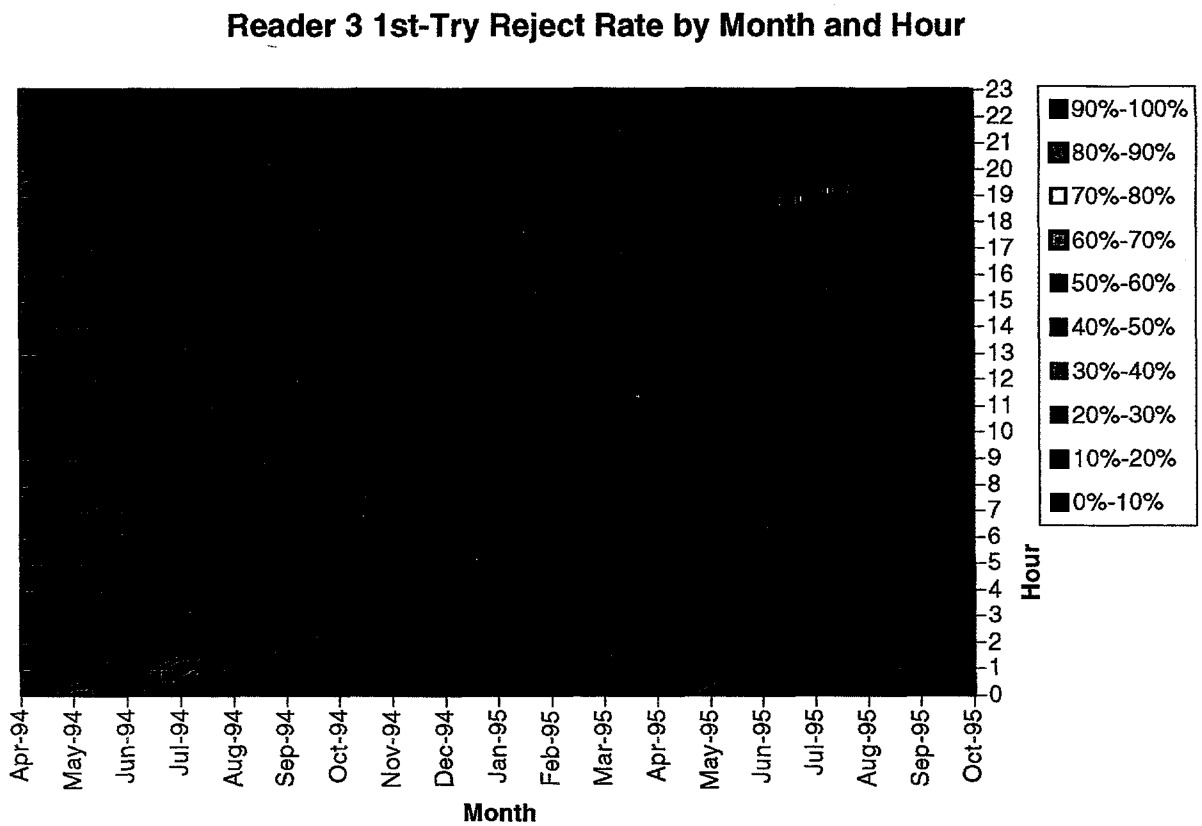


Figure 30

Figure 30 is just like the last two charts, but for reader 3. The only features present here are almost definitely from random noise amid too few transactions. Other than that, reader 3 is here shown to be a very uniformly pleasant reader, which may be due in part to its entirely indoor location.

7.0 Conclusions

There are some recommendations and conclusions that can be drawn from operating the HandKey units for almost two years.

The HandKey needs to be protected from direct, bright sunlight as stated in the manufacturer's literature. Reader 1 was mounted on a west facing wall and was difficult for users to operate properly during late afternoon sun. Users complained that the display was difficult to see and that the device rejected them more frequently during late afternoon sun. This was substantiated by the data as well as by reports from users.

The HandKey does need some maintenance, a substantial dust storm caused performance to be severely degraded during one week of the first year. The performance was so bad that users called to complain and the devices had to be cleaned.

We feel that users who install the HandKey would be well advised to purchase the option so that scores are reported. Even though this is not a standard feature, it seems that some simple monthly or quarterly analysis of the scores can help identify people who are having problems with the units, may identify the need for maintenance and cleaning, would allow maintenance personnel to investigate problems when they are reported, and would probably increase user satisfaction. With the degradation in performance near the end of the test period, we would expect users to become dissatisfied with the system and begin to complain. Probably, this did not happen because they knew that the system was soon to be replaced.

8.0 References

1. SNL Report, Report of Performance Testing and Evaluation of Eyedentify Identity Verifier, Russell L. Maxwell, Murray Silverman, April 12, 1984.

2. SAND87-0977, Russell L. Maxwell, Larry J. Wright, A Performance Evaluation of Personnel Identity Verifiers, July 1987.

3. SNL Report, Identity Verifier Performance, Russell L. Maxwell, October 14, 1987.

4. Informal Report, A Performance Evaluation of Biometric Identification Devices, James P. Holmes, Russell L. Maxwell, Larry J. Wright, July 1990.

5. Memo - "Additional Information on the Biometric Test Results", James P. Holmes, August 2, 1990.

6. SAND91-0276, A Performance Evaluation of Biometric Identification Devices, James P. Holmes, Larry J. Wright, Russell L. Maxwell, June 1991.

7. SAND96-1033, Laboratory Evaluation of the IriScan Prototype Biometric Identifier, Frank Bouchier, Janet S. Ahrens, Grant Wells, April 1996.

8. INMM Presentation, Entry Control Technology Biometric Field Evaluations, J. R. Rodriguez, J. S. Ahrens, D. L. Lowe, summer 1994.

Distribution:

1	MS 9018	Central Technical Files, 8940-2
5	0899	Technical Library, 4414
2	0619	Review & Approval Desk, 12690 For DOE/OSTI
1	0769	D. S. Miyoshi, 5800
3	0762	Safeguards and Security Library, 5807
24	0782	Contraband Detection Technologies, 5848