CONF-970684--2 SAND 97-0969 C

BIOMETRIC IDENTIFICATION DEVICES – LABORATORY TESTING VS. REAL LIFE

Janet Steele Ahrens, Senior Member of Technical Staff

Sandia National Laboratories, Department 5848, Albuquerque, New Mexico, 87185

PO Box 5800, Mail Stop 0782, Phone:(505) 844-8877, FAX:(505) 844-5569, e-mail: jsahren@sandia.gov

ABSTRACT

For over fifteen years Sandia National Laboratories has been involved in laboratory testing of biometric identification devices. Tests were conducted to verify manufacturer's performance claims, to determine strengths/weaknesses of devices, and to determine devices that meet the U.S. Department of Energy's needs. However, during recent field installation, significantly different performance was observed than was predicted by laboratory tests. Although most people using the device believed it operated adequately, the performance observed was over an order of magnitude worse than predicted. The search for reasons behind this gap between the predicted and the actual performance has revealed many possible contributing factors. As engineers, the most valuable lesson to be learned from this experience is the value of scientists and engineers with 1) common sense, 2) knowledge of human behavior, 3) the ability to observe the real world, and 4) the capability to realize the significant differences between controlled experiments and actual installations.

Introduction

Since 1984 Sandia National Laboratories research organizations have been involved in the laboratory testing of biometric identification devices. Testing has been conducted for the U. S. Department of Energy (DOE) to evaluate which devices meet the unique needs of the DOE. Multiple reports have been published and much valuable information has been gained from testing of these devices in the laboratory. During the past years the Recognition System, Inc. Hand Geometry Identification Unit, the ID3D model, has become the single unit which has been used in almost all of the separate tests to allow comparison of data between tests.

Historically, Sandia's operational security groups have not seen a need to install biometrics devices. However, during Fiscal Year 93 Sandia's security group requested the assistance of the research group in determining user acceptance of a biometric identifier by installing the devices at one building. Three HandKey units were installed and operated from September 1993 to October 1995. This paper discusses the data collected and the lessons learned during this field operation of the equipment.

Biometrics Background

One of today's challenges in security work is the proper identification of authorized personnel. All types of businesses - from grocery stores to super-secret government installations need to assure that workers are admitted quickly and that potential intruders are excluded from the facility. While badges, name tags, receptionists, guards, electronic tags, and combination locks have served this purpose for years; a more cost effective, more reliable, less easily defeated system is continually sought. The key concept of biometric identification devices is the ability for the system to identify some unique aspect of the individual rather than some object a person may be carrying or some password they are required to know.

The technology of biometric identification has been active (to various degrees) in the past twenty years. Technology has been developed to identify people based on many different personal characteristics including voice, handwriting, facial characteristics, eye retina blood vessel patterns, and iris patterns. One of the great challenges to a designer of a security system is to determine whether or not to use a biometric identification system, and to select the one most appropriate for the particular facility.

Prior Work by Sandia

During 1987, 1990 and 1991, Sandia was tasked by the DOE to test various biometric devices (see References). These tests were performed in the laboratory at Sandia and the tests were designed to provide statistically significant data. The tests were planned, and large amounts of data were collected and analyzed by highly qualified mathematicians. Test subjects were solicited from Sandia's employees and were encouraged to use each device several times per day. Early tests were very tightly controlled, with a test observer present to assure that the test subjects provided the proper input to each machine. Carefully controlled

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.

DISCLAIMER

attempts were made to test whether any test subject could pass as another test subject.

Several measures of devices have been developed in order to make comparisons. The obvious features for comparison are cost, size, number of people that can be enrolled, time required to enroll people, time required to use the device, and similar measures used when engineers compare any two technical devices. One of the key performance features developed during this time was the "equal error crossover rate", as illustrated in Figure 1.

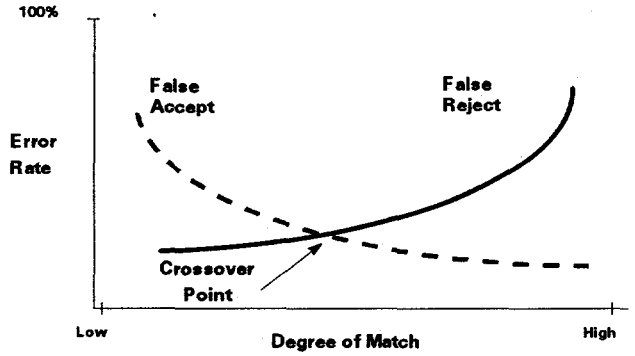

Figure 1

Many of the biometric devices have a threshold setting that determines the degree of "match" necessary for the person requesting identification to be accepted as matching the data base template. This setting is shown as the x-axis in Figure 1. If the equipment is set to require a high degree of match (i.e. to the right of the figure), it is very unlikely that the an unauthorized person will ever be erroneously identified as an authorized person. This is referred to as a "false accept" and is shown by the dashed line. If the device is set to be very tolerant, it is unlikely that any authorized person will ever fail to be identified by the system. This is referred to as a "false reject", characterized by the solid line. While the exact names of the these two curves can differ, many vendors use "equal error crossover point" to measure the accuracy of the device.

Testing of the Recognition Systems ID3D HandKey system is documented in References 1-4. As a result of the various laboratory tests, the curves for the HandKey predict a 0.2% error rate.

Field Testing

During 1993, Sandia's operational security organization approached the research organization to install biometric equipment at an operational facility. Based on previous laboratory testing, cost and ease of installation, the Recognition Systems HandKey ID3D was selected, as shown in Figure 2.

Figure 2

In September 1993, three HandKey devices were installed at a Sandia office building. Approximately 250 people required frequent access to the building. The building contained offices, conference rooms, lockers and equipment for the Sandia Protective Force Officers (SPOs). Three main entrances provided access to the building 24 hours a day, in addition to two doors which were only for emergency exit. Two of the entrances were exterior doors, however the west side of the building contained a gym accessible 24 hours a day from the outside, so the actual building entrance on the west side was an interior door. Three HandKey units were installed as shown in Figure 3 for an operational field test.

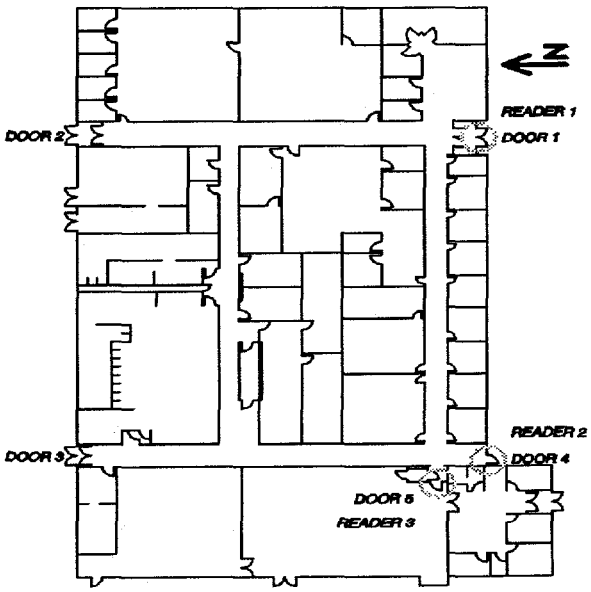

Figure 3

This equipment operated until the summer of 1995 when it was replaced with another prototype system. During the operation of the HandKey devices, no routine maintenance was performed, and no periodic testing was performed. One of the evaluation goals was to determine how much maintenance would be required. Essentially the system only received maintenance when the users complained that it was not operating properly.

Data Analysis

During the past several years, two studies were performed to analyze data from the HandKey field installation. Detailed data analysis is included in References 5 and 8. Reference 5 analyzed data from the test building for a period of 28 days including a total of 6000 transactions. Reference 8 includes data analysis for a total of 537 days of operation, including 117,213 transactions.

Some of the analysis results are exactly what would be expected as a result of installing the system outdoors. In fact, the manufacturer recommends that the device not be installed where it can be affected by direct sunlight and both the analysis and user observations support the fact that reader 1, which faced west, had a higher error rate during the times the setting sun was causing reflections. A good systems engineer should be very careful to follow the manufacturer's recommendations.

The analysis in Reference 5 showed a really significant increase in the error rates during a particular week. In fact, that week was the only week during which users called to complain that the system just was not working. Albuquerque had experienced a dust storm and the platen was covered with dust so that the device was unable to get clear measurement of the hand being presented. Reference 8 showed a degradation of performance over the last couple months of system operation. In an operational system, maintenance should be performed and provisions made for cleaning after bad storms before the users complain.

However, other data did not fit with our expectations. For instance, the error rates from Reference 5 are summarized in Table 1.

Test	1st Try Reject Rate	2nd Try Reject Rate	3rd Try Reject Rate
lab	1%	na%	0.1%
field	2%	43%	58%

Table 1

This points out a significant difference between the two types of testing. It is extremely important that engineers designing systems understand the difference between what happens in a carefully controlled situation and what happens when real people use the device in an actual application.

We believe we understand why there is a difference between the numbers for the 2nd and 3rd attempts. In laboratory tests, people who are not recognized by the device are usually more careful and they tend to place their hand more carefully 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. Since all the SPOs possessed keys to the building, 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 people did attempt a second or third try, it is quite 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.

Table 2 summarizes the results of this analysis from Reference 8. Depending on the data included in the analysis, the field data can yield different results.

Test	1st Try Reject Rate	2nd Try Reject Rate	3rd Try Reject Rate
lab	1%	na%	0.1%
field	4.7-9.4%	49-54%	64-69%

Table 2

Again we suspect that some users also possessed keys to the building and that others simply followed friends into the building. In an attempt to support this concept we studied the number of attempts to use the system by enrolled individuals.

Figure 4 includes a dot for each person using the device. To understand this representation of the data, it is important to know that a 2222 score represents a gross mismatch between the hand presented by the user and that stored in the template. The y-axis of the dot represents the number of times the person used the HandKey, the x-axis represents the number of times that individual received a score of 2222. Interestingly enough, this representation shows that a couple people used the device almost 1000 times and received a score of 2222, 25-30% of the time. In addition, a couple of people didn't use the device much but received scores of 2222 on every single use.

At first we could not understand why these people had not complained about the operation of the

system. We attempted to interview a few of them and they believed the system was operating well. We now believe that the people with 30% bad scores probably had a PIN which was very close to the PIN of several other people and these errors may be the result of miskeyed PINs. The documentation regarding user's of the system did not allow us to identify the people with 100%failure rates. It may well be that these were people who were enrolled, tried the system, got frustrated and simply used keys or tailgated.

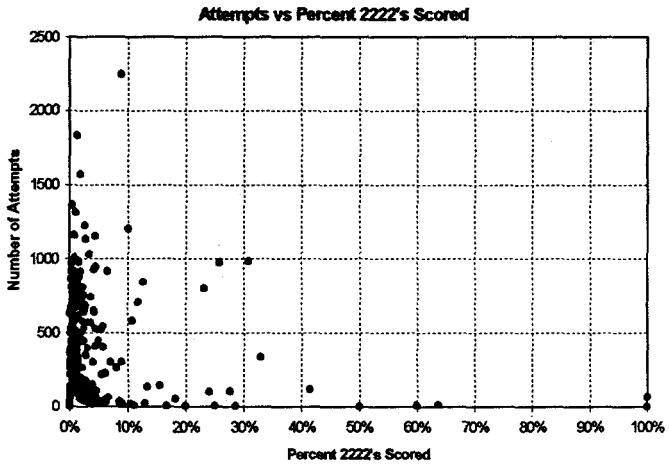

Figure 4

For the design engineer and the installer of equipment, the lesson to be learned from this particular analysis is that people will not necessarily be interested in your project and/or willing to assist you. People will make mistakes, people will try to circumvent the system. If you really want to know what is happening and what people think, you need to design the system to collect adequate data and try to include the people in the planning so that they understand the importance of reporting problems. In retrospect, if we really wanted to determine the false reject rate, we should have somehow verified the PIN (by use of a card reader instead or in addition to the PIN, or by video taping, or by observing). What we really studied was a combination of the error rate of a PIN entry process and the biometric process.

Another key observation for the engineer attempting to install a biometric identification device is the fact that the field installation is not capable of providing the data to substantiate the false accept rate for the device. This would be the number of people who were incorrectly identified as an authorized person, when in fact they were not actually that person. If this had happened in our field test, it would have been the result of one of two circumstances. If a person entered the wrong PIN and his hand closely matched that of the person who actually had that PIN, he would have been admitted and would never have realized he entered the wrong PIN. Our data analysis would show a successful entry for the person with the second PIN. The other situation would have been an actual intruder who guessed someone's PIN and whose hand was a close match. Based on the lab data, there is a slight chance this could happen. Of course, the intruder would never tell us, and our only chance of discovering it would be if the intruder were caught. Since this never happened, we are forced to say that the analysis shows a false accept error rate of 0%, while in reality there is no way of knowing.

Conclusions

The major conclusion to be drawn from this field experience it that the laboratory is not like the field, and the security engineer who is planning to install a biometric identification device would do well to not only carefully study the manufacturer's literature, but also be sure to understand the particular population and facility where the device is to be installed. It is important to not install the device and forget it, maintenance, user surveys, periodic analysis of data and periodic controlled attempts to test the system are key to assuring that the expected level of security is in fact provided by the biometric device.

References

1. SAND87-0977, *A Performance Evaluation of Personnel Identity Verifiers*, R. Maxwell, L. Wright, July 1987.
2. SNL Report, *Identity Verifier Performance*, Russell L. Maxwell, October 14, 1987.
3. Informal Report, *A Performance Evaluation of Biometric Identification Devices*, James Holmes, Russell Maxwell, Larry Wright, July 1990.
4. SAND91-0276, *A Performance Evaluation of Biometric Identification Devices*, James Holmes, Larry Wright, Russell Maxwell, June 1991.
5. INMM Presentation, *Entry Control Technology Biometric Field Evaluations*, J. R. Rodriguez, J. S. Ahrens, D. L. Lowe, summer 1994.
6. SAND97-0614, Hand Geometry Field Application Data Analysis, M. Ruehle, J. Ahrens,March 1997.

Biography: Janet Steele Ahrens graduated from New Mexico State University (BSEE), and Stanford (MSEE). Projects at Sandia National Laboratories have included providing a Sandia access control system, testing biometrics identification devices, developing a Sandia connection to the DOE badge database, developing a prototype security system for the FAA, and heading Sandia's involvement in testing W80 Cruise Missiles.