Password Security: An Empirical Study

MOSHE ZVIRAN AND WILLIAM J. HAGA

MOSHE ZVIRAN is Senior Lecturer of Information Systems at the Faculty of Management, The Leon Recanati Graduate School of Business Administration, Tel Aviv University. He received his B.Sc in mathematics and computer science and M.Sc and Ph.D in information systems from Tel Aviv University. He has held academic positions at Claremont Graduate University, Claremont, California, and the Naval Postgraduate School, Monterey, California.

Dr. Zviran's research interests include information systems planning, development and management of information systems, information systems security, and information systems in health care and medicine. His published works have appeared in *Journal of Management Information Systems*, *MIS Quarterly*, *Communications of the ACM*, *Information and Management*, *Omega*, *Information Systems*, *The Computer Journal*, *Strategic Information Systems*, *Computers & Security*, *Journal of Medical Systems*, and other journals. He is also coauthor of *Information Systems for Management* (in Hebrew).

WILLIAM J. HAGA is a Professor in the Department of Systems Management at the Naval Postgraduate School. He earned his Ph.D. in organization theory from the University of Illinois. He is the author of some thirty academic articles and conference papers. He has taught economics, research methods, marketing, advertising, and management of information systems. Dr. Haga has a minor reputation for his contributions to managerial role theory. He has been nominated three times for the Schieffelin Award for outstanding teaching. He was a Geisert Fellow and a National Science Foundation Fellow, and is a Fellow of the Inter-University Seminar on Armed Forces and Society.

ABSTRACT: Organizations are more dependent than ever on the reliable operation of their information systems, which have become a key to their success and effectiveness. While the growing dependence on information systems creates an urgent need to collect information and make it accessible, the proliferation of computer technology has also spawned opportunities for ill-intentioned individuals to violate the information systems' integrity and validity.

One of the most common control mechanisms for authenticating users of computerized information systems is the use of passwords. However, despite the widespread use of passwords, little attention has been given to the characteristics of their actual use. This paper addresses the gap in evaluating the characteristics of real-life passwords and presents the results of an empirical study on password usage. It investigates the core characteristics of user-generated passwords and associations among those characteristics.

KEY WORDS AND PHRASES: access control, information system security, passwords, user authentication.

Journal of Management Information Systems / Spring 1999, Vol. 15, No. 4, pp. 161–185

0742–1222 / 1999 $9.50 + 0.00

IN NOVEMBER 1988, ROBERT MORRIS, JR., A GRADUATE STUDENT at Cornell University injected an experimental, self-replicating, self-propagating program into the Internet. His program, later dubbed the Internet worm, contained a bug that caused it to propagate itself far faster than Morris intended. It raced across the network using automated password-guessing techniques to penetrate and bring down over 6,200 computer systems in the United States. While no known alteration or destruction of data occurred, the program filled all available memory space on infected computers, bringing them to a grinding halt. The cost of clearing memory space and restarting systems was estimated at 100 million dollars [49, 51].

A key element of the Internet worm involved attempts to discover user passwords. It exploited the tendency of users to choose easy-to-remember passwords and used lists of words, including the standard online dictionary, name lists, and combinations of four-digit numbers (standing for the last four digits of a social security number), as potential passwords. It compared them against the actual passwords stored in the system file and gained access whenever a match was found.

Because the use of passwords has always been one of the most common control mechanisms for authenticating users of computerized information systems, it was expected that the Morris incident would have raised the awareness of system users about the consequences of how they choose their passwords. However, despite the widespread use of passwords and their importance as the first line of defense in most information systems, little attention has been given to the characteristics of their actual use. This study looks at real-life user-selected passwords to address two research questions:

1. What are the characteristics of user-selected passwords, such as number of characters in a password, type of characters used in a password, frequency of changing passwords, and the method of choosing passwords?
2. What are the relationships among key password characteristics, data attributes, and password memorizability.

The Threat

ORGANIZATIONS ARE MORE DEPENDENT THAN EVER ON THE RELIABLE OPERATION of their information systems (IS), which have become a key to their success. Global competitive pressures and continuing innovations are forcing organizations to employ information technology to rationalize business processes, increase organizational effectiveness and productivity, and help them in gaining competitive advantage [65]. Consequently, IS has become essential to the welfare and even survival of many organizations. In many cases it is impossible to run an organization without proper and smooth operation of its information systems.

While the growing dependence on IS creates an urgent need to collect information and render it accessible, the proliferation of computer technology has bred opportunities

for ill-intentioned individuals to violate IS integrity and validity. Despite the scarcity of reliable information about the amount of computer crime that occurs and the nature and severity of the crimes [57], the available evidence suggests significant losses. Studies by LaPlante [37] and Ernst & Whinney [22] revealed that more than half of U.S. firms suffer major dollar losses annually from computer abuse. A study of *Fortune* 500 companies showed that 61 percent of the reported incidents of computer misuse were fraud and embezzlement cases that ranged from several thousands of dollars to over one million dollars [21]. Hoffer and Straub [30] found that one out of five organizations experiences at least one security breach in a three-year period, with one organization reporting a two-million-dollar loss. Moreover, since only 5 to 10 percent of all computer abuse is reported to law enforcement authorities [55], the extent of the problem is probably much larger. In addition to financial losses, computer crime can result in the loss of data or the disclosure of confidential data to competitors.

The following examples convey the variety and scope of computer crimes:

- In the fall of 1978, Stanley Rifkin obtained the electronic transfer code for the Security Pacific Bank in Los Angeles. Posing as a branch manager, he used the code to transfer $13 million from Security Pacific to his Swiss bank account [58].
- A group of German hackers penetrated dozens of military, government, and commercial computer systems by cracking passwords of legitimate users and system administrators. They were looking for military information that could be sold to the Soviet Union [53, 54].
- In April 1994, an English teenager penetrated Pentagon computers and set off a massive security alert when his Internet probing nearly provoked an act of war with North Korea. He entered the Department of Defense systems through the Air Force's Rome (New York) Laboratory using the default password *guest* [59].
- Early in 1998, a trio of Israeli teenagers hacked into the information systems of the Knesset, Israel's parliament. By guessing user passwords, they accessed 150 accounts. They left the data and system unharmed but sent the system administrator an e-mail message describing the system's security loopholes [18].
- Two California teenagers cracked passwords of several Pentagon computers to enter data on the Department of Defense payroll data and personnel files [19].

The introduction of Internet-based purchasing and customer service applications, along with the successful incorporation of the Internet in many extraorganizational IS applications, further enhances their vulnerability and provides evidence that new management-directed countermeasures are required to shore up system defenses [7].

A fundamental security mechanism in any information system is the ability to authenticate the identity of a system user. While research continues on more sophisticated methods of authentication, password mechanisms remain the predominant method of authenticating IS users [14, 38, 61, 63, 64]. Even popular encryption programs, such as PGP and RSA, rely on access authentication via passwords and passphrases [1, 25, 44]. However, while Morris's Internet worm [49], Stoll's experience

[54], and the Mitnick affair [23] caused considerable furor in the press and raised many troubling issues regarding computer security in general and user authentication in particular, password practices seem to be as lax as ever.

Hacker intrusion has raised user awareness about the consequences of lax security practices without addressing the core of the laxity: user password practices. Dependency on global and organizational information systems grows apace without commensurate sophistication in the management of access authentication [8, 14]. However, despite the fact that practically every penetration of a computer system, at some stage, relies on the ability to compromise a password, little attention has been given to the characteristics of their actual use. This paper assesses empirically the characteristics of real-life, user-selected passwords and looks at possible associations among these characteristics.

Evolution of Passwords

CRUCIAL TO ACCESS CONTROL IS THE ESTABLISHMENT OF A POSITIVE, unique identification for each person or entity to whom access is to be granted. While there are several methods for authenticating the identity of a user, the most common method requires a user to provide information that is supposedly known only to him or her [34, 35, 64].

A user typically logs onto a system and then provides a nominal, claimed identity, such as a user-name or an account number. The system then requests a password that is a single, mutually agreed-upon code word, assumed to be known only to the user and the operating system. In some cases a password is chosen by a user; in other cases, it is assigned by the security kernel of the operating system. When entered at log-on, a password is checked against the information stored in a password file in the operating system; if there is a match, the user is granted access to the system [6].

Most multiuser computer systems employ *user-selected* passwords. The advantage of user-selected passwords is that they are easily remembered by users. The disadvantage is that they are often weak. *System-assigned* passwords are usually stronger than user-selected passwords, but generally difficult for users to remember [6, 64].

The tradeoff between memorizability and security poses a dilemma for self-selected passwords. Passwords should be difficult to guess and easy to remember. For passwords to be difficult to guess, they should be selected from a large domain. However, if passwords are chosen to make them difficult to guess, they may also be difficult to remember. The most secure type of password is a random string of characters [5, 14, 27, 29]. Although long, random, arbitrary passwords are difficult for others to guess, users generally cannot remember them. Thus, most users will take the road of least resistance and resort to the minimum number of characters acceptable by their system and use meaningful details, such as their name, nickname, initials, or birth date [27, 28, 39].

A password that is difficult to remember invites users to write it down, ensuring they will not forget it but compromising its secrecy [45]. On the other hand, if a difficult password is not written down, it may well be forgotten, resulting in serious inconve-

nience [1, 3]. An organization should establish a password policy that strikes a balance between ease of recall and susceptibility to compromise [8].

A unique effort to reveal the characteristics of passwords used in a real-life system was made by Morris and Thompson [42]. They described the basic characteristics of user-generated passwords in a UNIX environment and analyzed the level of security provided by these passwords. Examining over three thousand passwords, they noted that over 85 percent of them fell into one of the following categories: words in an English dictionary, reverse spelling of words in a dictionary, first or last names, street names, cities, and social security or telephone numbers. These findings suggest that most passwords are easy to guess and fail to provide the required level of security to the systems they are supposed to protect.

Despite the growing importance of passwords as a first line of defense in most information systems, no follow-on research or additional empirical work on password usage has been reported since the Morris and Thompson study. The literature on password methods features lists of do's and don'ts [5, 33]. The "do" list recommends passwords that are long, composed of random characters, and frequently changed. The "don't" list warns against the use of names, initials, dates, and personal and environmental cues.

Key Issues and Exploration Hypotheses

TO ADDRESS THE MOST PROMINENT PASSWORD USAGE CONCERNS of management, this study focused on two research questions: (1) What are the characteristics of user-selected passwords, and (2) what are the relationships among key password characteristics?

The first question focuses on the actual implementation of real-life passwords. Password characteristics under investigation are *length* (number of characters in a password), *composition* (character domain: alphabetic, numeric, alphanumeric, or the entire ASCII character set), *lifetime* (frequency of changing passwords), and password *selection method*. Selection method means whether a password is based on a personally meaningful detail (user's last name, first name, nickname, child's name, or some other easily recalled bit of personal, biographical information), a combination of meaningful details (BILL89 or LOVEMARY), a pronounceable string of characters (2BFREE), string of random characters (H*DGFH8H), or some other basis [2, 14, 20, 28, 29, 34, 60, 64].

The second question focuses on possible relationships among data attributes, key password variables, and password memorability. Of specific interest are associations between data attributes and password characteristics, password characteristics and their memoraizbility, and computer usage pattern and password memorizability. Figure 1 summarizes the key variables and associations tested.

Our first hypothesis refers to possible associations between data importance and sensitivity and password characteristics. Data importance refers to the inherent value of the data to an individual user. Sensitivity means the degree to which problems would arise if the contents of their data files were known to others. To distinguish between importance and sensitivity of a data file, consider, for example, a data file containing

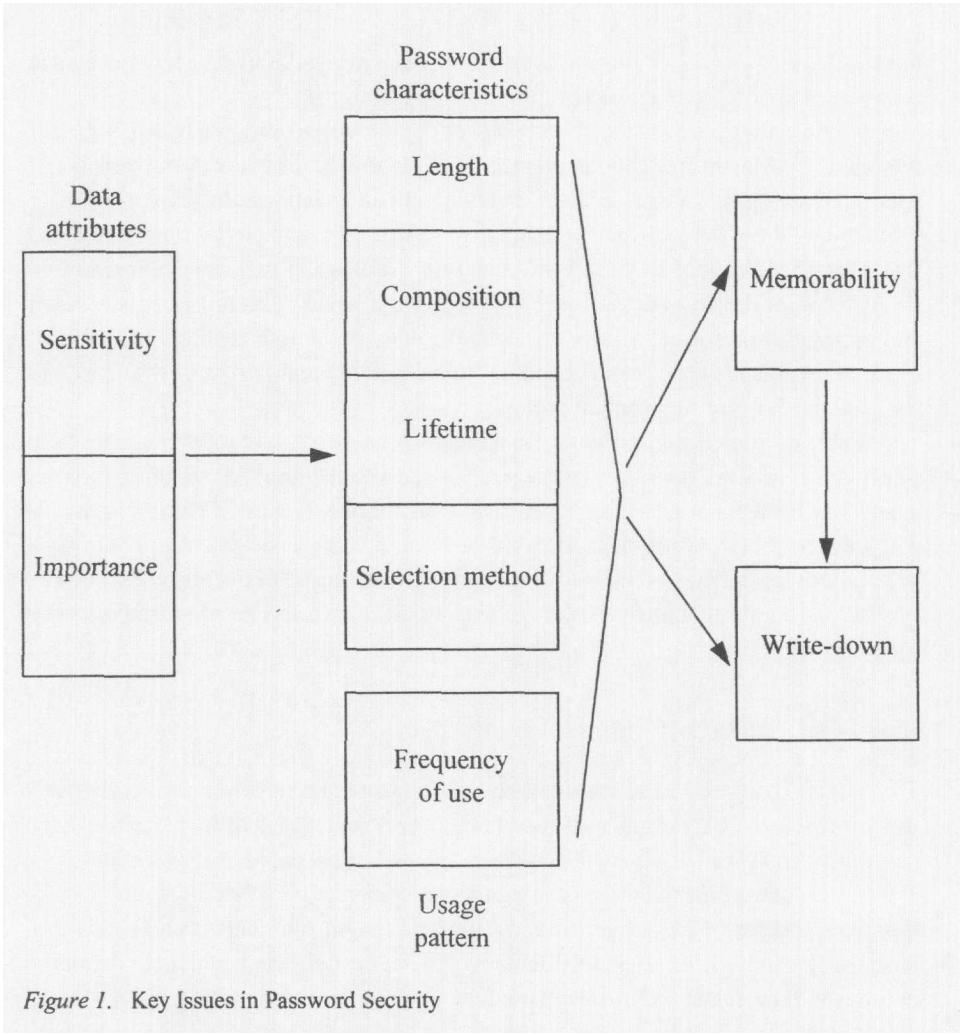


Figure 1. Key Issues in Password Security

the text of this research paper. It might not be publicly sensitive but it would be of major value to its authors. By comparison, a professor's data file containing student course grades would have little inherent value but would be highly sensitive to disclosure. In the United States, divulging such a list could violate laws regarding privacy of information.

Previous research reveals little about the possible associations between data attributes and password characteristics. Highland [29] and Hoffman [31] suggested that the level of security should be commensurate with the importance of the resources it protects, but this relationship has not been investigated empirically. This research will test if data attributes, such as data importance and sensitivity, affect certain characteristics of user-generated passwords.

> *H1: Password characteristics (length, composition, lifetime, and selection method) are associated with the importance and sensitivity of the data being protected.*

The next set of hypotheses investigates possible relationships between password characteristics and their memorizability and the tendency to write down passwords. It has been suggested that password characteristics affect memorizability [2, 3, 28, 45, 46]. It follows that, if a password is difficult to remember, it will be written down [3, 27, 28, 64]. Empirical evidence for such relationships is sought through the following hypotheses:

H2: Password characteristics (length, composition, lifetime, and selection method) are associated with difficulty in remembering the password.

H3: Password characteristics (length, composition, lifetime, and selection method) are associated with writing the password down.

H4: The difficulty of remembering a password is associated with writing the password down.

A third issue to be explored is whether password memorizability or the tendency to write down a password is related to frequency of use. Menkus [39] assumes that frequent use of a password enhances memorizability and reduces the need to write it down. Similar assertions are made by Ahituv et al. [2], Barton and Barton [5], and Morrey [41]. The following hypothesis addresses these relationships:

H5: The difficulty of remembering a password or writing it down is associated with the frequency with which it is used.

Methodology

Instrumentation

DATA FOR THIS STUDY WERE COLLECTED BY MEANS OF A QUESTIONNAIRE that asked for responses in five areas: user demographics, data attributes, computer usage pattern, password characteristics, and password memorizability. *Demographic items* included age, sex, and organizational affiliation. *Data attributes* were addressed in two questions that separately targeted the importance and the sensitivity of a user's data files using a five-point, Likert-type scale. Data importance referred to the inherent value of the data to an individual user, and the scale ranged from "nonvital" (1) to "highly vital" (5). For example, if the data were lost, what would be the impact on a respondent's ability to do his or her job? Sensitivity means the degree to which problems would arise if the contents of data files were known to others. The scale ranged from "nonsensitive" (1) to "highly sensitive" (5). For example, if the data were made public, what would be the impact on a respondent's organization? Data importance and data sensitivity are characteristics of the data only insofar as they are expressed through each user's attitudes about the data within the context of his or her organization.

To assess *computer usage patterns*, respondents were asked how often they used their password. The ordinal scale ranged from "never" to "several times a day."

Password characteristics involve length (number of characters in password), composition (multiple-choice option: alphabetic, numeric, alphanumeric, or the entire ASCII character set), lifetime (ordinal scale, ranging from "never" to "more than once a month"), and selection method (a meaningful detail, a combination of meaningful details, pronounceable, random combination of characters, or other). Selection method means whether a password is based on (a) a personally meaningful detail, such as user's last name, first name, nickname, child's name, (b) a combination of meaningful details (BILL89 or LOVEMARY), (c) a pronounceable string of characters (2BFREE), (d) a string of random characters (H*DGFH8H), or (e) some other basis that the respondent was asked to specify [2, 20, 28, 34, 60].

The fifth area, *password memorizability and write-down practice*, entails difficulty in remembering a password (yes or no), if the password was written down (yes or no), and, if so, where (wallet, notebook, calendar, desk, drawer, keyboard, monitor, etc.).

Instrument Validation

Before the questionnaire was administered, it was tested for validity and reliability [12, 16, 32, 56]. Content validity of the items was established through a literature review on user authentication in general and passwords in particular. This was used as a starting point for determining the central and tangential dimensions of the construct. As the first stage of a prestudy, a draft form of the questionnaire was presented to eight IS security experts who monitor and manage password practices, and to twelve users. We were thereby able to better interpret particular questions and assess the clarity and completeness of the questionnaire. As a result, we added one new item to the questionnaire and modified three others. The revised questionnaire was approved by the prestudy panel.

In the second prestudy stage, thirty-six randomly selected users were asked to judge the questionnaire. Specifically, we divided these users into two groups of eighteen individuals. The subject matter and the dimensions of the questionnaire were explained to the members of the first group who were then interviewed and asked to evaluate the items for their applicability to the respective dimensions of the questionnaire. The questionnaire was administered to the same individuals one month later [15]. The results of the interviews and corresponding questionnaire responses were then matched (correlation = 0.93, $p < 0.001$). Matching of the second group's two response waves yielded a correlation of 0.90 ($p < 0.001$). The matching results of the combined group also reveal a statistically significant correlation ($r = 0.92, p < 0.001$).

A third prestudy stage assessed the test-retest reliability of the instrument. The questionnaire was administered on two occasions separated by a six-week interval to a sample of eighty-nine randomly selected users. The correlation for this repetition of the overall instrument was $r = 0.88$ ($p < 0.001$). Test-retest correlations for individual items ranged from $r = 0.64$ to $r = 0.96$ ($p < 0.001$ for all).

Source of the Data

The source of the data for this study was computer users at a Department of Defense (DoD) installation in California. Most employees at this installation have access to one or more of the computing center facilities within the installation. These facilities can be accessed using a combination of a user ID (assigned by the computing center) and a user-generated password. Once authenticated and granted access, a user has immediate access to a variety of computing resources within the DoD through the attached communications network, as well as access to the Internet.

Sample Characteristics

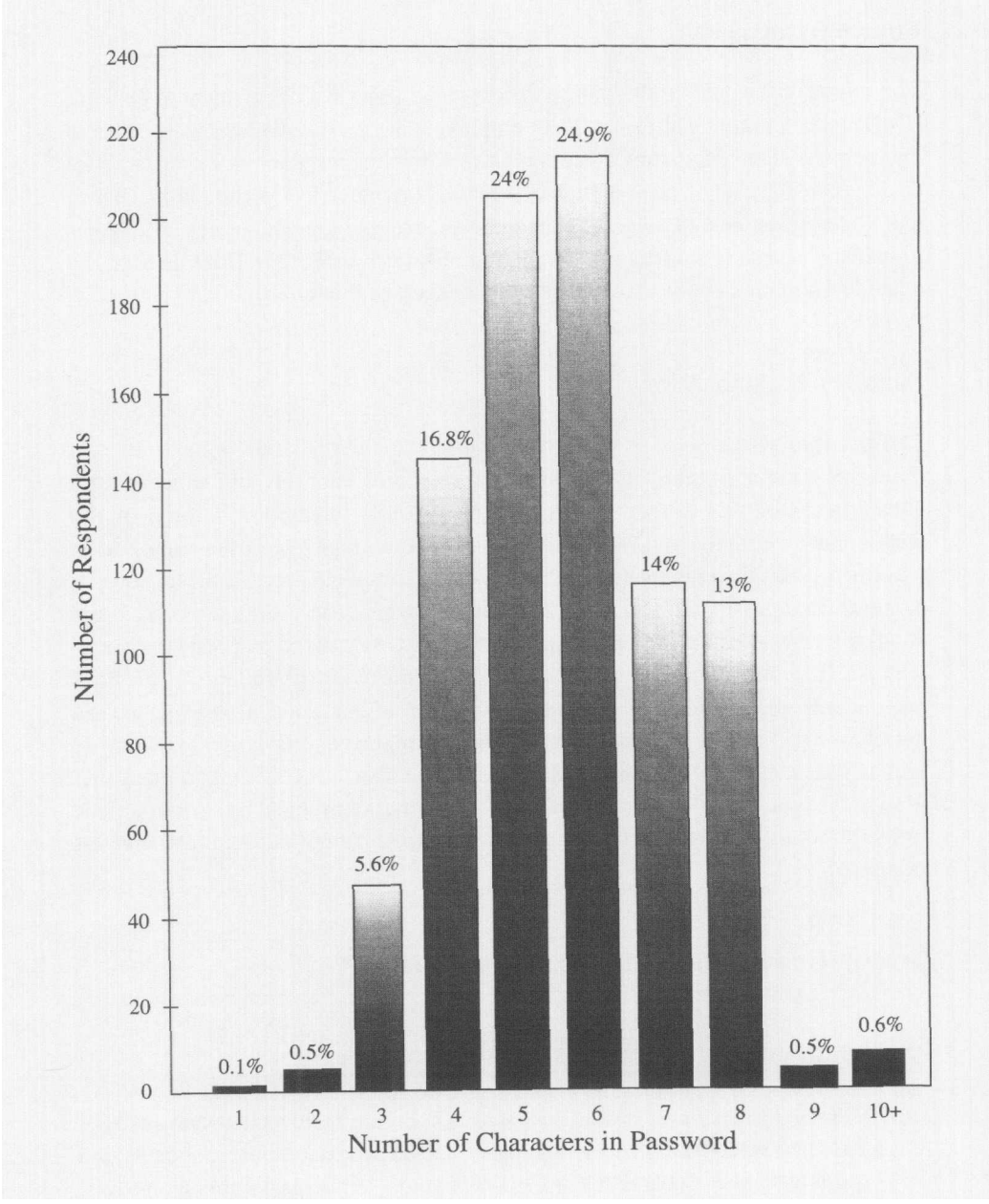
The questionnaires were distributed through the internal mail system to two thousand computer users at this installation. Nine hundred and ninety-seven questionnaires (49.9 percent) were returned. Seven hundred and three of the respondents were men and 294 were women. The average age of the respondents was thirty-four; the range was from twenty-three to seventy-six. Of these, 860 (43 percent) used passwords and were included in the analysis. Because asking users about their password practices is a sensitive topic, we treated responses as anonymous. This, of course, deprived us of any opportunity to follow up on nonrespondents and introduced bias into the findings of our study. However, we are not claiming that the findings here statistically represent any larger population. This is a case study of system users in one government organization at one time. We would speculate that our findings are probably not atypical of user practices and shortcomings at other places and times, but we cannot substantiate such an assertion.

Descriptive Findings: Password Characteristics

Password Length

THE AVERAGE NUMBER OF CHARACTERS WAS SIX. Figure 2 shows that passwords of five and six characters nearly tied in popularity, and 80 percent of the respondents said they used four to seven characters in their passwords. While passwords are recommended to be, at minimum, six characters in length [13, 20, 27, 47] and the ideal length is thought to be six to eight characters [39], 47 percent of the surveyed respondents failed to create a password that long. Windows NT and Unix support passwords up to fifteen characters long [36, 62], while the password length in RSA is up to thirty-two characters [1].

Morris and Thompson [42] suggest that a shorter password means less work for an intruder in a brute force attack to discover a user's password. Their view is based on search complexity and is supported in the literature [3, 34, 47].



Password Composition

As depicted in figure 3, 80.1 percent of the respondents claimed to prefer alphabetic characters for their passwords, while only 0.7 percent said they used the entire ASCII character set as a basis for their password. This finding lends further support to Morris and Thompson's [42] study and suggests that, despite the large

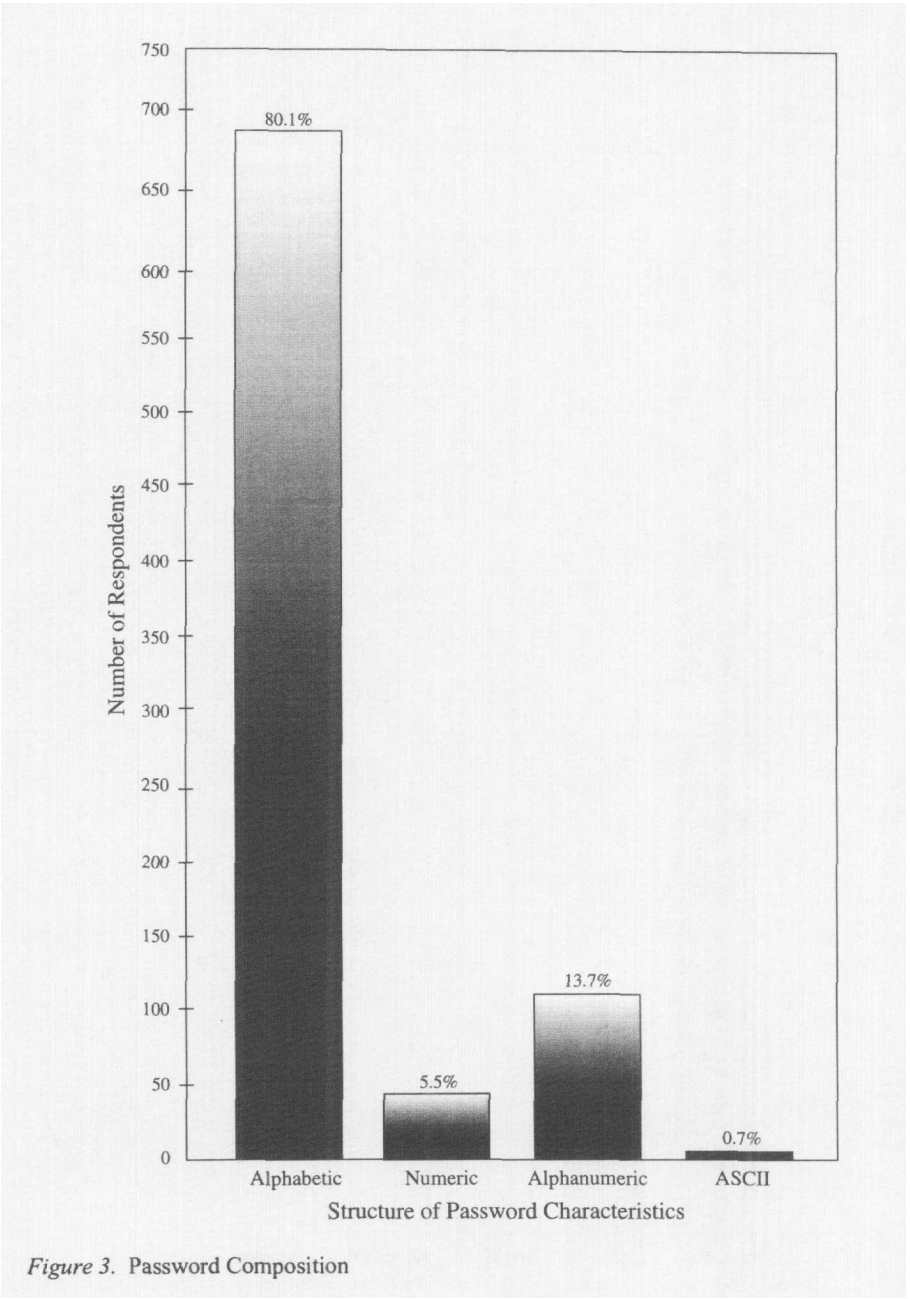


Figure 3. Password Composition

variety of characters available to users, they tend to avoid nonalphanumeric symbols in their passwords.

Password Change Frequency

The periodic changing of a password is a basic security measure [20, 34, 41]. Wood [60] asserts that passwords should be changed annually. Menkus [39] suggests every

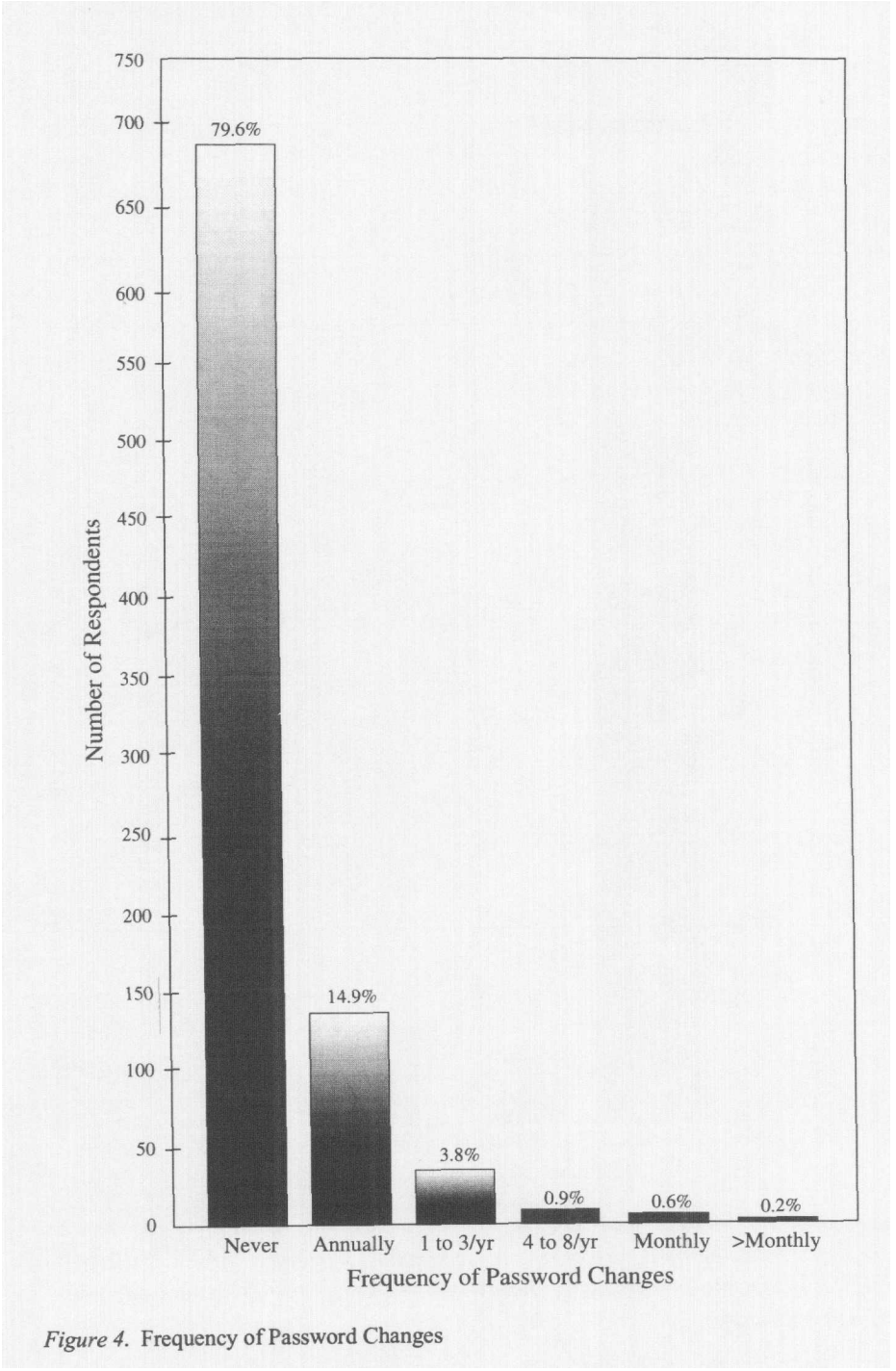


Figure 4. Frequency of Password Changes

thirty days. While research supports the frequent changing of passwords to reduce predictability, this study found that 79.6 percent of the surveyed users said they *never* changed their password (figure 4). Less than 5.5 percent of them said they changed their passwords more often than once a year.

Password Selection Method

Password compromises have resulted from information on computer bulletin boards, guesses about meaningful details about a user, environmental cues, and systematic intrusions [5, 14, 29, 42, 49]. Examples of meaningful details are names, nicknames, name of child, name of pet, name of spouse or relations, slang, profanity, car names, or birth dates. The item has meaning for the person using it, which should enhance its memorizability. However, selecting meaningful details as passwords increased "the hackers' comfort" [26], since it limits the number of guesses a penetrator has to make [35]. Figure 5 shows that users have a strong preference (78.4 percent) for passwords made up from a meaningful detail or a combination of meaningful details.

Password Memorizability and Write-Down

When a user writes down a password, he or she usually does so in an insecure location [14, 52]. Once a password is written down, it is no longer something to be guessed but becomes something to be located [48]. Consequently, searching through a user's notebook, desk, diary, or user's manual will usually reveal a password [3].

The *DoD Password Management Guidelines* recommend that "If passwords must be written, they should be protected in a manner that is consistent with the damage that could be caused by their compromise" [20]. While only 9.7 percent of the respondents reported difficulty remembering their password, 35.3 percent of them said they wrote down their passwords. The popular location of choice was the wallet (42.1 percent), followed by a notebook (21.3 percent), calendar or organizer (16.6 percent), and desk/keyboard/monitor (7.6 percent).

Relationships Between Password Characteristics

SELECTION OF THE APPROPRIATE TEST FOR EACH OF THE HYPOTHESES that follow depends on the nature of the particular variables being tested [50]. By convention, in reporting empirical findings, *associated* refers to a relationship between variables that does not assert or suggest causality [24, 64].

Relationship Between Data Attributes and Password Characteristics

Data importance refers to the inherent value of the data to an individual user. Sensitivity means the degree to which problems would arise if the contents of data files were known to others. We considered that respondents who scored their data as a 4 or 5 on the importance scale regarded it as at least "vital." Similarly, respondents who scored their data as a 4 or 5 on the sensitivity scale judged it at least "sensitive." While 42.5 percent of the respondents considered their data to be at least "vital," 49.3 percent judged their data to be at least "sensitive."

In testing the first hypothesis, we followed Hoffman [31] and Highland [28, 29] in assuming that data rated as sensitive or important would be surrounded with more security than data not so rated. For the sake of presentation, H1 was broken down into

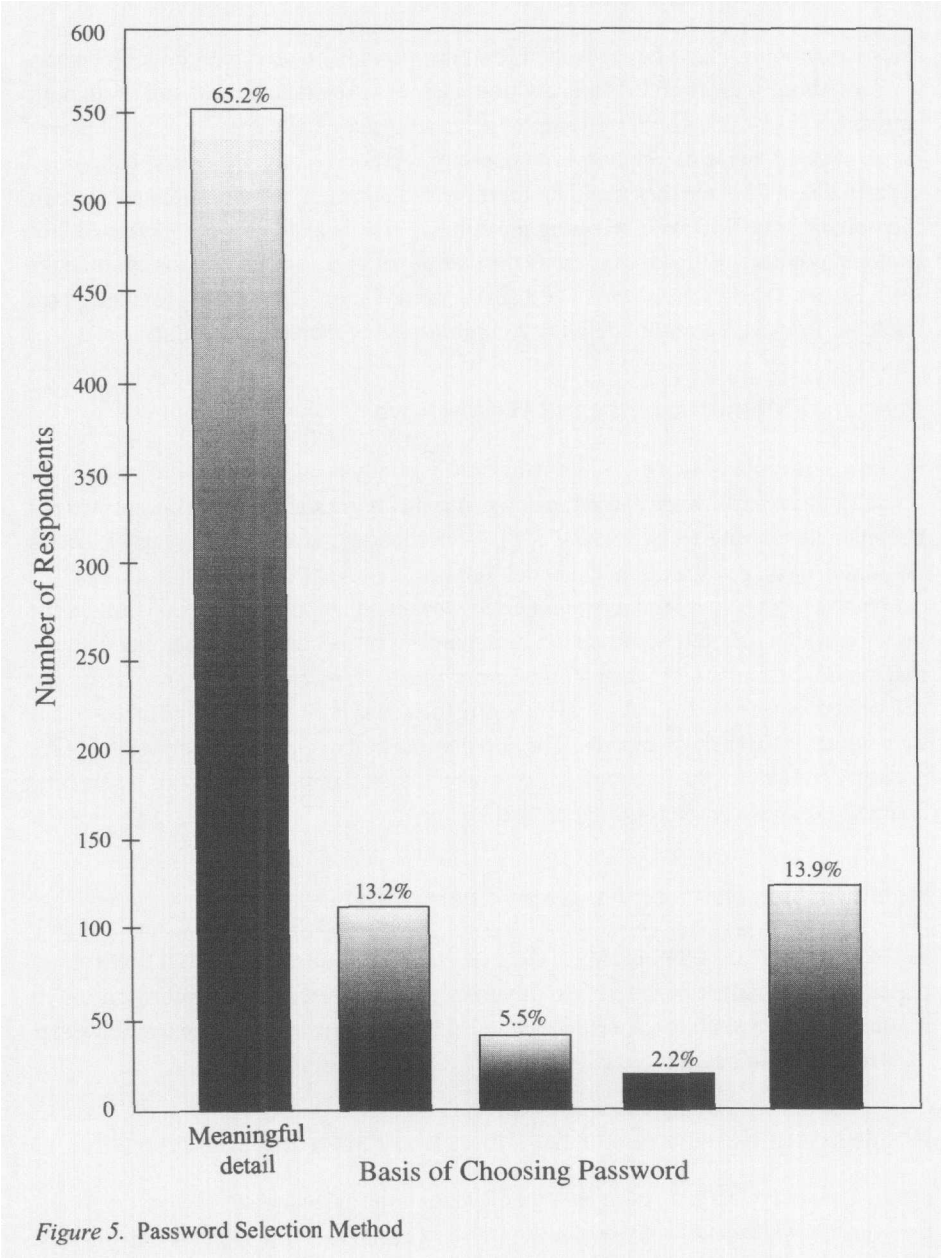


Figure 5. Password Selection Method

four subhypotheses, each of which refers to a single password characteristic.

H1a: The number of characters in a password is associated with the characteristics (sensitivity and importance) of the data being protected.

A null hypothesis of no association between the number of characters in a password and either the importance or sensitivity of the data it protects could not be rejected at

Table 1. Associations of Password Characteristics with Data Attributes

Hypothesis	Password characteristic	Data attribute	Level of measure	Test	Test value	Probability	Reject null hypothesis
H1a	Length	Sensitivity	Interval	ANOVA	1.388	0.236	No
		Importance	Interval	ANOVA	0.430	0.787	
H1b	Composition	Sensitivity	Nominal	Kruskal-Wallis	2.886	0.5771	No
		Importance	Nominal	Kruskal-Wallis	8.073	0.0889	
H1c	Lifetime	Sensitivity	Ordinal	Spearman's rho	0.1544	0.0000	Yes
		Importance	Ordinal	Spearman's rho	0.1916	0.0000	
H1d	Selection	Sensitivity	Nominal	Kruskal-Wallis	11.264	0.0122	Yes
		Importance	Nominal	Kruskal-Wallis	12.98	0.0114	

the 0.05 level of probability that this association could have occurred by chance alone. The finding is that no association existed between the length of a password and the importance or sensitivity of the data it protected. This finding contradicts the assumption that a user would treat sensitive or important data with care and select a longer password, comprised of a variety of alphanumeric and nonalphanumeric symbols.

H1b: The composition of a password is associated with the characteristics (sensitivity and importance) of the data being protected.

A null hypothesis of no association between the composition of a password and either the importance or sensitivity of the data it protects could not be rejected at the 0.05 level of probability that this association could have occurred by chance alone. The finding is that no association existed between the composition of a password and the importance or sensitivity of the data it protected.

H1c: The frequency with which a password is changed is associated with the characteristics (sensitivity and importance) of the data being protected.

A null hypothesis of no association between the frequency with which a password is changed and either the importance or sensitivity of the data it protects was rejected at the 0.05 level, suggesting that frequency of changing a password is related to the importance or sensitivity of the data it protects. These findings confirm Highland's finding [29] that users take care to choose passwords that are difficult to predict and take the precaution of frequently changing their passwords in order to protect sensitive or important data.

H1d: The method used to select a password is associated with the characteristics (sensitivity and importance) of the data being protected.

A null hypothesis of no association between the method used to select a password and either the importance or sensitivity of the data it protects was rejected at the 0.05 level. This lack of association between the selection method of a password and the importance or sensitivity of the data it protects might be explained by the likelihood that, when users choose a password, they are naive about the importance or sensitivity of the information they will be storing. Once they grasp the character of the data they will be protecting, users may then change their passwords accordingly.

Relationship Between Password Characteristics and Memorizability

The second set of hypotheses investigates possible relationships among password characteristics, their memorizability, and the tendency to write down passwords. It has been suggested that password characteristics affect memorizability [2, 33, 28, 45, 46]. Consequently, it follows that if a password is difficult to remember it will be written down [3, 27, 28, 52].

H2: The characteristics of a password (number of characters, composition, frequency of change, and method of selection) are associated with difficulty in remembering it.

Table 2. Association of Password Characteristics with Memorizability

	Association between password memorizability and:				
Variable	Level of measure	Test	Test value	Probability	Reject null hypothesis
Length	Interval	*t*-test	–0.38	0.706	No
Composition	Nominal	Cramer's V	0.1131	0.0110	Yes
Lifetime	Ordinal	Mann-Whitney	25363	0.0000	Yes
Selection	Nominal	Cramer's V	0.1221	0.0121	Yes

A null hypothesis of no relationship between the characteristics of a password and difficulty remembering it was rejected at the 0.05 level in the case of a password's composition, frequency of change, and method of selection. However, the null hypothesis could not be rejected in the case of password length. The finding is that a password's composition, frequency of change, and method of selection are related to how difficult it is to remember, while the number of characters in a password is not (Table 2).

The finding of a connection between password composition and recall supports Wood [60], who found that alphanumeric passwords based on a meaningful detail are more easily remembered than passwords generated from pseudo-random combinations. Users will select passwords from a simple domain of things meaningful to them or something from episodic memory [39, 60]. The number of characters in a password was expected to affect memorizability [40]. Barton and Barton [5], Highland [27, 29], and Menkus [39] suggest that the ability to recall a password decreases as its length increases. It has long been accepted that people can remember expressions of about seven characters in length [40]. Menkus [39] has proposed that passwords be in the range of six to eight characters. Respondents in this study did not adhere to these suggestions.

H3: The characteristics of a password (number of characters, composition, frequency of change, and method of selection) are associated with writing it down.

A null hypothesis of no relationship between a password's characteristics and writing it down could not be rejected at the 0.05 level for password length, frequency of change, and selection method. However, the null hypothesis was rejected in the case of password composition. The finding is that length, change frequency, and selection method are not related to writing down a password, while its composition is. The lack of empirical support here for a connection between password length and writing it down challenges Avarne's [3] suggestion that writing down a password is a function of its length (Table 3).

H4: The difficulty of remembering a password is associated with writing it down.

Table 3. Association of Password Characteristics with Writing Down Passwords

	Association between writing down a password and:				
Variable	Level of measure	Test	Test value	Probability	Reject null hypothesis
Length	Interval	*t*-test	–0.20	0.839	No
Composition	Nominal	Cramer's V	0.1194	0.0065	Yes
Lifetime	Ordinal	Mann-Whitney	65872	0.9899	No
Selection	Nominal	Cramer's V	0.0875	0.1584	No

A null hypothesis of no relationship between difficulty in remembering a password and writing it down was rejected at the 0.05 level ($\chi^2 = 38.45, p = 0.000$). The finding is that the difficulty of recalling a password is related to writing it down. These findings show that 9.7 percent of the respondents found it difficult to remember their passwords, while 23.3 percent wrote them down. Users who expect that they will not be logging into the system frequently may choose to write down their password for future reference [27]. Users may write down their password simply out of habit or because they anticipate changing their password frequently [2, 3, 39].

Relationship Between Frequency of Password Usage and Memorizability

The third issue explored is whether password memorizability or the tendency to write down a password are related to the frequency of password use. This issue is addressed by H5:

> *H5: The difficulty of remembering a password or writing it down is associated with the frequency with which it used.*

Using the Mann-Whitney test, the null hypothesis of no relationship between the frequency with which a password is used and difficulty in recalling it or writing it down was rejected at the 0.05 level. The finding is that usage frequency is related to recall difficulty and writing down a password. This finding supports the assumption that infrequent use of a password could lead to the password being forgotten and prompt a user to write it down, and it supports Spender's [52] suggestion that frequent access to an information system is related to password memorizability [2, 41, 52].

Discussion

USER-SELECTED PASSWORDS ARE RELATIVELY WEAK AND EASY TO GUESS, and their characteristics in the Internet era have not changed much from those in the pre–personal computer era identified by Morris and Thompson [42]. Despite efforts by information

system professionals to educate users about secure password practices, this study found that user-selected passwords are still being made up of the characteristics of personal details meaningful to the user, are relatively short, are comprised of alphanumeric characters, are rarely changed, and are usually written down. Passwords remain easy to remember and simple in structure.

While 8.6 percent of the users in this study rated their data files as "nonvital," we expected that those who rated them as vital would be careful in choosing and using their passwords. This study showed that how a password is chosen, the number of characters in a password, and password composition (alphanumeric or ASCII) were not affected by the level of data importance or sensitivity. This finding can be explained by the fact that most users are asked to devise a password when registering as new system users, long before they can know how important or sensitive their data files will be. As new users, most likely they lack information system security consciousness and discipline.

It is important, however, that we found that the frequency of changing a password is affected by the level of data importance and sensitivity. A user will choose to change his or her password more often if he or she is protecting important or sensitive data files.

The analysis of the relationships among the password variables and their memorizability and the tendency to write them down yielded both confirmations of the conventional wisdom about password practices as well as some surprises. The confirmations were:

1. Password selection methods affect password memorizability.
2. The frequency of changing a password, although it increases the level of security, hinders memorizability.
3. The more frequently a password is used, the less it is written down.
4. The more a password is used, the less difficult it is to remember.
5. Frequently changing passwords, necessary to reduce password predictability, nonetheless hinders recall.
6. Difficulty recalling a password is related to a user's tendency to write it down.

The surprises were:

1. Difficulty recalling a password or writing it down is not related to a password's length.
2. Whether a password was chosen in such a way as to make it easy to remember had no bearing on whether or not it was written down.

Many of the findings reported here are neither confirmations nor surprises because they introduce dimensions of password security previously not explored in the literature, such as data importance, data sensitivity, and frequency of use.

Conclusion and Recommendations

MANAGEMENT'S CONCERN WITH INFORMATION SYSTEM SECURITY has decreased over recent years. In 1981, it ranked as the fourteenth most important information management topic [4]. In a 1986 study, security was ranked eighteenth [9], and in a 1989 survey the issue had dropped to nineteenth place [43]. In a 1995 study [10], MIS executives dropped the security issue from their top twenty MIS issues, suggesting that either security had become less of an issue or they had implemented greater control.

The traditional user-selected password is still the common means of authentication for users. Endorsed by U.S. Government security agencies, user-selected passwords are popular because they are conceptually simple, inexpensive to administer, and user-friendly [11]. However, the empirical findings reported here suggest that computer users tend to violate secure password practices, resulting in passwords that are easy to guess.

User-selected passwords are supposed to be long, composed of characters beyond alphanumerics, changed frequently, and not based on personal details; however, almost 50 percent of the users surveyed in this study reported passwords composed of five or fewer characters; 80 percent used only alphabetic characters; 80 percent never changed their password; and 78 percent based their password on a combination of meaningful details. These findings indicate a need to look at the effectiveness of educational efforts to raise the security consciousness of system users. In addition, organizations should have a set of guidelines for selecting and implementing user-selected passwords and mechanisms that monitor their implementation.

Another avenue to overcome the problem of selecting easy-to-guess passwords is the use of password validation software such as Proactive Password Checker [6] for Unix systems, Baseline Software's Password Coach, Computer Oracle and Password System, and Crack, a public-domain password tool [17]. These programs evaluate each user-selected password for its characteristics to ensure that all users employ strong passwords. Crack uses its own dictionaries and rules to guess user's passwords. If a user fails to select a strong password in a given number of attempts, some systems assign a system-generated password that has the required characteristics. Also, in order to minimize the risk of systematic password-guessing trials, stricter user-authentication strategies, such as "three strikes and you are out" should be employed.

The details of passwords and their effectiveness warrant further research. First, the information systems community enjoys a surfeit of essays and nonempirical insights into what users ought to do about password practices. This community will benefit from channeling research efforts toward investigations of how users actually use their passwords. Second, the IS community would be well served if the procedures described here were replicated to challenge these findings in various user populations and organizations. A third avenue for future research is possible differences in the quality of passwords between organizations with a well-defined security policy and those without one. Finally, the security impact of using password validation systems in a user-generated password environment should be examined.

The fact that a sixteen-year-old boy in Colindale, England, could unwittingly make

it appear that the United States was attempting to access North Korean computers, just because the default password *guest* was all he needed to break into the Pentagon's data systems [59], should make it clear that the lessons about user password practices remain to be learned.

Acknowledgment: The authors thank the Editor-in-Chief and two anonymous reviewers for their valuable comments on earlier versions of this paper.

REFERENCES

1. Adams, S. How to keep a secret. *Forbes*, *157*, 7 (April 1996), 108–109.
2. Ahituv, N.; Lapid, Y.; and Neumann, S. Verifying the authentication of an information system user. *Computers and Security*, *6*, 2 (1987), 152–157.
3. Avarne, S. How to find out a password. *Data Processing & Communication Security*, *12*, 2 (Spring 1988), 16–17.
4. Ball, L., and Harris, R. SMIS member: a membership analysis. *MIS Quarterly*, *6*, 1 (March 1982), 19–38.
5. Barton, B.F., and Barton, M.S., User-friendly password methods for computer-mediated information systems. *Computers and Security*, *3*, 3 (1988), 186–195.
6. Bishop, M., and Klein, D.V. Improving system security via proactive password checking. *Computers and Security*, *14*, 3 (1995), 233–249.
7. Botting, J. Security on the Internet: authenticating the user. *Telecommunications*, *31*, 12 (December 1997), 77–79.
8. Bradner, S. But will they pay attention this time? *Network World*, *14*, 4 (January 1997), 32–34.
9. Brancheau, J.C., and Wetherbe, J.C. Key issues in information systems management. *MIS Quarterly*, *12*, 1 (March 1987), 23–36.
10. Brancheau, J.C., and Wetherbe, J.C. Key issues in information systems management: 1994–95 SIM/Delphi results. *MIS Quarterly*, *20*, 2 (June 1996), 225–242.
11. Broderick, J. Who knows who you are? *Infoworld*, *19*, 24 (June 1997), 108–112.
12. Cooper, D.R., and Emory, C.W., *Business Research Methods*, 5th ed. Dubuque, IA: Irwin, 1995.
13. Cooper, J.A. *Computer and Communications Security, Strategies for the 1990s*. New York: McGraw-Hill, 1989.
14. Corbitt, T. Ensure your datafiles are secure even if the Pentagon's are not. *Management Services*, *41*, 5 (May 1997), 24–26.
15. Cronbach, L. Test validation. In R.L. Thorndike (ed.), *Educational Measurement*, 2d ed. Washington, DC: American Council on Education, 1971, pp. 443–507.
16. Denzin, N.K. *The Research Act*, 3d ed. Englewood Cliffs, NJ: Prentice-Hall, 1989.
17. Dichter, C. Easy Unix security. *Unix Review*, *11*, 4 (April 1993), 42–51.
18. DiDio, L. Major hacks raise hackles, spur defenders. *ComputerWorld*, *32*, 13 (April 1998), 49–50.
19. DiDio, L. Cyberattack prompts DoD to boost security. *ComputerWorld*, *32*, 9 (March 1998), 14.
20. DoD. *Department of Defense Password Management Guideline*. Washington, DC: National Computer Security Center, CSC-STD-002–85, 1985.
21. Ernst & Whinney. *U.S. Computer Security Survey of Fortune 500 Industrial Companies*. Cleveland: Ernst & Whinney, 1987.
22. Ernst & Whinney. *The 1989 Computer Abuse Survey: A Report*. Cleveland: Ernst & Whinney, 1989.
23. Goodell, J. *The Cyber Thief and the Samurai*. Menlo Park, CA: Dell, 1996.
24. Graziano, A.M., and Raulin, M.L. *Research Methods*, 3d ed. Reading, MA: Addison-Wesley, 1997.

25. Grossman, W. *Net Wars*. New York: New York University Press, 1997.
26. Herschberg, I. The hacker's comfort. *Comity, 6*, 2 (1987), 133–138.
27. Highland, J.H. Demise of passwords. *Computers and Security, 9*, 4 (1990), 196–200.
28. Highland, J.H. How to prevent the use of weak passwords. *EDPACS, 18*, 9 (March 1991), 7–12.
29. Highland J.H. Changing passwords. *Computers and Security, 16*, 3 (1997), 183–184.
30. Hoffer, J., and Straub, D.W. The 9 to 5 underground: are you policing computer crimes? *Sloan Management Review, 30*, 4 (Summer 1989), 35–44.
31. Hoffman, L.J. *Modern Methods for Computer Security and Privacy*. Englewood Cliffs, NJ: Prentice-Hall, 1977.
32. Hunt Sparkman, R.D., Jr., and Wilcox J.B. The pretest in survey research: issues and preliminary findings. *Journal of Marketing Research, 19*, 2 (1982), 269–273.
33. Icove, D. *Computer Crime: A Crime Fighter's Handbook*. Sebastopol, CA: O'Reilly & Associates, 1995.
34. Jobusch, D.L., and Oldhoeft, A.E. A survey of password mechanisms: weaknesses and potential improvements, part 1. *Computers and Security, 8*, 7 (1989), 587–601.
35. Jobusch, D.L., and Oldhoeft, A.E. A survey of password mechanisms: weaknesses and potential improvements, part 2. *Computers and Security, 8*, 8 (1989), 675–689.
36. Kearns, D. Paying attention to passwords. *Network World, 13*, 23 (June 1996), 28–29.
37. LaPlante, A. Computer fraud threat increasing, study says. *Infoworld, 18* (May 1987), 47.
38. Loch, K.R.; Carr, H.H.; and Warkentin, M.E. Threats to information systems: today's reality, yesterday's understanding. *MIS Quarterly, 16*, 2 (June, 1992), 173–186.
39. Menkus, B. Understanding the use of passwords. *Computers and Security, 7*, 2 (1988), 132–136.
40. Miller, G.A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 3 (March 1956), 81–97.
41. Morrey, B. Beefing up NT's security out of the box. *Inforworld, 19*, 24 (June 1997), 122–124.
42. Morris, R., and Thompson, K. Password security: a case history. *Communications of the ACM, 22*, 11 (November 1979), 594–597.
43. Neiderman, F.; Brancheau, J.C.; and Wetherbe, J.C. Information systems issues for the 1990s. *MIS Quarterly, 15*, 4 (December 1991), 475–502.
44. Nelson, M. PGP's Business Security Suite spotlights corporate users. *ComputerWorld, 31*, 21 (October 1997), 88.
45. Paans, R., and Herschberg, I.S. Computer security: the long road ahead. *Computers and Security, 6*, 5 (1987), 403–416.
46. Parker, D. *Fighting Computer Crime*. New York: Charles Scribner's Sons, 1983.
47. Pfleeger, C.P. *Security in Computing*, 2d ed. Englewood Cliffs, NJ: Prentice-Hall, 1997.
48. Porter, S.N. A password extension for human factors. *Computers and Security, 1*, 1 (1982), 54–56.
49. Seeley D. Password cracking: a game of wits. *Communications of the ACM, 32*, 6 (June 1989), 700–703.
50. Siand Castellan, J.N. *Nonparametric Statistics for the Behavioral Sciences*, 2d ed. Boston: McGraw-Hill, 1988.
51. Spafford, E. The Internet worm: crisis and aftermath. *Communications of the ACM, 32*, 6 (June 1989), 203–227.
52. Spender, J.C. Identifying computer users with authentication devices (tokens). *Computers and Security, 6*, 6 (1987), 385–395.
53. Stoll, C. *The Cuckoo's Egg*. New York: Pocket Books, 1995.
54. Stoll, C. Stalking the willy hacker. *Communications of the ACM, 31*, 5 (May 1988), 484–497.
55. Straub, D.W. Computer abuse and computer security: update on an empirical study. *Security, Audit and Control Review, 4*, 2 (Spring 1986), 21–31.
56. Straub, D.W. Validating instruments in MIS research. *MIS Quarterly, 13*, 2 (June 1989), 146–169.
57. Straub, D.W., and Nance, W.D. Discovering and disciplining computer abuse in organizations: a field study. *MIS Quarterly, 14*, 1 (March 1990), 45–60.

58. Tom, P.L. *Managing Information as a Corporate Resource*, 2d ed. New York: HarperCollins, 1991.
59. Ungoed-Thomas, J. The schoolboy spy. *Sunday Times* (London) (March 29, 1998), section 5, 1–2.
60. Wood, C.C. Effective information system security with password controls. *Computers and Security*, *2*, 1 (1983), 5–10.
61. Wu, T.C., and Sung, H.S. Authenticating passwords over an insecure channel. *Computers and Security*, *15*, 5 (1996), 431–439.
62. Yager, T. Taking command of Windows NT. *Unix Review*, *15*, 12 (November 1997), pp.33–42.
63. Zviran, M., and Haga, W.J. Cognitive passwords: the key for easy access control. *Computers and Security*, *9*, 8 (1990), 723–736.
64. Zviran, M., and Haga, W.J. Evaluating password techniques for multilevel authentication mechanisms. *Computer Journal*, *36*, 3 (1993), 227–237.
65. Zwass, V. *Foundations of Information Systems*. Boston: Irwin/McGraw-Hill, 1997.

APPENDIX: Password Characteristics Questionnaire

A FUNDAMENTAL SECURITY MECHANISM IN ANY INFORMATION SYSTEM is the ability to authenticate the identity of a system user. While research continues on more sophisticated methods of authentication, password mechanisms remain the predominant method of authenticating IS users. However, despite the fact that practically every penetration of a computer system at some stage relies on the ability to compromise a password, little attention has been given to the characteristics of their actual use. This questionnaire is aimed at collecting empirical data to evaluate the characteristics of real-life, user-selected passwords and investigating possible associations among these characteristics. Please respond to the following questions without revealing your password. We thank you for your cooperation.

1. Do you use the organization's mainframe system or any of its local area networks (circle one)?
   No Yes

If no, please return this questionnaire anyway. Even if you do not use the system, we appreciate completed returns to this survey.

If yes, please continue.

2. How many characters are in your password?

3. How did you choose your password (circle one)?
   a. A meaningful detail. (e.g., name, date, street)
   b. A combination of meaningful details. (e.g., Bill1997, 4june63)
   c. A pronounceable password. (e.g., one4you, 2Bfree)
   d. A random combination of characters. (e.g., car8&t, dUCk*? +)
   e. Other (please specify)

4. What are the characteristics of your password (circle one)?
   a. Alphabetic (e.g., abdc, ERTIS).

11. How often do you log on? (circle one)
    a. Never.
    b. Annually.
    c. Quarterly.
    d. At least once a month.
    e. Several times a month.
    f. At least once a week.
    g. Several times a week.
    h. At least once a day.
    i. Several times a day.

12. Do you use any other information systems that require the use of a password? (circle one)
    No Yes

13. Do you use the same password on the other systems? (circle one)
    No Yes

Please place completed questionnaire in the self-addressed envelope provided and return as soon as possible.

Thank you for your cooperation.

b. Numeric (e.g., 1234, 5579).
c. Alphanumeric (e.g., a34d, Fo67YI).
d. ASCII (e.g., cd!Yx, Aci+t6).

5. Have you ever had difficulty remembering your passwords (circle one)?
No Yes

6. Very often, computer users find it convenient to write down their password for one of those unfortunate times when they forget it. Do you also practice this. (circle one)?
No Yes
If so, where do you write it down (users manual, calendar book, notebook, desk, drawer, keyboard, monitor, on something in your wallet, or other)?
Where:

7. How often did/do you change your password (circle one)?
a. Never.
b. Less than once a year.
c. Up to three times a year.
d. Four to six times a year.
e. About once every month.
f. More then once a month.

8. Have you ever changed your password because you felt it had been guessed by someone else (circle one)?
No Yes
If so, what led you to believe it had been guessed?

9. *Data importance* refers to the inherent value of the data to you or to the organization. On a scale of 1 to 5, how important are your data? (circle one)

1	2	3	4	5
Nonvital (not important, would not miss, life would go on)		Moderately vital		Highly vital (life-threatening, lose clients, etc.)

10. *Data sensitivity* means the degree to which problems would arise if the contents of data files were known to others. If the data were made public, what would be the impact on your organization? On a scale of 1 to 5, how sensitive are your data (circle one)?

1	2	3	4	5
Nonsensitive (nothing to hide)		Moderately sensitive (mildly embarrassing)		Very sensitive (embarrassing personally or to the organization