A User Authentication Based on Personal History
- A User Authentication System Using E-mail History -

Masakatsu NISHIGAKI
Graduate School of Science and Technology, Shizuoka University,
3-5-1 Johoku, Hamamatsu, 432-8011, JAPAN
and

Makoto Koike
Electronic Div.I, Denso Techno Co., Ltd.
1-714 Taisho-machi, Kariya, 448-0855, JAPAN

ABSTRACT

This paper proposes a user authentication using personal history of each user. Here, authentication is done by giving answers to questions about the history of user's daily life. Users do not have to memorize any password, since the passwords are what users already know by experience. In addition, everyday-life experience increases day by day, and thus the question could change on every authentication trial. In this paper, a user authentication system using user's e-mail history is shown as a prototype of our proposal, and some basic experiments to evaluate the availability of the system are carried out.

Keywords: User authentication, Personal History, E-mail history, Passwords.

1. INTRODUCTION

It is very difficult for users to memorize secure passwords (long and random strings). This is well known as the trade-off between usability and security of the password authentication system. However, as the concern about security increases in recent years, the necessity for an authentication system of satisfying both usability and security is strongly asked for. To actualize that, one of an effective approach is an authentication which uses an image as a password that human is comparatively skillful and/or easy to memorize [1][2]. This aim is to "reduce" the burden of memorizing passwords by utilizing human's memory characteristics. In this paper, the idea is developed further; the aim here is to "eliminate" the burden of memorizing passwords. Specifically, what users already know is used as a password, instead of letting users memorize new password. In other words, this paper proposes "a user authentication based on a history of user's daily life".

2. USER AUTHENTICATION USING HISTORY OF USER'S DAILY LIFE

Information that is already known
Information that users already know is classified roughly into the following two groups. One is the information specific to the user, and another is the personal history information based on user's experience.

For example, a birth date, an address, an occupation, a hobby, a favorite food, a favorite color and so on can be listed as the user's specific information. This kind of information is almost invariable and known also by a person around the user. Therefore, if it is used as a password as it is, other person may know it someday soon.

On the other hand, the following information can be listed as the user's personal history information;
- What did the user have for lunch yesterday?
- What did the user watch on TV at 8 PM last night?

We believe that passwords based on this type of information have some advantages. First, since a person's experiences increase day by day, the password space also increases day by day. Second, the answer could keep changing in everyday life, as "today's lunch" becomes "yesterday's lunch" when the next day comes. In other words, when such information is used as a password, it can be regarded as a kind of one-time password. Third, generally speaking, everybody has their private time, and usually nobody knows what the user was doing during his/her private time. So if it is used as a password, it is difficult for crackers to guess the password.

Judging from the discussion above, in this paper hereafter a user authentication using personal history of user's daily life as passwords is constructing.

Home computing and personal history information
They say, the era of the home computing is just coming, in which residential environment becomes highly intelligent. Every household appliance inside a home would be connected with a home computer, and autonomously or cooperatively under the control of the home computer they would back up our life [3].

In the home computing environment, it is conceivable that the logs regarding the condition and the use of various household appliances are being left to a home computer as shown in Fig.1. These logs can be said as "history of life of the user", since the logs are just caused by people at home using these appliances in their life. Therefore, this log information is expected to be able to use as passwords in the user authentication based on a history of user's daily life.

Figure 2 shows an example application to home security. The authentication procedure is as follows.

(1) As Bob comes home, Bob asks the authentication system to open the door.
(2) The authentication system selects one log randomly out of Bob's life history which has been stored in the home computer.

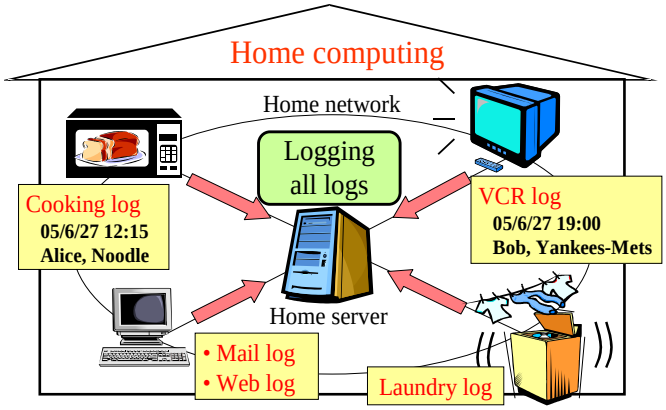

Fig.1: Log data stored in home computer.

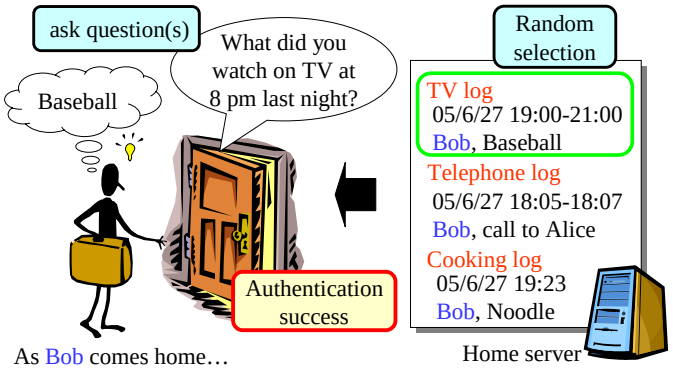

Fig.2: An application to home security.

(3) The question about the log is presented to Bob.
(4) If Bob can answer, the authentication system will open the door. (Procedures (2)-(4) are repeated, if necessary.)

Discussion

Handling of log information: In this system, it is not desirable to put the log information to the place other than the home computer, because the log information that is stored in the home computer relates to the privacy of the residents deeply. Therefore, a domestic authentication like home security as explained before is congenial to this system. When this system is used for a user authentication outside home, it seems necessary to inquire the authenticity of the user to the home computer from outside, in accordance with Liberty Alliance specifications [4].

In addition, there will be a problem that it is difficult to use the log information for user authentication of each individual, since the appliances in home are shared with the family members (except a user who lives alone). To deal with that, the use of some countermeasures is necessarily, such as, (i) the use of log information with respect to an alarm clock or a cellular phone, which are appliances of "one unit for per person" (not "one unit for per family"), or (ii) the use of log information with respect to each e-mail address since each family member usually acquires own e-mail address even though the family shares one personal computer.

Form of a question: In the user authentication using personal history, it is not well understood what kind of question is appropriate for how to generate the questions from the log information saved in the home computer. Specifically, the problems are as follows.

(a) A certain level of intelligence is necessary to generate a proper question from each log information automatically. Otherwise, all the questions must be fixed to be "What did you do at X o'clock, Y days ago?"
(b) It is impossible for people to remember all the experiences accurately, since human memory is uncertain.
(c) A person who knows the legitimate user well might be easy to cheat, since the person could understand or guess what the legitimate user did yesterday or some days ago.
(d) The security level and usability of the authentication system changes depends on the type of answer ("choose one out of alternatives" or "fill in the blanks") and the number of questions.
(e) When using "choose one out of alternatives"-type questions, alternatives of the answers for the question are presented in the authentication trial. This means that even when the impersonator was not authenticated, some privacy information leaks to the impersonator, since the right answer is displayed among the alternatives.

These problems above should be addressed in the future study.

3. USER AUTHENTICATION USING E-MAIL HISTORY

We believe that the user authentication based on a history of user's daily life could be one solution to achieve a user authentication of satisfying both usability (easy to memorize passwords) and security (difficult to have passwords cracked). However, the home computing environment has not yet available now. Moreover, it is not sure whether people really remember everything of everyday-life experience or not. Therefore, this section implements a user authentication system using user's e-mail history as a prototype system, and carries out some basic experiments to evaluate the availability of the idea proposed in this paper.

The system

The user authentication system using e-mail history is shown in Fig.3, and its procedure is as follows.

(1) The system classifies the user's e-mails into two mail boxes; the e-mails received within the last N days are put in "the recent mail box," while the e-mails received in the time before M days ago are put in "the old mail box."
(2) The system selects one e-mail randomly out of either recent or old mail box, and shows the mail body. The header information such as From:, To: and Subject: is not presented. Furthermore, if any word that shows the date

and the month is included in the mail body, it is overwritten by '+'.

(3) By reading the mail body, the user answers whether it is a recent mail or an old mail.

(4) Repeat procedures (2)-(3) i times.

(5) If the number of the correct answers is j or more, the user is authenticated.

The strategy

It is very important here that human memory is not always accurate. For instance, let us consider that the system shows you one of your e-mails of 6 days ago and asks you whether you received the e-mail within one week or not. In this case, it is almost impossible for you to give the exact answer to the system, since human memory is rather vague (Fig.4). To overcome the problem, the system here uses two thresholds N and M to exclude "vague e-mails" between the recent and the old. Let us consider again that the system shows you one of your e-mails of 6 days ago. Then, the system sets N as 7 and M as 30, and tell users "Don't be afraid. Any e-mail of from 8 to 29 days ago will not be used. So you can answer intuitively recent or old". This would help you considerably to make a clear decision to the answer (Fig.5), since you can get into the answer in the following way.

(i) Human memory is vague. Probably you can not remember the exact date of the e-mail received. However, human memory is not too vague. You will know roughly that you received the e-mail within around one week. At least you are sure that the e-mail was not received more than 30 days ago.

(ii) You know that the e-mails of from 8 to 29 days ago do not appear to the question in the system.

(iii) Judging from (i) and (ii), you can decide that this e-mail must be received within the last one week.

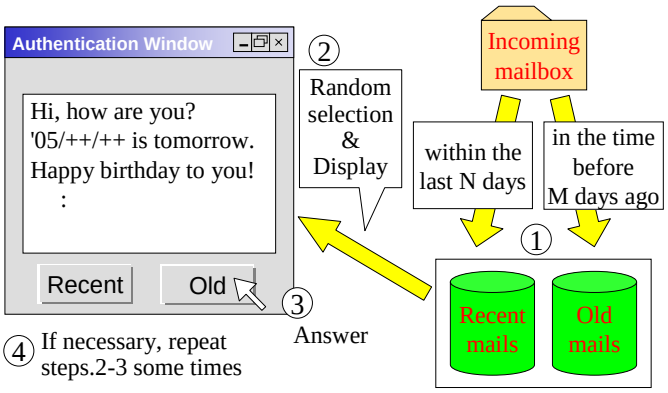

Fig.3: User authentication system using e-mail history

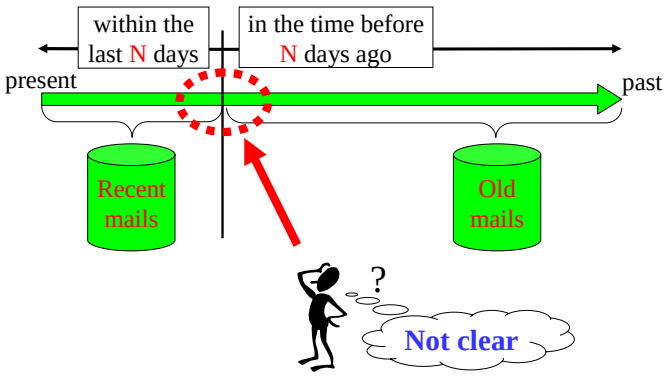

Fig.4: Users can not make a clear decision when N=M.

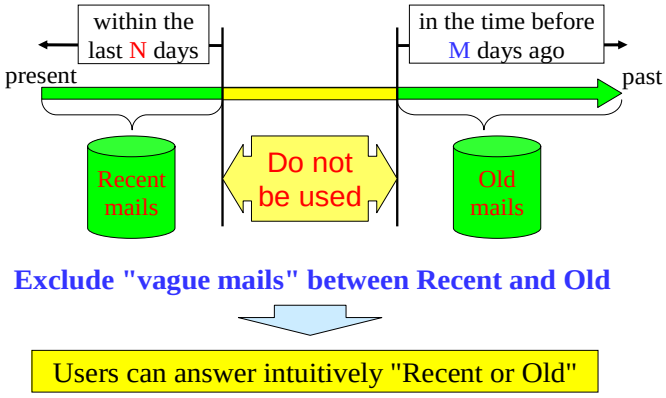

Fig.5: Users can answer intuitively by excluding vague mails between recent and old.

In addition, the system uses one more scheme to exclude "vague e-mails." Generally speaking, people would not forget an important or useful mail, while there would be nobody who remembers the spam mail and junk mail. Therefore, in this system, when the user is sorting the receiving e-mails into different folders, the system allows the user to choose the mail folders that the system is used for the authentication. In other words, the system uses only the e-mails in the chosen folders. It is considered to be effective to exclude the spam mails and junk mails by setting his/her trash folder as "Not used." To use only the unforgettable e-mails by choosing the folder for important mails would be further efficient.

4. BASIC EXPERIMENTS

Authentication by legitimate users

To examine the correct answer rate for the legitimate users (CAR_L) of the e-mail history authentication system, the authentication trial by the legitimate users is carried out. In the experiments, the correct answer rate for each e-mail is measured (in other words, both i and j in procedures (4) and (5) described in "*The system*" of Section 3 are set as 1s), as the purpose of this paper is to collect fundamental data for this kind of authentication system.

The subjects are eight male employees (A-H) of some information system companies between the ages of 26-45. The e-mails used in the experiment for each subject are shown in Table 1. In Table 1, " # of mails a day " is the average

number of receiving e-mails in a day, and "% of unnecessary mails" is the average percentage of the unnecessary e-mails (junk mails such as spam, virus and so on) out of all the receiving e-mails. "Sorting mails into folders" shows yes/no answer about the question; "Are you sorting your receiving e-mails into mail folders for each category?" For the subjects with YES answer, the further questions are asked; "Are you excluding the unnecessary e-mails by not choosing the trash folder in the experiment?" and "How many percent of the folders within all the mail folders are you choosing to use in the authentication?" These answers are given in "Excluding trash folder" and "% of mail folders used" in Table 1.

In this experiment, we set N as 7 and M as 30, i.e., the e-mails which were received within 7 days are "the recent mails", while the e-mails which were received in the time before 30 days ago are "the old mails". The system chooses an e-mail at random from either the recent or old mails of each subject, and shows the mail body to the subject. The subject is asked to answer whether the mail is recent or old. For each subject, this is repeated 30 times in an experiment with 30 e-mails randomly chosen each time, and the experiment is conducted four times with the interval of over one week between each experiment. Namely, each subject answers recent or old to 120 e-mails in total. Table 2 is the experimental result, where the CAR_L for each e-mail (how many times each subject succeeded to answer within 120 e-mails) is shown.

We think that 85% of the average CAR_L is not too bad but not high enough. From the interviews with the subjects after the experiment, it has been found that it was especially difficult to answer to the e-mails which are delivered periodically with a similar format such as mail news, mail magazines and so on.

Authentication by impersonators
To examine the correct answer rate for the impersonators (CAR_I) of the e-mail history authentication system, the authentication trial by the impersonators is carried out. Again, the correct answer rate for each e-mail is measured (in other words, both i and j in procedures (4) and (5) described in "*The system*" of Section 3.1 are set as 1s) in the experiments, as the purpose of this paper is to collect fundamental data for this kind of authentication system.

Here, the subject A shown in "*Authentication by legitimate users*" of Section 4 is supposed to be the legitimate user, and seven male impersonators (I-O) between the ages of 25-27 are trying to log on as the subject A. The subject A is a professor of a university, and the seven impersonators are graduates of the laboratory of the subject A. So, all the impersonators have known the person A well. The average numbers of receiving e-mails of I-O in a day are shown in Table 3. There are some impersonators who receive small number of the e-mails a day. We guess that they are not inadequate for this experiment, since the number of e-mails the impersonators receives in a day has got nothing to do with their attempts to impersonate A.

In this experiment, we set N as 7 and M as 30 again, i.e., the e-mails which were received within 7 days are "the recent mails", while the e-mails which were received in the time before 30 days ago are "the old mails". The system chooses an e-mail at random from either the recent or old mails of the subject A, and shows the mail body to the impersonators. Then the impersonators answer whether the mail is recent or old. For each impersonator, this is repeated 60 times with 60 e-mails randomly chosen each time from the recent or old e-mails of A. Table 4 is the experimental result, where the CAR_I for each e-mail (how many times each impersonator succeeded to answer within 60 e-mails) is shown.

Judging from the result (51% of the average CAR_I), it would seem that the impersonators could answered only by "guesswork" even though they have known the legitimate user A well.

5. AN IMPROVEMENT

From the result of the basic experiments in Section 4, it has been confirmed that the correct answer rate for the legitimate users (CAR_L) was not high enough. This system is trying to exclude the ambiguity of human memory by removing the e-mails in the period which is between recent and old. However, it seems to be difficult to eliminate the ambiguity completely.

Therefore, we are trying to add one more idea in order to even help the users to make a clear decision. The idea is to subdivide the alternatives of answer into four; "definitely new", "probably new", "probably old" and "definitely old". Then the authentication is carried out by using only the answers of "definitely new" and "definitely old".

Authentication by legitimate users
The same experiment as that carried out in "*Authentication by legitimate users*" of Section 4 but with four-alternatives is conducted. The result is shown in Table 5. (Actually, we have carried out only the experiment with four-alternative system, since it was not easy to ask the subjects to engage in our experiments for long time. Table 2 was obtained from Table 5 by viewing the answers of "definitely new" and "probably new" as "recent mails", while the answers of "probably old" and "definitely old" as "old mails".)

Table 5 says that the average CAR_L reached about 99% by using only the e-mails which the user can answer with "definitely".

Authentication by impersonators
The same experiment as that carried out in "*Authentication by impersonators*" of Section 4 but with four-alternatives is conducted. The result is shown in Table 6. (Actually, we have carried out only the experiment with four-alternative system, since it was not easy to ask the subjects to engage in our experiment many times. Table 4 was obtained from Table 6 by viewing the answers of "definitely new" and "probably new" as "recent mails", while the answers of "probably old" and "definitely old" as "old mails".)

From Table 6, it has been confirmed that the average CAR_L is still small enough in four-alternative system, since around 50% of the correct answer rate means that it seems to be "guesswork". Noted that 100 % of "CAR_I for definitely" of the impersonator J has no meaning since the number of e-mails that J can answer with "definitely" is only one out of 60.

Table 1: E-mails used in the experiment for each subject.

	A	B	C	D	E	F	G	H
# of mails a day	50	50	80	150	70	100	30	40
% of unnecessary mails [%]	60	0	70	60	1	30	1	30
Sorting mails into folders	yes	yes	no	yes	yes	yes	yes	Yes
Excluding trash folder	yes	yes	--	yes	no	no	no	Yes
% of mail folders used [%]	50	30	--	50	20	70	80	50

Table 2: Experimental result for CAR_L.

	A	B	C	D	E	F	G	H	total
# of mails used	120	120	120	120	120	120	120	120	960
# of correct answers	108	108	84	106	101	105	88	112	812
CAR_L [%]	90	90	70	88	84	88	73	93	85

Table 3: Average number of e-mails that each impersonator receives a day.

	I	J	K	L	M	N	O
Average # of mails a day	30	40	5	40	3	2	15

Table 4: Experimental result for CAR_I

	I	J	K	L	M	N	O	total
# of mails used	60	60	60	60	60	60	60	420
# of correct answers	35	25	28	37	33	31	27	216
CAR_I [%]	58	42	47	62	55	52	45	51

Table 5: Experimental result for CAR_L of four-alternative system.

	A	B	C	D	E	F	G	H	total
# of mails used	120	120	120	120	120	120	120	120	960
# of the answers of “definitely new” or “definitely old”	83	104	52	86	60	86	38	101	610
# of correct answers	83	100	49	85	60	86	38	101	602
CAR_L for definitely [%]	100	96	94	99	100	100	100	100	99
# of the answers of “probably new” or “probably old”	37	16	68	34	60	34	82	19	350
# of correct answers	25	8	35	21	48	19	50	11	210
CAR_L for probably [%]	68	50	51	62	61	56	61	58	60

Table 6: Experimental result for CAR_I of four-alternative system.

	I	J	K	L	M	N	O	total
# of mails used	60	60	60	60	60	60	60	420
# of the answers of “definitely new” or “definitely old”	16	1	40	60	48	35	29	229
# of correct answers	13	1	18	37	28	18	13	128
CAR_I for definitely [%]	81	100	45	62	58	51	45	56
# of the answers of “probably new” or “probably old”	44	59	20	0	12	25	31	191
# of correct answers	22	24	10	--	5	13	14	88
CAR_I for probably [%]	50	41	50	--	42	52	45	46

6. CONCLUSIONS AND FUTURE WORKS

This paper has proposed a concept of the authentication system using history of user's daily life which would satisfies both usability and security. This system uses "what users already know" as passwords instead of "letting users memorize new passwords". Here, the user authentication system using e-mail history has been constructed as a prototype of this system, and its availability of the system has been confirmed through some basic experiments. The experimental results have indicated that it is important to remove the ambiguity of human memory as much as possible. Although we have already gotten a certain level of the authentication rate by excluding the e-mails in the period between recent and old and by using the e-mails the users clearly remember, it will be necessary to find further knowledge so that we could even reduce the ambiguity of human memory. We believe that the accumulation of these knowledge must help us again when we eventually implement an authentication system which uses various information of the home computer and/or appliances.

The e-mail history authentication system presented in this paper has an essential problem regarding the privacy issue; the impersonators are able to read the mail body of the legitimate user when they are trying to log on as the legitimate user. In other words, anybody can read the mail body even if they failed to impersonate the legitimate user. We guess that "dummy mail injection" would be an option to reduce this problem. In the system with dummy mail injection strategy, the e-mails used in the authentication are randomly chosen from the recent mails (the e-mails received within the last N days), the old mails (the e-mails received in the time before M days ago) and the dummy mails (the e-mails that the legitimate user has never seen), then the user is asked to answer recent, old or dummy. It is supposed that the legitimate users are able to find the dummy mails, while that must be difficult for the impersonators. We have to begin with to address that in the future.

7. REFERENCES

[1] R.Dhamija and A.Perrig, "DejaVu: A User Study Using Images for Authentication", 9th Usenix Security Symposium, pp. 45-58 (2002).

[2] T.Takada and H.Koike, "Awase-E: Image-based Authentication for Mobile Phones using User's Favorite Images", Fifth International Symposium on Human Computer Interaction with Mobile Devices and Services (Mobile HCI 2003), 2003.

[3] I.S.Tenidis, "Intelligent Home", 39th European Telecommunications Congress (FITCE 2000), 2000, http://www.intracom.gr/downloads/pdf/news/publications/2000/icom_doc012_fitce2000.pdf .

[4] Liberty Alliance Project, "LIBERTY ALLIANCE PROJECT", http://www.projectliberty.org/ .