Studying Password Use in the Wild: Practical Problems and Possible Solutions

Philip Inglesant
Department of Computer Science
University College London
Gower Street, London WC1E 6BT, UK
p.inglesant@cs.ucl.ac.uk

M. Angela Sasse
Department of Computer Science
University College London
Gower Street, London WC1E 6BT, UK
a.sasse@cs.ucl.ac.uk

ABSTRACT

HCI research into usability and security over 10 years has repeatedly found that users are unable to cope when faced with unusable password policies. Yet to show the full impact of these policies, it is necessary to consider the context of use within the organisation. Password requirements which users cannot meet have a cost in terms of impact on users' primary task and, hence, loss of productivity. Conversely, organisational practices determine the numbers of passwords and the frequency of use. Retrospective accounts, questionnaires, and experimental methods fail to capture the full context of use.

We present our experiences from the use of a study which was designed to overcome these shortcomings. We devised a structured diary study of password use followed by detailed debrief interviews. We found that this study effectively elicited participants' main password uses and brought to light details of the context of use. However, the study did not capture accurate measures of workload or time taken in password use; these are better measured through other methods. Finally, our research leads us to conclude that there are further impacts of passwords in the workplace which can only be fully understood from richer ethnographic methods.

Categories and Subject Descriptors

H.5.m. Information interfaces and presentation (e.g., HCI): D.4.6 Security and protection, K.6.5 Authentication.

General Terms

Experimentation, Security, Human Factors.

Keywords

Diary studies; passwords; ethnographic studies; semi-structured interviews

1. INTRODUCTION

Passwords interrupt the smooth flow of work; coping with these interruptions calls on other skills and knowledge on the part of password users. Indeed, it is partly the intention of passwords to act as "barriers" to action [6], and the challenge is to minimise the impact of these barriers for legitimate users while maintaining them for those who should not have access.

Symposium On Usable Privacy and Security (SOUPS) 2010, July 14-16, 2010, Redmond, WA, USA.

Our goal is to reduce some of these barriers and to understand the true impact – the user cost - of security policies, and how users cope with them to complete their primary tasks.

Users' ability to develop methods for devising, remembering, and using multiple passwords is an unremarked but remarkable skill which employees have mainly learned from improvisation, from pieces of often conflicting advice, and from talking to colleagues. However, these coping methods are symptomatic of failures in password policy, rather than a justification for ever more stringent password requirements [1, 10].

In coming to this understanding, laboratory studies of human behaviour in HCI can be criticised for being too distinct from the normal context of use. Ethnographic approaches based on observation or participation enable the researcher to study and document work practices in everyday settings [5]. However, as in our case, traditional ethnographic methods such as observation are often impractical or too invasive in the workplace, for various reasons.

We present an approach which we took to capture, as far as possible, the contextuality of the ethnographic approach within the practical limitations of the resources and access to participants available to us.

1.1 The Ethnography of Passwords

Innovative technologies have signally failed to achieve universal usable security and privacy. Passwords remain the most widely-used form of personal authentication, despite many promising alternatives.

Understanding the reasons for this demands a consideration not only of the technologies, but also of the practices into which these technologies are incorporated [6]. This all the more so where, as in the case of passwords, the technology is simple but remains very widespread.

Even though typing a password is a relatively bounded act, it nevertheless exists within a context of social criteria and values [5]. Only by understanding these factors can the full impact of password use be understood.

For example, the basis of organisational password policies rests on assumptions, often unstated, about the perceived need for a given level of security.

At the same time, for users the focus is not on security but on their primary task, which in the workplace means getting their work completed, often under time pressure.

Table 1: Participants in the diary & interview study

Organisation A	
A large research-intensive university; participants are administrators, and lecturers or researchers in disciplines removed from Computer Science and HCI, and teaching staff:	
Administrative staff	9
Lecturing and research (all apart from 1 are non-Computer Science/HCI)	6
Organisation B	
A large financial services organisation. Participants are members of a security team and Human Resources administrators. The security team members are of interest to us as a sample of more security-aware users	
HR Administration	5
Security team	12

1.2 Reasons for the Study Design

Sensitivity considerations, especially in financial services meant that in our case, video recording was not an option.

Ethnography would suggest traditional pencil-and-paper observation. We believe that this could be very effective; however, an initial trial was found to be too intrusive by participants.

We therefore chose a structured diary study as a way to gather "*in the wild*" [6] data, in an objective and generalisable way, but with minimal disruption and interruption to participants' daily work. It is important to stress that laboratory, ethnographic, and diary studies are not alternatives methodologies but are best used in combination, for a rounded view of the questions of interest.

Diaries rely on user-generated data and therefore depend on the reliability of the participants. We considered other forms of quantitative data which could in some cases be obtained from non-user sources, such as analysis of system logs, or direct methods such as client software to record aspects of password use, such as was done by Florêncio & Herley [7]. However, we wanted to build a full picture of password use in the workplace, and neither of these methods was available to us across all applications. These less user-dependent forms of data capture could nevertheless have been useful to give a measure of the accuracy of the diaries.

The study produced some important substantive findings which we have presented in another paper [8]. In this paper, we expand on our experiences – with reflective self-criticisms – of the approach we took.

2. DESIGN OF THE DIARY STUDY

Diary studies have become an accepted research method in HCI, using either paper or other methods of recording [3, 9]. In the sense that we used the diaries prior to a debrief interview, our method is similar to Brown et al. [3]; we used diary capture initially, which then supported richer data collection through in-depth debrief interviews.

It is important to be clear, however, that in our study the diary was highly structured, with only limited free space fields in which participants were invited to record free-text descriptions of the background task and the specific reason for the password use. By choosing a diary study, we were choosing a level of research with a far smaller set of participants than large-scale studies such as Florêncio & Herley [7]. On the other hand, we were able to asks why users use particular passwords; it is far more in-depth and, in a positive way, more subjective.

The process of keeping the password diary does not consist only of the period while participants are actually keeping the diary. Our data gathering had three parts:

1. Introducing the requirements of the diary to the participant; this is a key element in success
2. The participant kept the diary for one week, with regular contact from the researchers
3. A final debrief asked structured questions about each of the passwords, and allowed for unstructured probing about the circumstances of password use and specific incidents recorded in the diaries.

2.1 Two Contrasting Organisations

Because we were concerned with organisational password policies, we recruited participants from among staff members within two organisations. Over a period from December, 2008 to August, 2009, we recruited 32 participants who each kept the diary for 4-5 working days – see Table 1. Participants were all volunteers and were compensated with gift certificates.

2.1.1 Diary-Keeping Only in the Workplace

We were interested particularly in organisational password policies and their impact on primary tasks. It is also important to avoid invading the privacy of participants by investigating non-work related issues. However, there is a large interplay between work-oriented and non-work oriented password use. The same password may be used for different purposes, or the same service, such as airline booking or Webmail, may be used for both personal and work-related purposes.

For this reason, we asked our participants to record *every* password use during their *working day*, including non-work use, for one week. Some participants routinely work from home, or outside normal office hours; we asked them to include passwords during what they regard as "work" time, regardless of time or location.

2.2 Three forms of "In the Wild" Data

Our collection methods yielded three distinct forms of data:

2.2.1 Qualitative data from the diary studies

Unlike conventional diary studies, our diaries were highly structured. Nevertheless we chose to keep the background task and the specific reasons for use as free-text fields; the space available was a hint to participants that around 2-8 words were expected.

The use made of this field showed, as expected, many commonalities, but was also a strong reminder of the wider context of password use. For example, "*Coming back from lunch*" or from a meeting; starting work; or logging into a portal.

The free-text fields showed the impact of different password practices in the two organisations: in one, users are required by policy to lock desktop PC's every time they move away from their desks, even for a short period, but complete shut-down is generally only required at weekends (when upgrades are often run by systems staff). In the other organisation, by contrast, there is no such rule, but participants habitually close down their desktop PCs overnight. On the other hand, in this organisation, staff are mainly working on internal online services, which in some cases must be accessed via a virtual desktop, so that logging onto the local desktop PC is only one part of the initial daily login process.

2.2.2 Qualitative data from the debrief interviews

As soon as possible following the completion of one week's diary-keeping, participants were interviewed in depth in a debrief interview about their password use. The interviews were highly structured around a questionnaire, which ensured that each of the passwords was discussed fully, but participants were encouraged to expand on their responses.

The debrief interviews were voice-recorded and fully transcribed to capture the richness of the data. We recorded 17.4 hours of interviews from these participants. Although this is a relatively small sample size, it is appropriate for an in-depth analysis of diaries and qualitative data.

Contrary to our initial expectations, we found that the debrief interviews were the most useful source of insights into password use. We chose to make the password, rather than events in password use, as the unit of study; this allowed us to expand on the use of the password as it fits into the wider context, since reasons for use were given in the diaries. This was partly for practical reasons, since, even with this higher level of focus, interviews lasted between 30 minutes and one hour (about 10-15 minutes for each password) – we restricted ourselves to one hour as the maximum for a single interview.

In this sense, the password diaries themselves were not successful as a way to record password interruptions, but were useful as a record of which passwords are actually used in working life. There are other useful lessons to be learnt for future diary studies, which we discuss this in more detail below.

However, although the debrief interviews were highly structured, they yielded rich qualitative data *around* the open questions. For example, by asking participants about the sensitivity, in their perception, of the data or services being protected, and scaling this on a 5-point Likert scale, we also elicited the reasons why they might consider them to be sensitive, and the "bad things" which might happen if a password was to be compromised.

2.2.3 Quantitative Data

Our diary study provides data which are amenable to quantitative analysis, for example around the frequencies of various kinds of password events. For two participants, we found that the diary-keeping was too unreliable and discarded it from the quantitative analysis, but retained the qualitative interviews.

The diaries generated 982 password events involving 196 distinct passwords. Some of these distinct passwords are identical to others, either by constraint of the system architecture or by users' choices; taking this into account gives 137 unique passwords.

Both the diaries and the debrief interviews yielded large amounts of quantitative data about the nature of the passwords, their frequency of use and of changing, classifications about how passwords are chosen and reasons for use, and other factors.

However, it was not possible to make meaningful correlations between, for example, strength of the password and sensitivity of the data being protected, because of enforced strength rules; particularly in Organisation A, passwords are constrained according to policy-makers' views of what makes a good, strong password.

But this rather negative finding is in reality a useful finding in itself, since it shows a tension between the enforced password rule and the perception of what is necessary, and what is humanly possible in terms of memorability:

> *... it's got to the point where it's so difficult to remember, so difficult to make one up, and difficult to remember, that I have to write it down until I've learnt it, but then we're asked to change it again, so I have to write the new one down cos I can't remember the new one.*
>
> *So, I think it gets to a point where they restrict it too much, it makes it less secure – Administrator, Organisation A*

2.3 Analysis Using Grounded Theory

We analysed the transcribed recordings using a variant of Grounded Theory [4], using Atlas TI [11] to aid our qualitative analysis. Because we were interested in a number of dimensions in password use, rather than attempting to find one core category, we developed codes around the following categories:

- Factors impacting on perceptions of sensitivity of the protected data or service
- Estimated strength of the password (we also developed a measure of strength from entropy)
- How participants cope with the demands of policies
- The "ecology" or context surrounding password use

We also coded for a similarity between passwords, which had emerged as an issue during the interviews; in one organisation, many passwords were *constrained* to be identical by the architecture of the authentication system, whereas in the other, there was a true single sign-on password for most organisational applications.

From the diary data free-form field, we also had a rich set of reasons for password use; a total of 436 different entries, which we categorised to reflect commonalities, as we interpreted them. A small sample of these uses, and the 15 "generic" reasons by which we categorised them, is shown in Table 2.

3. HOW THE STUDY INFORMED OUR RESULTS

Recall that our main interest is in the *impact* of password policies within organisations; thus, we emphasise the organisational context as well as the policies, rather than considering passwords in simple generic terms.

Participants were remarkably open about discussing their passwords. It was important that we made clear to them that interviews were in confidence, and that any findings will not be personally attributable.

3.1 Findings from the Study

By way of illustration of the positive results from our methodology in throwing light on passwords in use, we discuss here some of our key findings. For a fuller presentation of these findings, see [8].

3.1.1 Wider Use of Passwords

One clear finding, which is often overlooked in studies which do not take the full context of password use into account, is that passwords are used for a wide variety of different uses; not simply for authenticating to services, but also as for uses such as for securing files. These are often ad-hoc and used in situations, such as sending files through email, where users are aware of a potential risk, but are not aware of more reliable means by they might overcome the risk. It is interesting that this use differs from that found by Dourish et al [6], who found users obscuring their email in other ways, such as by using a vague subject line or using institutional means to secure data by legal notices in email signatures.

Other emerging uses of passwords include the increasing use of "personal work-related" passwords. By this we mean, applications such as the use of social networking for work-related contacts, or the use of personal email for work uses.

3.1.2 The Ecology of Password Use

Answering the "bigger questions", such as "why are you using this password *at all*"? requires looking at the password use in context, such as preferences for particular email services, the need to use specific services to complete a work task, and conflicts between users' needs and password policies.

For example, in one of the organisations, some of the internal services can only be accessed by first logging in to an online virtual workstation. This provides a fully-functional working environment, which for some users is their main digital working space. For such users, this largely overcomes differences between working at home or in the office. Indeed, it adds an extra layer of security for home working. For others, however, the conventional desktop PC remains as the main workspace, and so logging on to the virtual workstation is simply an extra step before the required service can be accessed.

Finally, but significantly, we also found that insecure practices, such as writing passwords down in plain text in places which could be accessed by others, are still prevalent. This is a good example of a finding that can only be captured by asking people directly.

4. OVERCOMING METHODOLOGICAL ISSUES

The password diaries were found to be very useful as a way to capture all the passwords people actually use (or use within the workplace); much more reliable than simply remembering them retrospectively. However, to achieve satisfactory data gathering while minimising the impact on participants required a number of iterations of the diary design.

4.1 Some Problems with the Structured Diary

4.1.1 Balancing Richness with Participant Compliance

Conventional diary studies are very time-consuming, not only for researchers but also for participants. Since passwords are an interruption in themselves, we regarded it as important to avoid, as far as possible, adding extra interruptions to participants' primary tasks.

Table 2: Sample Reasons for Password Use

Reason for password use	Generic reason for use
About to continue with day's work	Unlocking PC or other device
About to login for the day - unlocking PC	Logging-on to PC and/or domain or other device - routine
About to read email	email - internal
Access database - compensation details secure	Local file protection
Access database - data is confidential	Local file protection
Access organisational calendar - add meetings	Authentication for online service - company internal
Access to internal DRM	Authentication for online service - company internal
Access to policies – policy service	Authentication for online service - company internal
Access website - look at financial news	Authentication for external service
Accessing application	Authentication for external service
Accessing application for list of work queries	Authentication for external service
Accessing case management application	Authentication for online service - company internal

To achieve this, we designed several iterations of a structured form to minimise interruption and to guide participants to provide the data in which we were most interested. We asked participants to record each password use against a short nickname for the password, which they chose for themselves. Then, in the debrief interviews, we categorised each password according to its main uses, as reported by the participant.

One early "tick-sheet" version of the structured diary was successful in achieving compliance, but at the cost of almost no information about the services being accessed. This was particularly problematic where participants recorded the same password used for multiple applications as one password (an ambiguity which we discuss below).

On the other hand, the free-form text fields in which participants recorded the reason for password use provided much more richness, but also raised the problem that participants would

record the reason at different contextual "levels"; for example, "*starting work*" or "*logging onto PC*" are both reasons for password use, but the larger task provides the context for the more specific one.

To overcome this ambiguity, we included some examples on the form, but this raises the new risk of biasing the data towards the examples given rather than the participants' actual experience.

As a final consideration of the recording method, we also offered participants the choice of recording their diary using a voice-recorder, making very short recordings for each password event. This option was chosen by only one participant, however. We speculate that, in a noisy office, talking into a voice recorder might be considered too intrusive, and that participants might prefer not to risk being overheard as they record their password diary.

4.1.2 Reliability of completion

We suspect there is under-reporting of password problems, especially of re-tries where the final result was success; it seems that participants are so used to re-typing a password that they hardly even register it.

Accurate recording is especially problematic during busy periods – which are particularly interesting occasions for password research. Users sometimes think they can fill the diary in later from memory; they cannot. It is important to make this clear to participants.

As well as the space for participants to describe in free text what they were doing during password use, the diary asked them to estimate the disruption, on 5-point Likert scale, and the time taken in seconds. However, these measures, while interesting for comparison, are very subjective; they do not provide an accurate measure of the time actually taken (which would be better measured experimentally), and participants often did not perceive interruptions, unless there had been some major problem.

4.1.3 The same passwords used for many services.

We needed to know which service was being accessed, but we also needed to know in which cases the same password was used for several services. This should be uncovered in the debrief. We found in practical use of the diary that this raised a problem of consistency; without guidance, some participants chose to record different services with identical passwords as different passwords, while others recorded them as the same one.

Ambiguity is more likely if identical passwords are enforced by the architecture (single-sign-on or a single password). It is important to make the distinction, because participants' perceived sensitivity of the password may be different for the different services being protected.

4.2 Overcoming Problems in Diary Studies

We relied on the diaries as a cue to the debriefs. It was all the more important, therefore, for them to be completed as accurately as possible, and to be available to the interviewer during the interview.

Unless the diaries are available before the debrief is conducted, it is easy to miss passwords, especially as there is pressure to maximise the use of volunteer participants' time. For the same reason, we chose in general, not, to follow up specific incidents, such as a password failure, during the debriefs.

4.2.1 Prepare Initial Instructions for Participants.

Wherever possible, instruct participants face to face. Doing this in a group is acceptable, as long as there is an opportunity for participants to ask questions. Where this is not possible, then participants must be followed up early in the diary study. This initial instruction is also the opportunity for some informal observation of the workplace and the participant's context of use.

Visit participants at the end of the first day whenever possible to check for any misunderstandings. In addition to this, keep in close contact by phone or email. A freeform diary study would involve daily debriefs; this is not necessary with a very structured dairy, but close contact with participants is nevertheless essential. There is only one chance to capture this data, so poorly completed diaries are wasted and cannot be re-done.

4.2.2 Pilot the Diary, as Many Times as Necessary.

Pilot the diary with a sample of "real" users; check especially for ambiguities. In the pilots, we realised the importance of ensuring that tick-items have options for "yes" and "no" (or "success" and "failed" etc.), rather than simply one tick-box for negative events. Otherwise, if a box is unchecked, there is no way to know if this means "ok", or that the participant simply forgot to check it.

Once the diary is designed, following the pilot studies, resist the temptation to "improve" the empirical instrument once the data-capture has started. As well as potentially reducing rigour, this introduces the risk of new problems and ambiguities. An example of such ambiguities is the case of a participant who completed the field "*Why are you using this password?*" by responding "*because it is the correct password*" – interpreting the field to mean "*Why are you using <u>this</u> password (rather than some other password)?*".

4.2.3 Allow Plenty of Time for the Debrief Interviews

Debrief time of 10-15 minutes for each password is reduced a little where passwords are identical, but even in those cases it is still necessary to enquire about services for which each password is used, and the level of sensitivity and the circumstances surrounding each.

Rather than regarding this as a problem, this is indicative of the very rich findings from even this very structured form of interview. Failing to allow enough time to capture this richness would limit the valuable opportunity to discuss attitudes and elicit narratives around password use.

4.2.4 Careful Recording of Results

It is worth giving careful thought to the recording of the results, both the structured quantitative data, and the recordings of interviews.

We recorded the diary and the quantitative details of passwords in a structured Microsoft Access database, which gave more flexibility than a spreadsheet. Although initial set-up of the database is more complex, this allows analysis in many different ways, and forms can be designed to make data entry very fast.

To ensure confidentiality - necessary to gain the confidence of participants and as basic ethical good practice - there is a need to secure the recordings and transcripts from unauthorised reading; this can be done simply by storing on an encrypted drive or in password protected files. Participants and their associated files are anonymised, with identities known only to core researchers. The same level of security is necessary when files are in transit; for example, an encrypted USB drive should be used.

5. REFLECTIONS ON EFFECTIVENESS AND LIMITATIONS OF THE STUDY

5.1 Diary-Keeping as Classification

In the design of the diary form, we invited participants to record events according to pre-defined categories in terms of location (home, office, mobile, or other); numbers of attempts; and whether the password is written down, or is automatically entered by a browser. We pre-defined error types: with passwords and/or the user Id, leading to re-tries, total forgetting, or to resets by a helpdesk.

Having allowed a free-text field for participants to describe the reasons for password use, we then categorised these reasons into a much smaller set of "categories" of use, for various quantitative analyses.

In these ways, we implicitly assumed that all passwords and events around passwords could be uniquely and unambiguously identified. Yet were aware that in doing this we were pre-making *classifications* into which local, circumstantial, unpredictable real-world events were to be made to fit..

This well-studied issue in society and technology stems from the seminal work of Bowker and Starr [2]. Categorisations are the basis for analysis in research, for making sense of the world, but are also inevitably reductionist and hide implicit assumptions. Used carelessly, categorisations valorise some viewpoints while silencing others; in our research towards usable security, we must seek to ensure that the viewpoints we strengthen are those of the *users*, rather than our own or which are inherent in the design of the technology.

5.2 Insights to improve future studies

Although we have made some interesting and innovative findings from our study, we are aware that our study was made under some practical constraints.

To the extent that we were able to overcome these practical difficulties, we offer this paper as a methodological contribution showing novel techniques in collection and analysis of a broad qualitative and quantitative dataset.

On the other hand, to the extent that we wish to avoid reductionism, we believe that there is much work still to be done to address the full impact of unusable password policies. We suggest that this should take the form of deeper ethnographic studies.

Referring again to an earlier study which provides inspiration for our own, Brown et al. [3] have shown a way to combine diaries with photographic recording. Now that video recording is increasingly easy and affordable, this provides an exciting possibilities for further research. Suchman [12] showed the potential of detailed analysis of video recordings in a controlled setting, but this is now a realistic possibility for understanding routine action which is not only situated, but is recorded in the context of use.

6. REFERENCES

[1] Adams, A. and Sasse, M. A. (1999) Users Are Not The Enemy, Communications of the ACM 42 (12 December), 41-46. DOI=http://doi.acm.org/10.1145/322796.322806

[2] Bowker, G. C. and Star, S. L. (1999) Sorting things out: classification and its consequences, The MIT Press, Cambridge, MA, USA

[3] Brown, B. A. T., Sellen, A. J., and O'Hara, K. P. (2000) A Diary Study of Information Capture in Working Life In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00, The Hague, The Netherlands April 2000), 438-445. DOI=http://doi.acm.org/10.1145/332040.332472

[4] Charmaz, K. (2006) Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis, SAGE Publications, London, UK

[5] Clancey, W. J. (2006) Observation of Work Practices in Natural Settings, Ch. 8 in The Cambridge Handbook of Expertise and Expert Performance, Ericsson, Anders K (Eds.), Cambridge University Press, New York, NY, USA, 127-146

[6] Dourish, P., Grinter, R. E., Delgado de la Flor, J., and Joseph, M. (2004) Security in the wild: user strategies for managing security as an everyday, practical problem, Personal and Ubiquitous Computing 8(6), 391-401. DOI=http://dx.doi.org/10.1007/s00779-004-0308-5

[7] Florêncio, D. and Herley, C. (2007) A Large-Scale Study of Web Password Habits In Proceedings of WWW 2007 (Banff, Alberta, Canada May 2007), 657-666. DOI=http://doi.acm.org/10.1145/1242572.1242661

[8] Inglesant, P. G. and Sasse, M. A. (2010) The True Cost of Unusable Password Policies: Password Use in the Wild in Proceedings of the 28th international conference on Human factors in computing systems (CHI 2010, Atlanta, GA, USA April 2010), 383-392. DOI= http://doi.acm.org/10.1145/1753326.1753384

[9] Palen, L. and Salzman, M. (2002) Voice-Mail Diary Studies for Naturalistic Data Capture under Mobile Conditions in Proceedings of the 2002 ACM conference on Computer supported cooperative work (New Orleans, LA, USA November 2002), 87-95. DOI=http://doi.acm.org/10.1145/587078.587092

[10] Sasse, M. A., Brostoff, S., and Weirich, D. (2001) Transforming the 'weakest link' - a human/computer interaction approach to usable and effective security, BT Technology Journal 19 (3 - July), 122-131. DOI=http://dx.doi.org/10.1023/A:1011902718709

[11] Scientific Software Development, 2006 ATLAS.ti The Knowledge Workbench, Berlin, Germany

[12] Suchman, L. A. (2007) Human-Machine Reconfigurations: Plans and Situated Actions, 2nd Edition, Cambridge University Press, New York, NY, USA