Evaluating the evaluation criteria for account-recovery procedures in passwordless authentication

Manuel Keil\textsuperscript{1,3} and Alf Zugenmaier\textsuperscript{1,2}

Abstract: Passwordless authentication avoids the weaknesses of password based authentication such as guessable passwords and password reuse. However, when passwordless authentication becomes impossible for the user, e.g. due to loss of the security token, an account recovery method has to be used. Kunke et al. [Ku21] analysed these recovery mechanisms in respect of criteria they extracted from the literature. However, these criteria in the literature were based on researchers’ opinions and were not grounded in practical experience.

To achieve this grounding, semi-structured interviews were conducted with practitioners in various industries. These experts were asked to rate the existing criteria and contribute additional criteria if required. The result is a weighted list of criteria that can be used in future to evaluate account recovery procedures.

Keywords: passwordless authentication, account recovery, requirements evaluation

1 Introduction

Password-based authentication is vulnerable to a multitude of attacks, including phishing and dictionary attacks [PMA22; Ra12]. Initially password managers were deployed to mitigate these vulnerabilities[Bo12]. Password managers pose problems when they are to be used on different devices by the same user, as a synchronization of the password manager must be implemented or a device with the password manager needs to be carried at all times [Bo12]. The centralization of online password managers introduces additional security risks [Ar23]. Two-factor or multi-factor authentication making use of the factors knowledge, ownership and biometrics also mitigates the vulnerabilities of password based authentication, but still has the same usability issues. The logical next step is to move to passwordless authentication, which eliminates the factor knowledge. However, passwordless authentication shares a problem with password based authentication: it still requires a process for recovering account access in case of lost hardware tokens or biometric feature changes, e.g. due to injury. Processes for account recovery already exist, and a few of them have been specifically designed for passwordless authentication, e.g. the recommended practices published by the FIDO alliance, like using multiple authenticators per account [GLS19]. This is necessary because the security advantages of passwordless authentication should not be undermined by the recovery process [Ku21]. According to Kunke et al. [Ku21]

\textsuperscript{1} Munich University of Applied Sciences, Munich, Germany,  
\textsuperscript{2} Concordia University, Montreal, Canada,  
\textsuperscript{3} ARCA-Consult GmbH, Pfaffenhofen a. d. Ilm, Germany,
none of the established methods fulfill all their requirements, either in terms of usability or in terms of security. Kunke et al. draw their requirements and evaluation criteria from prior publications, which are mostly based on those defined by Bonneau et al. [Bo12], i.e. defined by researchers. As passwordless authentication is gaining traction in business contexts, the question to be addressed here is whether the requirements set out more than 10 years ago are still relevant and complete for business use cases.

To address these issues, a mixed methods study was conducted with experts from various industries to verify known criteria, identify new ones, and examine existing criteria definitions. The wide range of interviewees ensures applicability to a broad spectrum of industries. The results of these interviews were analyzed using a qualitative, category-guided text analysis [Ma10] to determine which of the already known criteria are no longer relevant for use in business environments, which criteria are still considered in principle but no longer fall under the definition of Kunke et al. [Ku21], and which additional, not yet documented criteria are of relevance. This paper also proposes further categorization for the catalog of criteria into general and optional criteria.

2 Theoretical Background

2.1 Definitions

The definitions used in this paper for passwordless authentication and account recovery are analogous to the ones given by Kunke et al. [Ku21]. For convenience of the reader, these are given in Appendix A. Likewise, the definitions of the three categories into which criteria are grouped, namely usability, deployability, and security, agree with those used by Bonneau et al. [Bo12] and Kunke et al. [Ku21], and are also given in Appendix A.

As this work is based on the evaluation criteria defined by Kunke et al. [Ku21], the reader is referred to that paper or to Appendix A, where these are summarized.

2.2 Related Work

The most important studies on this topic are the aforementioned ones by Kunke et al. [Ku21] and Bonneau et al. [Bo12], which defined criteria for evaluating authentication and recovery methods as well as proposed practices for evaluation. The criteria mentioned by those studies seem plausible, but are solely defined by the authors of those papers and not backed by input from practitioners. Nevertheless, their results and methodology for evaluation as well as the definitions for criteria are the basis for this study. Kunke et al. [Ku21] already collected possible criteria from multiple sources and reduced previously mentioned criteria to those related to account recovery with passwordless authentication. Other earlier papers like those by Saltzer and Schröder [SS75] Nielsen [Ni94], and Stajano [St11] proposed and evaluated criteria for security and usability of processes.
Gerlitz et al. [Ge23] measured deployment of various recovery mechanisms. They used their experience from these measurements to suggest enhancements to existing recovery mechanisms. However, they stop short of defining actual criteria. This paper and another by Amft et al. [Am23], published almost concurrently, are very similar, as they both measured the deployment of recovery mechanisms and provided recommendations. Li et al. [Li22] are proposing a new recovery scheme, and are evaluating it in respect of security and performance. Because account recovery should be a procedure that is only rarely invoked, the performance evaluation seems superfluous. Wahab et al. [Wa21] also propose inclusion of keystroke dynamics into the recovery procedure, and analyse it in respect of false positive and negative rate.

The closest work to this work is probably the thesis by Tiller [Ti20], which performs an online user study with questionnaires to determine user preferences of recovery methods for two-factor authentication, and in which correlations between preferences and various demographic factors are explored. The preferences in this work are structured slightly differently, and are often specific to individual recovery authentication methods, but are also covered by the criteria used in this paper. One novel preference from the two open questions is to not be intrusive to an emergency contact. This requirement is also specific to a particular recovery method and wouldn’t be applicable in an enterprise setting.

It should be noted that in all the above work, the focus is on recovery for individual accounts. Recovery from mass compromise as analysed by Fritsch [Fr23] is not considered here.

3 Methodology

This study aimed to validate the criteria for the evaluation of account recovery by conducting interviews with experts in industry responsible for the fields of IT administration, digitalization, and controlling. These semi-structured interviews were set up according to guidance given in Helfferich [He22]. The experts were in positions that would be involved in procurement or deployment decisions on authentication solutions. The selection of the interview partners was trying to cover the three fields as well as trying to cover a variety of industries. The study aimed to avoid bias, by searching on Google for companies in relevant industry and size and then addressing the relevant person, and by searching in LinkedIn for potential interview partners, aiming at covering the different fields and industries. There may still be some self selection bias, as the response rate for interview requests was about 10%.

The interviews were conducted between 27 October 2023 and 06 December 2023 with seven experts from three different fields in six companies in Germany. The interviews were conducted in different industries, including education, consulting, manufacturing, and banking. The size of companies ranged from 20 to over 100 000 employees. The interviews

---

4 This is on the low end of interview partners according to Creswell [Cr98], but still within the given range of 5-50 participants for qualitative research.
were conducted online as well as in person, with the majority of the interviews conducted via phone conference. With consent of the interview partners, the interviews were recorded and transcribed using an audio recorder and transcription software. The transcripts were later manually corrected and formatted.

The interviews serve to answer the research question of which criteria are relevant for the evaluation of account recovery processes for passwordless authentication in companies. The analysis is conducted using a mixed methods approach. The qualitative part is analyzed using category-oriented text analysis, focusing on the criteria that decision-makers in companies consider relevant, and can influence the final decision for a new procedure. The interview partners' responses are coded to provide additional information and a simple evaluation. The coding was performed by one person, with a second person performing spot checks to validate the coding. In the quantitative part the experts evaluated the relevance of the existing criteria on a 5 point Likert scale.

The interview process was structured according to a questionnaire which was made available to the participants ahead of time, so they could understand the definitions of the existing criteria. In part one of the interview, the interview partners were asked to describe their own experiences with account recovery, and to describe the current process for account recovery in their company. The interview partners were then asked if they have any experience with or whether the company is using or is planning to switch to passwordless authentication. The interviewees were also asked about optimizations that can be implemented to improve the process.

In part two of the interview, the criteria selected by Kunke et al. [Ku21] were evaluated and criteria are expanded upon. Each of the criteria was rated and the rating was usually given with justification. The scale for evaluation was from 0 to 4, with 0 indicating the criteria is irrelevant, and 4 indicating high relevance. As interview partners also gave a reason for their rating it was possible to determine whether the rating would be specific to the company or industry or whether they would need to be redefined to fit within the current requirements for authentication and account recovery.

The questionnaire was pre-tested to gauge the duration and whether there are issues with the questions which led to improvements before the actual interviews. The results from the pre-test were not included in the final analysis.

4 Results

4.1 Interview part 1: experience

The first part of the interview, in which experts were asked about current processes and their experiences with passwordless authentication and account recovery showed that:
1. The most commonly used form of account recovery at the moment is helpdesk-based, meaning the user is not able to recover their accounts on their own. In one case, the procedure for authenticating to help desk personnel is highly complex, even requiring governmental documents.

2. Most of the respondents would prefer a form of self-service for their account recovery procedures.

3. Passwordless authentication is rarely deployed in corporate environments. Where it is used the most common forms are platform authenticators like mobile phones or Windows Hello. However, most of the respondents have personal experience with passwordless authentication.

4. Companies, that do not offer passwordless authentication themselves but are using it externally usually do not have account recovery procedures defined for accounts with passwordless authentication.

5. Where security is of the essence, password managers, Single-Sign-On and multi-factor authentication are implemented as a more secure alternative to passwordless authentication.

For example, a statement supporting item 2 given by an interviewee when talking about his experience with passwordless authentication would be:

"[self-service] would of course be more practical, because as I said, the [help desk] resources would be wasted, which in the end, [...] will cost something" (Translated from German)

4.2 Interview part 2: relevance of criteria

In the second part of the interview, the experts evaluated the criteria defined by Kunke et al. [Ku21] and have a chance to propose what they believe to be missing. Table 1 provides the average importance of the individual criteria together with the responses standard deviation, which indicates how well the experts and the environments for which they would evaluate such a recovery mechanism are aligned. The scale to rate the criteria was: 0: irrelevant, 1 low relevance, 2: medium relevance, 3: relevant, and 4: very relevant.

To keep the length of this paper in check, we will only discuss the most and least important criteria further, as well as some honorable mentions.

- Resilient-to-Phishing was rated as very relevant by all interviewees, saying that it should certainly be able to withstand a phishing attack, attack surface in general should be minimized and that users don’t always have the experience to recognize such an attack.

- Resilient-to-Targeted-Impersonation like the following two criteria was rated very relevant, calling targeted impersonation one of the greatest threats currently and
<table>
<thead>
<tr>
<th>Criterion</th>
<th>Avg. rating</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resilient-to-Phishing</td>
<td>4,00</td>
<td>0,000</td>
</tr>
<tr>
<td>Resilient-to-Targeted-Impersonation</td>
<td>3,86</td>
<td>0,378</td>
</tr>
<tr>
<td>Easy-to-Learn</td>
<td>3,86</td>
<td>0,378</td>
</tr>
<tr>
<td>Resilient-to-Leaks-from-other-Verifiers</td>
<td>3,86</td>
<td>0,378</td>
</tr>
<tr>
<td>Memorywise-Effortless</td>
<td>3,86</td>
<td>0,378</td>
</tr>
<tr>
<td>Negligible-Cost-per-User</td>
<td>3,71</td>
<td>0,756</td>
</tr>
<tr>
<td>Resilient-Theft</td>
<td>3,71</td>
<td>0,488</td>
</tr>
<tr>
<td>Resilient-to-Physical-Observation</td>
<td>3,57</td>
<td>0,535</td>
</tr>
<tr>
<td>Complete-Mediation</td>
<td>3,57</td>
<td>0,787</td>
</tr>
<tr>
<td>Resilient-to-Internal-Observation</td>
<td>3,57</td>
<td>0,787</td>
</tr>
<tr>
<td>Accessible</td>
<td>3,57</td>
<td>0,787</td>
</tr>
<tr>
<td>Scalable-for-User</td>
<td>3,57</td>
<td>0,787</td>
</tr>
<tr>
<td>Implemented</td>
<td>3,43</td>
<td>0,976</td>
</tr>
<tr>
<td>Work-Factor</td>
<td>3,43</td>
<td>0,787</td>
</tr>
<tr>
<td>No-trusted-Third-Party</td>
<td>3,43</td>
<td>1,134</td>
</tr>
<tr>
<td>Unlinkable</td>
<td>3,29</td>
<td>1,253</td>
</tr>
<tr>
<td>Physically-Effortless</td>
<td>3,14</td>
<td>1,464</td>
</tr>
<tr>
<td>Nothing-to-Carry</td>
<td>3,00</td>
<td>1,155</td>
</tr>
<tr>
<td>Requiring-Explicit-Consent</td>
<td>2,71</td>
<td>1,603</td>
</tr>
<tr>
<td>Browser- Compatible</td>
<td>2,57</td>
<td>1,272</td>
</tr>
<tr>
<td>Non-Proprietary</td>
<td>2,43</td>
<td>1,134</td>
</tr>
<tr>
<td>Match-System-to-Real-World</td>
<td>2,29</td>
<td>1,496</td>
</tr>
<tr>
<td>Open</td>
<td>2,00</td>
<td>1,527</td>
</tr>
</tbody>
</table>

Tab. 1: Average rating and std. dev. according to the experts

relating it to cyber-bullying. One expert said it is only relevant, justifying it by mentioning that not everything can be the most important.

- **Easy-to-Learn** was rated highly because multiple experts explained that a process that is easy to learn can reduce training costs and does not stress users with complexity. One expert rated the criterion as just relevant, but gave no explanation. Therefore it is likely just personal preference.

- **Resilient-to-Leaks-from-other-Verifiers** was rated just as highly with only one expert categorizing it as only relevant calling the criterion unattainable because attackers have access to highly sophisticated attacks and regularly leak data to the darknet. Others called it essential or repeated that it should not be possible to get user data from other service providers.

- **Non-Proprietary** is closely related to **Open** but considers the roll-out effort and costs more than security. With ratings ranging between 1 and 4 on the scale, there is no clear answer for the relevance of this criterion. For some, cost is of higher value than
security. Others say licensing costs are expected, and with free offers, the developer bears no liability.

- *Match-System-to-Real-World* even received ratings between 0 and 4, covering the whole scale. Generally the ratings were more on lower side, though two experts stated it would be very relevant. One said, an unknown process could unsettle the users and therefore rated it highly while the lowest ratings came with the explanation that it would not be necessary, and that the real world would change so often that processes would constantly need to be adjusted.

- *Open* received the lowest average rating of all criteria, with some experts giving similar reasoning to *Non-Proprietary*. High ratings were not accompanied by explanations while experts who gave lower ratings reasoned, that a proprietary, but certified system would be chosen just as likely. One even mentioned that open-source software would carry a negative impact since their cyber-insurance would not cover damages caused by open-source software.

- *Implemented* was viewed as relevant by most of the experts, with especially experts in large corporations reasoning, that software development is not their business and that self-developed software could also cause problems regarding the cyber-insurance. Rather surprisingly, experts from smaller companies expressed that hiring someone to develop this process would be acceptable as well.

For example, one expert from a large company gave the following statement in regards to the criterion *Implemented*:

"[...] we are not a software company […]. I think there are other people who are better equipped for this and [we] are not […] hiring someone here who only takes care for this topic." (Translated from German)

5 Discussion

The quantitative analysis of the expert opinions suggests that there are some mandatory requirements that an account recovery procedure needs to fulfill, while others could be considered optional. This is corroborated by the interviews. For this paper, we defined the requirements with a standard deviation of > 1 as optional, and with an average rating of < 2.5 as not relevant, which is reflected in Table 1.

The inductive analysis also allows to infer new criteria, which are defined below:

1. **Reliable (Usability):** The user can expect successful execution of the recovery procedure if followed properly. A procedure should receive a good rating if it is already widely used. (Mentioned by 1 expert)

2. **Ease-of-Use (Usability):** The procedure is designed not to complicate the user-experience and does not require additional devices. (Mentioned by 1 expert)
3. **Trusted-Vendor (Security):** The manufacturer of a possession factor like a hardware token can be trusted. Its adherence to security standards has been proven. (Mentioned by 1 expert)

4. **Cost-to-Benefit (Deployability):** The procedure offers more benefits for the same price or the same benefits at a lower price than a chosen reference procedure. (Mentioned by 3 experts)

5. **Resilient-to-Failure (Deployability):** The mechanism offers built-in redundancy as a measure against system failure or can be rolled out redundantly. (Indirectly mentioned by 1 expert)

6. **Regulatory-Compliant (Deployability):** The process meets the company’s regulatory requirements. This includes, for example, security standards such as NIS2 [EU22], IT-GS (in Germany) [BS17] or the ISO27000 series [IS18a]. (Indirectly mentioned by 1 expert)

The criteria “Resilient-to-Failure” and “Regulatory-Compliant” were not defined explicitly by the interviewed experts. The above definitions for these two criteria were therefore generalized from the conversations. Newly added criteria are categorized as optional, since other experts could not confirm their relevance. This was due to the difficulty of acquiring experts willing to donate time for research, and due to the small number of new criteria.

### 6 Conclusion

The paper grounds evaluation criteria for account recovery mechanisms for enterprise use in requirements coming from practitioners in industry. It was possible to uncover several new requirements, as well as evaluate the importance of the requirements given in existing literature. The first part of the interview gave a snapshot of state of deployment of passwordless authentication and account recovery procedures at the time the interviews took place.

As might be expected, some requirements in literature are not considered to be as important as others. Interestingly, the criteria “open” seems to have a negative impact on insurability, or is at least perceived as such. This shows that the criteria in the category “deployability” may need to be extended further to include even more business related aspects as experience and interdependence with other aspects increase.

Future work in this area may include surveying how the relevance of criteria changes as companies collect more experience with passwordless authentication and the associated processes. A re-run of the interviews with a larger sample size and including the newly discovered criteria may result in a more conclusive ranking of the criteria. Additionally, collecting non-experts’ views, i.e. users’ views, may provide a different perspective.
References


Appendix - Definitions from literature

Passwordless authentication is a primary authentication method that does not rely on knowledge. Security tokens, smartcards or biometrics are examples [PMA22]. Multi-factor authentication that use knowledge (e.g. password) as one factor are not considered passwordless.

Account Recovery is a process to recover access through an alternative authentication method (also called secondary or fallback authentication) when the primary user authentication method can’t be used, e.g. due to lost security token or forgotten password, [Ge23; Ku21; LWS18]. Since Account Recovery is also a method of authentication on a system, it should be noted that the procedure needs to provide at least equivalent security compared to the primary authentication method.

Usability criteria represent improvements in efficiency and user satisfaction. [IS18b]. It considers the simplicity of the mechanism to provide users with a smooth, positive experience without additional support.

Deployability criteria consider the use of corporate resources for the deployment and maintenance of the resources required for the mechanism. The term “deployability” was coined by Bonneau et al. [Bo12]. The aim is to keep the necessary resources as low as possible so that companies of all sizes are able to use a considered process.

Security and its criteria address the method in terms of its resilience to various attacks as well as potential security risks that may arise from third parties or a closed system. In addition, this category also lists criteria dealing with privacy and data protection [Bo12].

Kunke et al. [Ku21] defined the following criteria, which are taken as the basis for this work. The suffixes indicate which category the criteria fall under: U for Usability, D for Deployability, and S for Security.

(1U) Memorywise-Effortless: The user does not have to remember an authentication secret.
(2U) Scalable-for-User: No additional burden is introduced when using the mechanism with hundreds of services.
(3U) Nothing-to-Carry: The user does not need to carry any additional physical item to use the recovery mechanism at any time.
(4U) Physically-Effortless: Users do not need to perform any physical activities during the process beyond pressing a button.
Easy-to-Learn: The mechanism is intuitively designed and thus easy to learn.

Match-System-to-Real-World: The access recovery mechanism is based on real world concepts. The user can operate it intuitively because it is based on real world operations.

Accessible: Users must be able to use this mechanism, even with physical limitations.

Negligible-Cost-per-User: The financial cost per user must be very low.

Browser-Compatible: Mechanism can be used with any standard web browser without installing additional plugins or other software.

Non-Proprietary: The mechanism can be used at no additional cost for royalties and are not protected by patents or other trade secrets.

Implemented: The mechanism must be implemented as a practical application. It must not exist only as a theoretical concept.

Resilient-to-Physical-Observation: Despite observing the user while using the mechanism, attackers fail to successfully legitimize themselves as the user.

Resilient-to-Targeted-Impersonation: The attacker can not impersonate the user to the mechanism with background knowledge, which he may be able to obtain, e.g., via social networks.

Resilient-to-Internal-Observation: Despite intercepting user input at participating devices, e.g., smartphone or desktop PC, it is impossible for an attacker to imitate the user.

Resilient-to-Leaks-from-Other-Verifiers: A user uses other services that use the same or similar mechanism but whose data is made public. The attacker cannot impersonate the user with the obtained data at that service.

Resilient-to-Phishing: Attacker is able to fake a legitimate mechanism and convince the user to use the faked version, but cannot successfully impersonate the user to the service with the resulting data.

Resilient-to-Theft: Refers to mechanisms that require the factor possession in the form of an object as proof of legitimacy. If attackers gain possession of a user’s object, they must not succeed in legitimizing themselves as the user to the mechanism.

No-trusted-Third-Party: The mechanism for checking the authorization of the access recovery process is not based on a third trusted party, which could have been taken over or manipulated by an attacker to become an untrusted party.

Requiring-Explicit-Consent: The access recovery mechanism must only be performed with the user’s conscious consent. It must never be started accidentally or automatically.

Unlinkable: The information processed by this mechanism cannot be used to draw conclusions about what other services a user is using.

Open: The code or at least the functionality of the mechanism must be openly accessible to everyone.

Work-Factor: The mechanism should be designed in such a way that an attacker has to invest many resources to falsely successfully legitimize against the mechanism.

Complete-Mediation: The authorization to use the mechanism must be verified every time. It is not enough to assume that the person that operates the mechanism during an open session is the legitimate user.