Keep on Rating - On the Systematic Rating and Comparison of Authentication Schemes

Verena Zimmermann[1], Nina Gerber[2], Peter Mayer[2], Marius Kleboth[1], Alexandra von Preuschen[1], Konstantin Schmidt[1]

[1]FAI - Work and Engineering Psychology - Technische Universität Darmstadt
[2]SECUSO - Security, Usability, and Society, Karlsruhe Institute of Technology
e-mail: zimmermann@psychologie.tu-darmstadt.de

Abstract

Purpose: Six years ago, Bonneau et al. (2012) proposed a framework to compare authentication schemes to the ubiquitous text password. Even though their did not reveal an alternative outperforming the text password on every criterion, the framework can support decision makers in finding suitable solutions for specific authentication contexts. The aim of this paper is to extend and update the data base thereby discussing benefits, limitations and suggestions for continuing the development of the framework.
Approach: This paper revisits the rating process and describes the application of an extended version of the original framework to an additional 40 authentication schemes identified in a literature review. All schemes were rated in terms of 25 objective features assigned to the three main criteria usability, deployability, and security.
Findings: The rating process and results are presented along with a discussion of the benefits and pitfalls of the rating process.
Research implications and limitations: While the extended framework in general proves suitable for rating and comparing authentication schemes, ambiguities in the rating could be solved by providing clearer definitions and cut-off values. Further, the extension of the framework with subjective user perceptions that sometimes differ from objective ratings could be beneficial.
Value: The results of the rating are made publicly available in an authentication choice support system named ACCESS to support decision makers and researchers, and to foster the further extension of the knowledge base and future development of the extended rating framework.

Keywords

Authentication Scheme, Password, Rating, ACCESS, Objective Features, Subjective Perceptions

1. Introduction

Authentication has long become an integral part of daily life. Every single authentication process provides access to private data like emails, account data, personal documents, or photos. A loss thereof to an unauthorized third party can thus have a huge impact on private life or businesses.

The password as an authentication scheme still is ubiquitous but suffers from shortcomings such as the high cognitive load of memorizing different passwords for multiple accounts. As a coping strategy, users often choose the same password across accounts, keep an insecurely stored record or choose unsecure dictionary passwords (e.g., Adams *et al.* 1997, Johnson and Grawemeyer, 2011, Wash *et al.*, 2016).

To mitigate the issues associated with text passwords, many alternative schemes have been developed including biometric or token-based schemes. Bonneau *et al.* (2012) compared these to the text password across a variety of features and, surprisingly, found that replacing the password was not as easy as imagined. None of the analyzed schemes received high scores in all of the three evaluated categories usability, deployability, and security. Still, the comparison has proven to be very helpful in identifying authentication schemes best-suited for a certain purpose or certain requirements in research and practice alike. Thus, the initial work by Bonneau *et al.* (2012) serves as a basis for the evaluation of further authentication schemes. To realize an even more objective evaluation with an increased differentiation between authentication schemes additional sub features have been formulated by Mayer *et al.* (2016). The sub features were formulated as partially exclusive axioms to clearly allocate a scheme to a certain class of features.

However, while the results by Bonneau *et al.* (2012) and Mayer *et al.* (2016) demonstrate the suitability of the rating for researchers and practitioners, the coverage of authentication schemes by their work is still very limited. Mayer *et al.* (2016) applied their finer-grained ratings only to the original data set from Bonneau *et al.* (2012) and an additional ten schemes. Compared to these 45 schemes, a far greater number of schemes have been proposed in the literature and decision-makers in research and practice would greatly benefit from an update and extension of the data set to choose suitable authentication schemes from. In order to advance the diversity of authentication scheme in the rated pool, this paper describes the process and results of a rating of 40 additional authentication schemes identified in the literature. Further, the benefits, limitations, and suggestions for further improving and extending the framework are discussed.

The core contributions of this work are three-fold:

1. The pool of authentication schemes rated using the same methodology is significantly extended from 45 to 85. Thereby, not only the number, but also the diversity in the pool of available schemes is increased. This extension offers decision makers a greater selection when choosing appropriate authentication schemes for their specific application scenarios.

2. The ratings are integrated into the free, online authentication choice support system ACCESS (Renaud et al., 2014; SECUSO, 2016) so that practitioners and researchers can easily benefit from our results.

3. The advantages and pitfalls of the rating process and the choice support system ACCESS are discussed to support others in the future rating of authentication schemes and to provide a starting point for solving ambiguous results within the community.

The remainder of the paper is structured as follows: Section 2 describes the methodology of the rating process. Section 3 exemplarily presents the rating process and results of four different authentication schemes. Due to space constraints, the complete rating results are made available online and within ACCESS (c.f. contribution 2). In section 4, use cases for the rating process are presented. Section 5 discusses limitations of the rating process and presents the rating system ACCESS. Finally, section 6 concludes the paper.

This paper is a revised and extended version of a paper presented at the 12th International Symposium on Human Aspects of Information Security & Assurance 2018 (Zimmermann *et al.* 2018).

2. Method

One of the primary goals of this research was to supplement and update the original rating of authentication schemes by Bonneau *et al.* (2012). To that end, a literature search via Google Scholar was conducted which revealed a total of 164 relevant publications dealing with authentication schemes. All publications addressed or evaluated the user interaction with or perception of the authentication schemes. Papers only describing technical aspects or algorithms were not considered. From the analysis 40 authentication schemes which were not already included in the rating by Bonneau *et al.* (2012) could be extracted. Even though all schemes were extracted from research papers, a significant number of these schemes are actually used in practice, e.g., Challenge Questions, Face Recognition, Passphrases, and Google's Android Pattern Unlock.

The second step was to rate the schemes according to the 63 sub features refined by Mayer *et al.* (2016) and shown in Appendix A. These were derived from the original 25 features of authentication schemes as defined by Bonneau *et al.* (2012). The sub features are extensions of the original features and provide a more detailed way to evaluate authentication schemes. For example, the feature "Accessible" is split into the three sub features "Accessible with Read/Write-Impairments", "Accessible with Visual Impairments" and "Accessible with Physical Impairments". They are also partially exclusive in that a scheme can only fulfil one of the sub features but not two at the same time. This allows for the allocation of schemes to distinctive classes. An example for this is the feature "Proprietary" with the sub features "Proprietary" or "Non-Proprietary".

The rating process was structured as follows: Similar to Bonneau *et al.* (2012) three of the authors each rated a subset of the 40 identified authentication schemes in terms

of every sub feature. Any arising questions or problems were discussed within the research group including an additional three independent researchers. Whenever possible, the rating was based on the description of the scheme or other data provided by the authors in the original publication. Where the original publication was not available or sufficient, e.g. where the scheme was only described in a review paper, additional literature describing the scheme was considered. In case a publication did not provide any specifics regarding a criterion, e.g., because the scheme was presented only on a conceptual level, the rating was logically derived from the description of the scheme. For example, even though some descriptions of biometric schemes did not actually state the number of secrets to remember to rate the feature "Memorywise-Effortless" the information was logically derived from the conceptual approach which is based on detecting biometric features that users carry with them naturally and do not have to remember. All ratings were conducted for using the authentication scheme with a PC or laptop.

In general, the ratings of authentication schemes widely used in various forms and without an identifiable "original" publication such as the fingerprint scheme or different password schemes were based on the *concept* of the scheme, rather than the specifics of a *certain implementation*. A scheme should not be excluded by a decision-maker beforehand due to a low rating based on a single implementation if someone deciding to use such a scheme could easily adapt certain aspects of an implementation according to the context of use. An example is setting a limit to the number of login attempts allowed before temporarily blocking an account, which affects the rating of the feature "Resilient-to-Throttled-Guessing". To preserve internal consistency, all new schemes were also compared to the ones that had already been rated by Bonneau *et al.* (2012) and Mayer *et al.* (2016) thus giving similar authentication schemes identical ratings. Examples include the already rated "Iris Scan" that shares features with the newly added "Retina Scan".

3. Results

Due to space constraints, the rating results will be presented exemplarily for four authentication schemes: the biometric scheme Retina Scan, the graphical scheme Android Pattern Unlock (Google Inc, 2011), the text-based scheme Associative Questions (Irakleous *et al.*, 2002) and the token-based scheme Cronto (OneSpan Inc., 2019). While the former three add to the existing data base, the scheme Cronto has already been rated by Bonneau *et al.* (2012) but is presented here for the purpose of comparison with a token-based authentication scheme. The rating results of the four exemplary schemes are depicted in Appendix A. The complete rating results of all newly added schemes as well as a description of the schemes and the rating features can be accessed online and via ACCESS (see Appendix B).

Retina Scan is a biometric authentication scheme that identifies the user by his/her unique patterns on the retina blood vessels (Figure 1a). The patterns are detected optically by casting an unperceived beam of low-energy infrared light into the user's

eye and measuring the absorption levels of light. In general, an appropriate sensor is required to perform the authentication. The Retina Scan is different from the Iris Scan where near infrared images of the iris are used for authentication. Similar to the Finger Print the Retina Scan is a general concept with a variety of implementations.

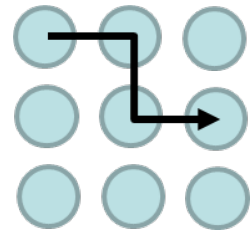
Android Pattern Unlock (Google Inc, 2011) is a proprietary recall-based graphical authentication scheme mainly used on mobile phones. To authenticate, the user draws a memorized path visiting up to nine dots on a 3x3 grid. Each dot can only be visited once (Figure 1b).

Associative Questions is a knowledge-based authentication scheme. Users name and memorize one association for each of the 20 given keywords (e.g. blue, house and fire). For authentication, they must recall their associated word for five randomly chosen keywords out of the set of keywords as shown in Figure 1c (Irakleous *et al*., 2002).

Cronto (OneSpan Inc., 2019) is a proprietary token-based scheme, also known as Photo-TAN, which is often used for online banking. It requires the user to operate a smartphone application with a stored secret key to scan a visual pattern similar to a QR-code (Figure 1d). This visual pattern holds encrypted transaction details as well as a one-time code. For the authentication, the system checks the validity of the one-time code typed in by the user.

a)

b)



c)

word	Association
blue	water
peace	pigeon
love	heart
eat	pizza
distant	

d)

Figure 1: Depictions of a) Retinal blood vessels (Pixabay.com, 2019), b) Android Pattern Unlock Scheme, c) Associative Questions, and d) visual pattern similar to that used for Cronto (Wikimedia.org, 2016).

3.1. Usability

Memorywise-Effortless. This feature is split into the three exclusive sub features: “No-Secret-to-Remember”, “One-Secret-to-Remember” and “More-than-One-Secret-to-Remember”. As Retina Scans solely rely on measurable characteristics of the user, they are assigned the feature “No-Secret-to-Remember”. This is not the case for Android Pattern Unlock, which requires the user to create a new, individual secret for each verifier and consequently was rated "More-than-one-Secret-to-Remember". The secret in the scheme Associative Questions even consists of a set of words and different sets for multiple verifiers and is thus classified "More-than-one-Secret-to-Remember". The scheme Cronto is more difficult to rate. It does not require the user to remember a secret and is thus rated as “No-Secret-to-Remember”. Still, a new key for extracting the secret from the visual pattern can be generated for every service and stored in the application.

Scalable-for-Users. In line with the “Memory-wise-Effortless” rating only the schemes Retina Scan and Cronto are granted “Scalable-for-Users” as they do not increase the cognitive load for authentication with different verifiers. The number of secrets to remember and thus the cognitive load, however, increases with the number of accounts when using Android Pattern Unlock or Associative Questions.

Nothing-to-Carry. The only scheme of the four that requires the user to carry an additional object, a smartphone, is the scheme Cronto (“Phone-to-Carry”).

Physically-Effortless. The scheme Retina Scan does not require particular physical effort and is thus rated “No-Physical-Effort”. The associated words used by Associative Questions and the one-time code used in Cronto need to be typed in and are thus classified as “Type-to-Enter”. The Android Pattern Unlock scheme requires the drawing of a pattern (“Scribble-to-Enter”).

Easy-to-Learn. All of the schemes are “Easy-to-Learn” as they consist of few process steps that are easy to recall.

Efficient-to-Use. All four schemes are granted the feature “No-Obstructive-Latency” as all process steps are directly related to the secret and the user does not need to wait for information from the verifier that is, e.g., sent via SMS. Only the scheme Cronto requires some “fiddling” as the smartphone application needs to be started and the smartphone aligned with the visual Cronto pattern to extract the secret. Similarly, only Cronto requires the transcription of a secret, the extraction of and typing in of the one-time code encrypted in the visual pattern. It is thus not granted “No-Fiddling-Tasks” and “No-Secret-to-Transcribe”.

Infrequent-Errors. The schemes Associative Questions, Android Pattern Unlock and Retina Scan are classified as “Not-Susceptible-to Assignment-Errors” and “Not-Susceptible-to-Transmission-Errors”. Only Cronto involves an additional smartphone

application that must be handled. Even though the error rate is low, problems might result from that issue.

In contrast, none of the schemes is granted “Not-Susceptible-To-Input-Errors” as Cronto, Associative Questions and Android Pattern Unlock might be prone to typing or drawing errors. Retina Scan might be prone to problems with the sensor or changes of the retina, e.g. due to a medical condition, leading to input errors.

Easy-Recovery-from-Loss. The schemes Android Pattern Unlock and Associative Questions were rated as offering “Easy-Recovery-from Loss”. Forgotten or stolen secrets can easily be replaced by new ones without having to overcome unreasonable burdens, e.g. by sending a recovery link via email. In contrast, Retina Scan was rated as “No-Easy-Recovery-from-Loss” as a compromised account or a physical inability to further use the scheme results in having to replace the scheme with an alternative one. Similarly, Cronto was rated “No-Easy-Recovery-from-Loss” as replacing the smartphone, rendering existing keys invalid, and generating new keys pose a significant burden.

3.2. Deployability

Accessible. Retina Scan is awarded the two sub features “Accessible-with-Read/Write-Impairments” and “Accessible-with-Physical-Impairments” as the only requirement is that the user can correctly position his/her retina in front of the scanner. However, visual impairments (e.g. due to a medical condition) can prevent people from being able to use Retina Scan.

Visual or physical impairments make it impossible to use Cronto, where users have to transfer a one-time code, type, and use a physical device. If we assume the one-time code to only consist of numbers that have to be found and selected on a keyboard, the scheme could be viewed as “Accessible-with-Read/Write-Impairments”.

Android Pattern Unlock does not involve text and, thus, is accessible with read/write impairments. Associative Questions could be used with a speech-to-text engine and can thus be used with visual or physical impairments. However, as the scheme requires the reading of random keywords and the typing of associations it is not granted “Accessible-with-Read/Write-Impairments”.

Negligible-Cost-per-User. Negligible-Cost-per-User is awarded to all schemes but Retina Scan. Retina Scan requires a scanner at each login point, which is possibly at every user’s computer. All other schemes only rely on existing hardware and the provided software and therefore do not involve extra cost. Android Pattern Unlock and Cronto are proprietary, patented schemes. It is thus expected that a license fee has to be paid. It was assumed that this fee does not increase with every new user or device as in this case they would lose the Negligible-Cost-per-User attribute.

Server-Compatible. Server compatibility is given for Android Pattern Unlock as the sequence of connected dots in the Android Pattern Unlock can also be represented as a password. Retina Scan requires a more sophisticated way of comparing the scanned image with the stored image and is thus not compatible to passwords. Similarly, the random selection of words in the scheme Associative Questions requires additional server-side modifications. Cronto relies on an additional device, a visual code, and a one-time code, being much more complicated than just a password.

Browser-compatible. Native browser compatibility is given for Cronto, Associative Questions, and Android Pattern Unlock. All three can be easily implemented in a web browser using common technologies such as HTML, CSS, and JavaScript. Retina Scan on the other hand is not browser compatible as additional software is needed for using the scanner and reading the image, for example.

Mature. Mature schemes that are repeatedly adopted, inside and outside academics are Android Pattern Unlock, Cronto, and Retina Scan. All three are popular in certain domains, e.g. Android Pattern Unlock for mobile phones, Cronto for online banking, and Retina Scan for security purposes. Associative Questions in this form is only presented and implemented in few academic papers (Haga and Zviran, 1991; Irakleous *et al*., 2002).

Non-Proprietary. Cronto and Android Pattern Unlock are proprietary schemes as they have been developed and patented by companies. Associative Questions and Retina Scan are freely-available concepts that do not necessarily involve proprietary software.

3.3. Security

Resilient-to-Physical-Observation. As a device-based challenge-response scheme Cronto offers all the associated sub features simply because there is no secret static element that could be observed. Retina Scan offers all sub features because the possibilities to capture the retinal blood vessel structures were deemed too limited and there are no residual traces or sound to record. Associative Questions utilizes a portfolio approach that only uses a few of the previously registered associations per authentication, so that an attacker would need to record multiple login procedures to obtain the entire secret. Because of this argumentation it was granted all four sub features. Still, without a clear cut-off value the rating of this feature might be ambiguous.

Of the four discussed schemes only Android Pattern Unlock was deemed “Non-Resilient-to-Visual-Recording” and “Non-Resilient-to-Shoulder-Surfing” as the secret is wholly observable from a single authentication. Android Pattern Unlock is also susceptible to smudge attacks if used on a touchscreen leading to a “Non-Resilient-to-Residual-Traces-Recording” rating.

Resilient-to-Targeted-Impersonation. Cronto and Retina Scan were rated as "resilient-to-targeted-impersonation" since the possession of personal information gives an attacker no advantage. However, the rating of Associative Questions and Android Pattern Unlock is more difficult. The chosen associations and patterns do not need to be based on personal information, but it is likely that users choose secrets associated with themselves, e.g. the first letter of their name as pattern or pets' or friends' names as associations. Considering this possibility, Android Pattern Unlock and Associative Questions were rated "Non-Resilient-to-Targeted-Impersonation".

Resilient-to-Throttled-Guessing. Schemes with a password space larger than 10^6 (Florêncio *et al.*, 2014) were rated to be "Resilient-to-Throttled-Guessing". Still, the four examples discussed here show that it is often difficult to assess the size of the password space of a scheme. The password space of Android Pattern Unlock was calculated with 2^19 possible patterns by Uellenbeck *et al.* (2013) and was thus rated as "Non-Resilient-to-Throttled-Guessing".

Because of its high vulnerability against dictionary attacks Associative Questions was not awarded that feature either. The resilience of Retina Scan is harder to assess since it is highly dependent on the false acceptance rate (FAR) – false rejection rate (FRR) trade-off in the specific implementation. Additionally, its accuracy changes with technological progress. Nevertheless, Retina Scan was granted the feature due to similarities with the scheme Iris Scan that falls into the same category (Bonneau *et al.*, 2012). Because Cronto utilizes a stored key and randomly generated one-time codes it was also rated as "Resilient-to-Throttled-Guessing".

Resilient-to-Unthrottled-Guessing. For the unthrottled-guessing category a minimum password space of 10^14 (Florêncio *et al.*, 2014) was set as threshold, hence Android Pattern Unlock and Associative Questions were not granted this benefit. Additionally, because we were unable to find sufficient evidence and in line with the rating of Iris Scan the scheme Retina Scan was not granted this benefit either. Cronto, however, was awarded the benefit as the key stored in the smartphone is assumed to be resilient to unthrottled-guessing. Even a successful guess of the generated one-time code, as suggested by the name, would only grant access once, not to the account in general.

Resilient-to-Internal-Observation. Of the four schemes only Cronto offers "Resilience-to-Eavesdropping" since tapping the communication is not sufficient to pose as the user. The secret key of the user held by the associated smartphone is required as well. However, it does not offer protection against malware as smartphones are easily infected with malicious software. All other schemes are based on static secrets, which makes them vulnerable to both malware and eavesdropping.

Resilient-to-Leaks-from-other-Verifiers. As Android Pattern Unlock, Retina Scan and Associative Questions are based on static secrets, and users might tend to reuse patterns or associations across accounts, a successful attack on one verifier might

affect other accounts of the users. Protection against possible leaks from other verifiers again is only offered by Cronto through the usage of unique private keys for each verifier.

Resilient-to-Phishing. Again, only Cronto is rated "Resilient-to-Phishing" as the user's key is not part of the challenge-response protocol and never entered into an interface on a website.

Resilient-to-Theft. Associative Questions, Android Pattern Unlock and Retina Scan do not involve physical objects that could be stolen and were rated to be "Resilient-to-Theft". Only Cronto utilizes the user's phone that could be stolen. Unless an optional second factor as assumed by Bonneau *et al.* (2012) is used for authentication, the possession of the phone is sufficient to pose as the user since the key is stored inside the phone ("Non-Resilient-to-Theft").

No-Trusted-Third-Party. None of the four schemes involves a third party besides the user and the verifier.

Requiring-Explicit-Consent. All four schemes were rated to require explicit consent. This rating was also granted to Retina Scan because of the invasiveness of the authentication procedure.

Unlinkable. Only for Retina Scan it is possible to link two accounts solely through the verification details, as the blood vessel pattern in the retina is regarded as unique to each person.

4. Application

The following section presents examples for the application of the rating by researchers and practitioners.

4.1. Application of the Rating by Researchers

The results of the rating process can be useful for authentication research as they allow researchers to quickly identify appropriate authentication schemes for study purposes or software applications. It further allows a thorough comparison of newly developed authentication schemes with a variety of existing approaches. One practical example for the use of the rating is a project on user-friendly authentication and encryption within the Centre for Research in Security and Privacy (CRISP). Within the project certain limitations for the choice of authentication schemes exist, e.g., it should be cost-free for the user and deployable in web browsers. First, the rating process described here allowed for excluding authentication schemes that did not meet the criteria set in the project and rank others in terms of the remaining objective security, usability and deployability features. Second, the rating was used to identify the best performing schemes out of five different categories, such as

knowledge-based and biometric schemes. The resulting schemes were analysed in terms of user perceptions in a laboratory study to identify the most suitable one for this use case.

4.2. Application of the Rating by Practitioners

Practitioners may use the results of the rating for similar purposes as researchers, e.g. for study purposes or for comparing own with existing approaches. Apart from that, the rating may support practitioners in identifying an appropriate authentication scheme for their service, web application, or product. It provides an overview over a range of existing schemes and, similar to the research example described above, allows excluding schemes that do not meet the requirements given by the product or the target user group.

5. Discussion of the Rating Process and ACCESS

As described above, the rating process provides a number of benefits for researchers and practitioners alike: support in the choice of an existing authentication scheme for one's own application or study, a comparison of new schemes with existing ones, and requirement- as well as context-based ratings of authentication schemes.

Still, the rating process in its current form and the results described here suffer from several shortcomings that should be acknowledged and addressed in the future. The next sections will discuss limitations of the framework in general as well as ambiguities of certain features that were discovered in the rating process presented in this paper. Each challenge is accompanied by suggestions for improvement, e.g., in terms of increasing the clarity of the features or possible extensions of the rating criteria Further, the authentication choice support system ACCESS will be presented and discussed.

5.1. Framework

This section concerns the rating framework as proposed by Bonneau *et al.* (2012) and refined by Mayer *et al.* (2016) in general.

Concept vs. Implementation. The rating process was based on the literature available to us. Some schemes were only described in a few papers or on a conceptual level. In particular, some details and technical information necessary for the rating were not available, so that the rating had to be based on similar schemes and/or logically derived from the conceptual approach. On the one hand, the rating would thus benefit from being reviewed by the developers of the authentication schemes that are experts for their work. On the other hand, the ratings might be more favorable if they are not done by an independent party. This points towards a general problem with rating immature schemes.

Other schemes, however, were described in many papers and in many different forms or implementations. One example is keystroke dynamics, where various implementations and service providers exist. In cases where it was not possible to rate the scheme independently from a certain implementation, we searched for review papers or a "common" way of implementation. Still, future work might benefit from rating different implementations separately and including the reference to the developers of that implementation.

Objective Ratings vs. Subjective User Perceptions. Currently, the framework aims for a high level of objectivity in the rating of authentication schemes. Still, research shows that subjective user perceptions of authentication schemes can differ from objective or technical features (Bhagavatula *et al.* 2015; Ur *et al.*, 2015; Zimmermann and Gerber, 2017). Subjective perceptions, e.g., security perceptions, are important as they may influence the acceptance and hence the actual use of these schemes (Huang *et al*., 2011). This is not to say that objective features should be replaced by subjective perceptions, but that an extension with subjective perceptions could support decisions between equally suitable schemes in favor of the user perspective. Further, subjective perceptions could offer explanations for low acceptance of highly rated schemes or reveal potential for improving aspects of certain authentication schemes.

A challenge of including subjective perceptions would be that results of user studies are not available for all schemes and that different samples, methods, and constructs are used in different studies. It would therefore be difficult to, e.g., rate security perception as given or not given. As a preliminary solution results of user studies could be provided in a qualitative form as additional decision support. Future research could explore more suitable ways of organizing and comparing subjective perceptions of authentication schemes.

Advances and Changes. The rating presented here was conducted at one point in time and with certain search terms and thus does not claim to cover an exhaustive list of existing authentication schemes. Besides, it is possible that schemes have been developed and improved further or that schemes are not available any more. Also, schemes already included in the knowledge based may meanwhile have reached new maturity levels or the cost for using schemes may have decreased with increasing spread or technological advances. Thus, to provide a valuable and actual resource for researchers and practitioners it would be beneficial if the knowledge base was regularly checked and updated by members of the community. One way to do so is provided by ACCESS, the authentication choice system that is presented in more detail in section 5.3.

5.2. Features and Sub Features

This section focuses on the definitions and cut-off values of certain features and sub-features within the rating framework.

Memorywise-Effortless and Scalable-for-Users. There seems to be a dependency between the features “Memorywise-Effortless” and “Scalable-for-Users”, e.g. schemes that require a new knowledge-based secret for every verifier are not scalable in terms of cognitive load. Similarly, only schemes that require no or one knowledge-based secret overall can be scalable. Hence, it might be suitable to either combine these features or make the distinction more explicit, e.g. to not only rate scalability in terms of cognitive effort but perhaps also in terms of other resources.

Resilience-to-(Un-)Throttled-Guessing. Generally, the features ”Resilient-to-Throttled-Guessing” and “Resilient-to-Unthrottled-Guessing” are hard to rate and the assumed cut-off values somewhat arbitrary, because the password space largely depends on the specific implementation and the availability of real-use data. For example, implementations allowing 32 vs. 128 characters for an answer, e.g. to an associative question, provide a different password space and resilience rating. Another issue is that, in practice, most users’ secrets cover only a fraction of the theoretical password space which could be influenced by user preferences or cultural aspects. To get an idea of the practical password space, though, user studies must be conducted, which is rarely the case especially for all freshly proposed immature systems.

A possible measure could be a refinement of the definition: One feature could rate the theoretical password space of one pre-dominant implementation. Where available, the actually used password space could be rated separately.

Resilience-to-Targeted-Impersonation. The exemplary rating results in section 3.3 show that the rating of “Resilience-to-Targeted-Impersonation” is ambiguous if the secret does not need to be based on personal information but if the user can freely choose the secret or might tend to base the secret on personal information. In such cases, it would be helpful if the definition of the feature was more precise or if an additional sub feature would account for the individual influence that users may have on the secret.

Nothing-to-Carry. While authentication methods like Face or Fingerprint recognition are supported by many current devices, schemes like Hand Vein Triangulation or Retina Scan require additional devices not commonly available. In these cases, not the secret per se but an additional device needs to be carried or provided at each login point. While a “device-to-carry” category already exists this sub feature is currently intended to account for devices like USB-dongles that directly carry the secret used for authentication.

Solutions to cover the remaining cases would be to extend the definition of the sub feature “device-to-carry” or to add another sub feature, such as “scanner-to-carry”. Schemes in this category require special devices or scanners that cannot be expected in current personal computers or phones. If these schemes are used only for fixed

points of access that include a scanner (e.g. a door with a retina scanner), however, users would not be required to carry any additional device.

Negligible-Cost-per-User. In line with the discussion of the feature "Nothing-to-Carry", the criterion "Negligible-Cost-per-User" has to be given a closer look. If only used for fixed access points, the cost for schemes which would usually require expensive authentication hardware (e.g. high-resolution cameras for retina scans, infrared cameras for hand-vein-triangulation), could be highly reduced by stationary devices available to all users. In this case, the cost would not increase with each user. Consequently, a distinction could be made by listing the cost factor that emerges per user and per access point. For example, the cost for using Retina Scan for door locks on the one hand is high per access point, but low per user. Security tokens on the other hand involve high cost per access point (door) and per user (token).

5.3. ACCESS Rating System

The rating process described in this paper was purposefully based on the refined rating framework by Mayer *et al.* (2016) which has been implemented as an online authentication choice support tool called ACCESS (Renaud *et al.*, 2014; Mayer *et al.*, 2018).

ACCESS is a freely available[1] open source platform[2]. It enables authentication researchers and practitioners to choose suitable authentication schemes for their application scenarios as well as to discuss the properties of authentication schemes. For researchers and practitioners alike, the major benefit of ACCESS is that it presents the results of the rating process described in this paper in a comprehensive and easily manageable form. Its functionality is provided by three different modules.

The information module allows systematic access to the information on the authentication schemes included in its knowledge base. The available information includes (a) a short description of the scheme, (b) the features provided by the authentication scheme (Mayer *et al*., 2016), and (c) a timeline of the changes to the knowledge base with respect to each scheme.

The decision support module guides through the process of choosing the most suitable authentication scheme according to the specified requirements. The requirement specification comprises only two steps. First, the authentication scheme features have to be ranked from most to least important in a drag-and-drop fashion. Then, hard constraints based on the sub-features have to be specified (e.g. whether the scheme has to be browser-compatible). After these two steps, the authentication schemes are listed ranked by their suitability (see Mayer *et al.*, 2016 for details).

[1] https://access.secuso.org/
[2] https://github.com/SECUSO/ACCESSv2

The collaboration module allows updating and extending the knowledge base with additional authentication schemes as well as changing the feature ratings of the schemes in the knowledge base.

Extending the original knowledge base to include the 40 schemes rated in the course of this work also allowed for refining the collaboration module, i.e. the process of adding and modifying the knowledge base. Additionally, areas for future work on the platform arose, e.g. handling potential changes to the set of features (e.g. through the inclusion of the subjective user perceptions as discussed in section 5.1) would require extending the platform.

Conclusion

This paper describes the rating process of 40 authentication schemes in terms of the three categories usability, security and deployability based on the framework introduced by Bonneau *et al.* (2012) and refined by Mayer *et al.* (2016). The rating offers researchers as well as practitioners an aid in the choice of appropriate authentication schemes for their specific application scenarios and allows comparisons with newly developed schemes.

To make the results easily available for the community, the rating results have been included in ACCESS (SECUSO, 2016), an authentication choice support system that allows the requirement-based rating of the authentication schemes. Thereby the number of included authentication schemes increased from 45 to 85. With the provision of our results within ACCESS we hope to allow a large number of researchers and practitioners to benefit from our work. Further, we hope to encourage other members of the community to add further schemes to the platform and participate in discussing and solving potential ambiguities in the rating process via the collaboration module.

The advantages and pitfalls of the rating process were discussed to support others in the rating of authentication schemes and to provide a starting point for future research. Suggestions for further improving the rating framework include the provision of more precise feature definitions and an extension with subjective perceptions.

6. Acknowledgement

The research reported in this paper was supported by the German Federal Ministry of Education and Research (BMBF) and by the Hessian Ministry of Science and the Arts within CRISP (www.crisp-da.de/). This work was further supported by the German Federal Ministry of Education and Research in the Competence Center for Applied Security Technology (KASTEL).

7. References

Adams, A., Sasse, M. A., and Lunt, P. (1997), “Making passwords secure and usable.”, in *Proceedings of HCI on People and Computers XII*, Springer, London, UK, pp. 1-19.

Bhagavatula, C., Ur, B., Iacovino, K., Kywe, S. M., Cranor, L. F., and Savvides, M. (2015). “Biometric authentication on iPhone and Android: Usability, Perceptions, and Influences on Adoption”, in *Proceedings USEC 2015*, Internet Society, pp. 1-2.

Bonneau, J., Herley, C., van Oorschot, P.C. and Stajano, F. (2012), “The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes”, in *Proceedings of the IEEE Symposium on Security and Privacy 2012*, IEEE, 2012, pp. 553-567.

Florêncio, D., Herley, C., and Van Oorschot, P. C. (2014), “An Administrator's Guide to Internet Password Research”, in *LISA*, Vol. 14, pp. 35-52.

Google Inc. (2011). *US Patent No. 20110283241,* Touch Gesture Actions From A Device's Lock Screen. Washington, D.C.: US Patent & Trademark Office.

Grawemeyer, B. and Johnson, H. (2011), “Using and managing multiple passwords: A week to a view. “, *Interacting with Computers,* Vol. 23 No. 3, pp. 256–267, doi: 10.1016/j.intcom.2011.03.007.

Haga, W. J., and Zviran, M. (1991). “Question-and-answer passwords: an empirical evaluation”, *Information systems*, Vol. 16, No. 3, pp. 335-343.

Huang, D.-l., Rau, P.-L.P., Salvendy, G., Gao, F. and Zhou, J. (2011), “Factors affecting perception of information security and their impacts on IT adoption and security practices”, *International Journal of Human-Computer Studies*, Vol. 69, No. 12, pp. 870-883.

Irakleous, I., Furnell, S. M., Dowland, P. S., and Papadaki, M. (2002), “An experimental comparison of secret-based user authentication technologies”, *Information Management & Computer Security*, Vol. 10 No. 3, pp. 100-108.

Mayer, P., Neumann, S., Storck, D. and Volkamer, M. (2016), “Supporting Decision Makers in Choosing Suitable Authentication Schemes”, in *Proceedings of the Tenth International Symposium on Human Aspects of Information Security & Assurance (HAISA 2016)*, pp. 67-77.

Mayer, P., Stumpf, P., Weber, T., Volkamer, M. (2018), “ACCESSv2: A Collaborative Authentication Research and Decision Support Platform”, *Who Are You?! Adventures in Authentication Workshop 2018*.

OneSpan Inc. (2019). “CRONTO”, available at: https://www.vasco.com/products/two-factor-authenticators/cronto/transaction-signing.html (accessed 17 January 2019).

Pixabay.com (2019), “Retina”, available at: https://pixabay.com/de/users/rdowns-2091986/ (accessed 07 January 2019).

Renaud, K., Volkamer, M. and Maguire, J. (2014), “ACCESS: Describing and Contrasting”, in *International Conference on Human Aspects of Information Security, Privacy, and Trust*, Springer, Cham, pp. 183-194.

Uellenbeck, S., Dürmuth, M., Wolf, C., and Holz, T. (2013), “Quantifying the security of graphical passwords: the case of android unlock patterns”, in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, ACM, pp. 161-172.

Ur, B., Noma, F., Bees, J., Segreti, S. M., Shay, R., Bauer, L., ... and Cranor, L. F. (2015, July). “I added ‘!’at the end to make it secure”: Observing password creation in the lab”, in *Proceedings SOUPS 2015,* Usenix, pp. 123-140.

Wash, R., Rader, E., Berman, R. and Wellmer, Z. (2016), “Understanding password choices: How frequently entered passwords are re-used across websites”, in *Proceedings SOUPS 2016*, Usenix, pp. 175-188.

Wikimedia.org (2016), “PhotoTAN mit Orientierungsmarkierungen.svg”, available at: http://commons.wikimedia.org/wiki/File:PhotoTAN.svg (accessed 07 January 2019).

Zimmermann, V., Gerber, N., Kleboth, M, von Preuschen, A., Schmidt, K. and Mayer, P. (2018), “The Quest to Replace Passwords Revisited – Rating Authentication Schemes”. In *Proceedings of the 12th International Symposium on Human Aspects of Information Security & Assurance (HAISA 2028)*.

Zimmermann, V., and Gerber, N. (2017). “If It Wasn’t Secure, They Would Not Use It in the Movies–Security Perceptions and User Acceptance of Authentication Technologies”, in *International Conference on Human Aspects of Information Security, Privacy, and Trust* Springer, Cham, pp. 265-283.

Appendix A: Rating Features and Exemplary Rating Results

[Insert Table 1 here]

Appendix B: Online-Appendix

The complete results and a description of the rated authentication schemes and rating features can be accessed with the following link: http://www.arbing.psychologie.tu-darmstadt.de/home/forschung_4/forschungsergebnisse_fai.de.jsp

The rating results are further integrated in ACCESS: https://access.secuso.org/

Table 1: Rating results of the Authentication Schemes Associative Questions, Retina Scan, Android Pattern Unlock and Cronto.

Category	Feature (Bonneau *et al*, 2012)	Sub feature (Mayer *et al*, 2016)	Associative Questions 1		Retina Scan		Android Pattern Unlock (Google)		Cronto	
			Bonneau	Mayer	Bonneau	Mayer	Bonneau	Mayer	Bonneau	Mayer
Usability	Memorywise-Effortless	No-Secret-to-Remember				x				(x)
		One-Secret-to-Remember								
		More-than-One-Secret-to-Remember	Not	x	Full		Not	x	Not	(x)
	Scalable-for-Users	Scalable-for-Users				x				x
		Non-Scalable-for-Users	Not	x	Full		Not	x	Not	
	Nothing-to-Carry	No-Object-to-Carry		x		x		x		
		Phone-to-Carry								x
		SmartCard-to-Carry								
		Document-to-Carry								
		Device-to-Carry	Full		Full		Full		Almost	
	Physically-Effortless	No-Physical-Effort				x				
		Speak-to-Enter								
		Type-to-Enter		x						x
		Scribble-to-Enter						x		
		Gesticulate-to-Enter	Not		Almost		Not		Not	
	Easy-to-Learn	Easy-to-Learn		x		x		x		x
		Non-Easy-to-Learn	Full		Full		Full		Full	
	Efficient-to-Use	No-Obstructive-Latency		x		x		x		x
		No-Fiddling-Tasks		x		x		x		
		No-Secret-to-Transcribe	Full	x	Almost	x	Full	x	Almost	
	Infrequent-Errors	Not-Susceptible-to-Input-Errors								
		Not-Susceptible-to-Assignment-Errors		x		x		x		?
		Not-Susceptible-to-Transmission-Errors	Full	x	Not	x	Almost	x	Almost	x
	Easy-Recovery-from-Loss	Easy-Recovery-from-Loss		x				x		
		No-Easy-Recovery-from-Loss	Full		Not	x	Full		Not	x
Deployability	Accessible	Accessible-with-Read/Write-Impairments				x		x		x
		Accessible-with-Visual-Impairments		x						
		Accessible-with-Physical-Impairments	Full	x	Almost	x	Not		Not	
	Negligible-Cost-per-User	Negligible-Cost-per-User		x				?		?
		Non-Negligible-Cost-per-User	Full		Not	x	Full		Almost	
	Server-Compatible	Server-Compatible						x		
		Non-Server-Compatible	Not	x	Not	x	Full		Not	x
	Browser-Compatible	Compatible-to-Native-Browser		x				x		x
		Compatible-to-Extended-Browser								
		Non-Browser-Compatible	Full		Not	x	Full		Full	
	Mature	Adopted-beyond-Academics				x		x		x
		Adopted-Repeatedly				x		x		x
		Adopted-in-Academics	Not	x	Full	x	Full	x	Full	x
	Non-Proprietary	Non-Proprietary		x		x				
		Proprietary	Full		Full		Not	x	Not	x
Security	Resilient-to-Physical-Observation	Resilient-to-Visual-Recording		x		x				x
		Resilient-to-Shoulder-Surfing		x		x				x
		Resilient-to-Residual-Traces-Recording		x		x				x
		Resilient-to-Sound-Recording	Not	x	Full	x	Not	x	Full	x
	Resilient-to-Targeted-Impersonation	Resilient-to-Targeted-Impersonation				x				x
		Non-Resilient-to-Targeted-Impersonation	Not	?	Full		Full	?	Full	
	Resilient-to-Throttled-Guessing	Resilient-to-Throttled-Guessing				x				x
		Non-Resilient-to-Throttled-Guessing	Not	x	Full		Not	x	Full	
	Unthrottled-Guessing	Resilient-to-Unthrottled-Guessing								x
		Non-Resilient-to-Unthrottled-Guessing	Not	x	Not	x	Not	x	Full	
	Resilient-to-Internal-Observation	Resilient-to-Eavesdropping								x
		Resilient-to-Malware	Not		Not		Not		Almost	
	Resilient-to-Leaks-from-Other-Verifiers	Resilient-to-Leaks-from-Other-Verifiers								x
		Non-Resilient-to-Leaks-from-Other-Verifiers	Not	x	Not	x	Not	x	Full	
	Resilient-to-Phishing	Resilient-to-Phishing								x
		Non-Resilient-to-Phishing	Not	x	Not	x	Not	x	Full	
	Resilient-to-Theft	Resilient-to-Theft		x		x		x		
		Non-Resilient-to-Theft	Full		Full		Full		Full	x
	No-Trusted-Third-Party	No-Trusted-Third-Party		x		x		x		x
		Trusted-Third-Party	Full		Full		Full		Full	
	Requiring-Explicit-Consent	Requiring-Explicit-Consent		x		x		x		x
		Non-Requiring-Explicit-Consent	Full		Full		Full		Full	
	Unlinkable	Unlinkable		x				x		x
		Linkable	Full		Not	x	Full		Full	

Note: The results are based on the rating framworks proposed by Bonneau *et al.* (2012) and Mayer *et al.* (2016).