Future Generation Computer Systems 80 (2018) 430–447

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Evaluation of transaction authentication methods for online banking☆

Sven Kiljan [a,b,c,*], Harald Vranken [a,c], Marko van Eekelen [a,c]

[a] Faculty of Management, Science & Technology, Open University of the Netherlands, P.O. Box 2960, 6401DL Heerlen, The Netherlands
[b] Economy & Management, NHL University of Applied Sciences, P.O. Box 1080, 8900CB Leeuwarden, The Netherlands
[c] Faculty of Science, Radboud University, P.O. Box 9010, 6500GL Nijmegen, The Netherlands

HIGHLIGHTS

- We introduced a qualitative evaluation mechanism for online banking authentication.
- Seven raters examine 12 online banking authentication methods with our mechanism.
- Bank-issued authentication devices overall have a qualitative very good fit.
- Most user-owned devices fit poorly compared to bank-issued devices.

ARTICLE INFO

Article history:
Received 1 September 2015
Received in revised form
17 May 2016
Accepted 22 May 2016
Available online 30 May 2016

Keywords:
Online banking
Authentication
Evaluation

ABSTRACT

Authentication is a major research topic in the information security field. Much has been written about assessing entity (user) authentication methods, but there is a lack of literature concerning the evaluation of financial transaction authentication in online banking. Entity authentication methods have been systematized by quantifying their qualitative aspects, but there is no evaluation mechanism which also places the additional characteristics of transaction authentication in a user-centric context. Based on an existing mechanism which quantifies accessibility, memorability, security and vulnerability characteristics in entity authentication methods, we propose feasibility as an additional dimension which quantifies aspects related to the secure usability of transaction authentication methods. We also propose the use of this evaluation mechanism by multiple raters to reduce personal bias. Four implemented and eight proposed authentication methods for online banking were evaluated by seven experts. The results indicate that the mechanism can be applied on a wide range of authentication methods, since it is able to evaluate methods based on different information schemes. However, care must be taken that evaluations are performed by multiple experts, due to the amount of subjectivity inherent in the mechanism and in the different opinions of the raters.

1. Introduction

Two forms of authentication can be used in online banking to authorize financial transactions [1]. Entity authentication is concerned with proving the identity of an online banking user, similar to authentication for other online services (email, instant messaging, etc.). Transaction authentication concerns the certainty that financial transactions (the destination account number, the amount of money, etc.) are deliberately authorized by the user. Current evaluation mechanisms of entity authentication methods do not take the specifics of online banking environments into consideration. A mechanism which also evaluates and compares aspects specific to transaction authentication is missing. Such a mechanism should take into account that transaction authentication methods can rely on an active role of the user to provide the security the method needs. Banks slowly start to introduce transaction authentication methods which require users to verify information received by the bank on bank-issued trusted devices and on user-owned mobile devices. The possible reliance on the user's actions and the trustworthiness of what the user observes should also be considered when comparing authentication methods.

The goal was to evaluate different implemented and proposed online banking authentication methods to identify points for

☆ This article is a product of the Dutch Research Program on Safety and Security of Online Banking. The research program is funded by the Dutch banking sector (represented by the Dutch Banking Association), the Police Academy, and the Dutch National Police.

* Corresponding author at: Faculty of Science, Radboud University, P.O. Box 9010, 6500GL Nijmegen, The Netherlands.
E-mail address: sven@kiljan.org (S. Kiljan).

http://dx.doi.org/10.1016/j.future.2016.05.024
0167-739X/

improvement. Our contribution includes an examination of different proposed evaluation mechanisms and our own proposal. We extended an existing mechanism with aspects related to the feasibility of using an authentication method securely. The new aspects cover the taxation of the user's cognitive capacity through expansion of the user's work flow, the ability for security to be (willingly or unwillingly by the user) circumvented and the lack of function and information clarity through the user interface and in- and output channels. The mechanism we propose can be used to evaluate online banking authentication methods in a way which takes the active role of the authenticating user into consideration. Seven raters performed an evaluation of 4 implemented and 8 proposed authentication methods.

The rest of this paper is structured as follows. Section 2 starts with an overview of the background material our work builds on. This includes sources for the evaluation mechanisms we examined, articles about secure usability aspects in information security, and proposals for transaction authentication methods. Different proposed evaluation mechanisms are compared and our choice for Renaud's mechanism is explained in Section 3. We give an overview of Renaud's mechanism in Section 4. The new feasibility dimension is introduced in Section 5, which accounts for the secure usability of the authentication method. In Section 6 it is noted how Renaud's mechanism and our expansion can be used by multiple raters to come to a single answer with less personal bias. We apply the original mechanism and the new dimension on four implemented and eight proposed online banking authentication methods, which are briefly described in Section 7. Considerations for the evaluations are noted in Section 8 and the results can be found in Section 9. We wrap up with limitations, discussion and further research in Section 10, and our concluding remarks in Section 11.

2. Background and related work

In this section, we note the most influential past work on which we base our contribution.

2.1. Authentication evaluation mechanisms

Renaud introduced a mechanism which quantified the qualitative characteristics of user authentication systems [2]. Aspects related to security and usability are given values based on qualitative characteristics to calculate a deficiency value over the aspects' respective dimensions. This approach allows comparisons of authentication methods by comparing weighted values without losing sight of important details. Values can be compared on three levels: aspect, dimension and overall. Since the environment in which an authentication method is used can have a positive or negative effect on its security and usability, Renaud also introduced environmental factors. These are modifiers that represent the influence an environment has on each dimension, and allow comparisons of authentication methods in their respective environments.

Mihajlov et al. present a conceptual framework, which uses Renaud's quality criteria and their own predefined quantification approach [3,4]. Differences with Renaud include an alternative mathematical model, and a reduction in the number of evaluated dimensions.

Another framework with a similar goal to Renaud's mechanism was proposed by Bonneau et al. [5]. Aside from security and usability, their framework also took deployability aspects into account. This framework only evaluates aspects on a single level and does not assign numerical values.

All noted evaluation mechanisms and frameworks are further discussed in Section 3.

2.2. Secure usability aspects

Yee provides a list of design principles for a secure usable design of systems [6]. A criteria of each principle was that it is fairly obvious that its violation would equal the introduction of a security vulnerability. They are proposed as guidelines for system designers to keep in mind.

Herley promotes the idea that users are economical instead of lazy in their decision to follow security instructions [7]. The cost of direct damage is seen as a risk when security advice is ignored, but the far greater cost of indirect damage due to actually following security advice is often not considered. It is this larger cost that makes users reject security advice, since the trade-off (in terms of the (perceived) reduced risk versus the (perceived) increase in user effort) is not considered worthwhile. In a follow-up, Herley explains how valuable a user's time is and how security has to compete for this time in today's information overloaded society [8]. He gives valuable advice to increase the acceptability of security instructions. The advice that relates most to our research results is that users should only be given instructions of which it can be expected that they will be followed.

2.3. Proposed online banking transaction authentication methods

Many authors have proposed conceptual improvements for transaction authentication in online banking. The proposals of Starnberger et al., AlZomai et al., Weigold and Hiltgen and Li et al. present different approaches to protect against attacks in which transaction data created by the user is modified before it reaches the bank for further processing [9–12]. While the approaches are conceptual, they are defined in enough detail to analyze qualitatively.

3. Choosing an evaluation method

For our survey, we wanted to compare different authentication methods implemented by banks and from academical proposals on both security and usability related aspects. We chose a qualitative approach, in which the availability or lack of specific characteristics would be observed. An advantage of this approach is that it produces comparable results. It also scales well when comparing more authentication methods, since only qualitative data is collected and analyzed. Measuring quantitative characteristics takes more effort for each evaluated authentication method and has a risk that the higher level of detail will not provide added value for comparisons. A disadvantage of examining qualitative characteristics is that results may not be reproducible since the observation is never completely objective. However, variance between observers can be reduced by stating the characteristics clearly.

We initially looked at rubrics as a base for the evaluation method. Rubrics are structured scoring guides which consist of specific pre-established performance criteria used to evaluate the quality of student work [13]. A holistic rubric provides a score based on the overall quality, proficiency or understanding of the specific content and skills. This rubric type evaluates student work on a single level. There are also analytic rubrics, which give scores for specific aspects of student work and a summed total score, representing assessment on two levels. Holistic rubrics take less time to use while analytic rubrics provide specific performance feedback, giving insight in a student's strengths and weaknesses. An overview of both is shown in Fig. 1.

As noted, we wanted to evaluate methods based on both security and usability. Only an overall score for each authentication method would not tell us whether something is either secure or usable. Therefore, the analytic approach seemed more suitable.

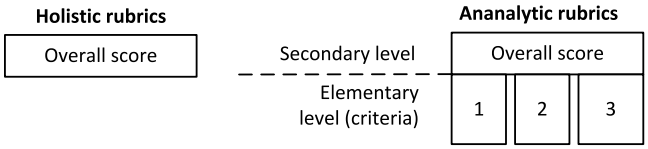

Fig. 1. The levels and outputs of different rubrics types.

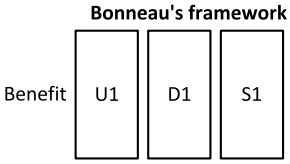

Fig. 2. The single level outputs of Bonneau's framework.

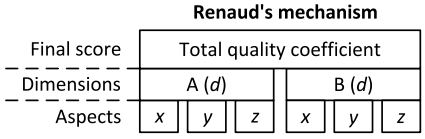

Fig. 3. The levels and outputs of Renaud's mechanism.

Instead of starting from scratch, we evaluated different proposals for evaluating authentication methods qualitatively to see what we could use as a base for our work.

Bonneau et al. introduced a framework (hereafter referred to as Bonneau's framework) for comparative evaluation of web authentication methods with a specific focus on user authentication on the web through uncontrolled client computers [5]. An overview of the framework's outputs is shown in Fig. 2. The 25 criteria in the usability, deployability and security dimensions represent what could be perceived as the characteristics provided by an ideal authentication method, and are therefore referred to as benefits. The deployability dimension is a combination of usability criteria (e.g. accessible to users with disabilities, independence of the installed browser, etc.) and economical criteria (e.g. negligible cost per user, and whether the authentication method is non-proprietary), and would definitely be useful when considering authentication methods which need to be deployed to a large number of users (such those of banks).

One disadvantage of Bonneau's framework is that the output is only on a very detailed level, and lacks a 'total' score which allows easier overall comparisons between evaluated authentication methods. Each criteria gets an 'offers the benefit', 'almost offers the benefit' or 'does not offer the benefit' value, which for some criteria is quite ambiguous and can therefore be interpreted in multiple ways by different raters. As noted earlier, it is possible to reduce the variance in observations, but only if the criteria are very narrowly defined. Furthermore, the authors recognize that weights of criteria can change based on specific goals for which authentication methods are compared, and see this as a reason not to assign weights to the individual criteria at all.

We also looked at a mechanism introduced by Renaud (hereafter referred to as Renaud's mechanism), which is used to compare the quality of web authentication methods [2]. Renaud's mechanism can be used for feature analysis of authentication methods and provides quantified scores on overall, dimension and aspect levels. Four equally weighted dimensions are recognized: accessibility, memorability, security and usability. Each dimension has three equally weighted aspects, each represented by a value that is constructed from either multiple criteria or from a single criteria which can have one of three or four specifically defined values.

Renaud's mechanism is closer to the idea of analytic rubrics compared to Bonneau's framework, which can be seen by comparing Figs. 1 and 3. Like analytic rubrics, Renaud's mechanism applies pre-established and specifically defined performance criteria to qualitatively rate different aspects of some work while also providing an overall score. Aside from the most detailed 'aspects' (represented by x, y and z in each dimension), it also provides intermediate deficiency (d) values which can be used for comparisons between authentication methods based on specific dimensions. As noted by Mertler who cites Trice [13,14], the process of converting rubric scores to student grades and descriptive feedback involves more logic than math. In the case of Renaud's mechanism the resulting quantified values have no mathematical context, nor is the source arbitrary. The values are only used as weights for easy comparisons, and are established using specific and detailed rules.

Another concept of Renaud's mechanism is the environmental factor. Each dimension has one or two environmental factors which act as modifiers for the dimension's deficiency value. These factors allow aspects from a dimension to weigh heavier or lighter depending on how well the environment supports the dimension. Environmental factors make it possible to compare authentication methods, where the environments' influence is included in the comparison.

We were inclined to use Renaud's mechanism as a base for our work, since it closely matches analytic rubrics. It is vital that the to-be observed criteria are described in as much detail as possible, and Renaud's mechanism gives a more detailed description on more levels compared to Bonneau's framework. The output of the mechanism allows comparisons of authentication methods on different levels, which makes it easier to spot where a method is strong and where it could be improved. While environmental factors are not directly relevant for our research, they can be used by other researchers to compare our results with evaluated authentication methods from fields other than online banking. This does not imply that Bonneau's framework is completely inappropriate. The economical aspects of the deployability dimension are something that Renaud's mechanism does not have. A bank would most certainly be interested in comparing the economic feasibility of authentication methods.

We also considered the work of Mihajlov et al. [3,4], who presented a conceptual framework (hereafter referred to as Mihajlov's framework) partly based on the qualitative characteristics provided by Renaud for usability [3,4]. In this framework, the number of dimensions are reduced to two: security and usability. One other difference is that the conceptual framework allows raters to more explicitly define how several criteria apply to an authentication method.

Mihajlov's framework has a heavy focus on the values of its dimensions and, derived from these two values, the total quality value as an end result of an evaluation. This is similar to Renaud's mechanism. However, the reduction in number of dimensions reduces the output of the framework. Renaud's mechanism provides separate output values for quality criteria related to usability (through the total values of the accessibility and memorability dimensions) and security (through similar values for the vulnerability and security dimensions), while Mihajlov's framework only provides overall values for usability and security. This makes Renaud's mechanism more transparent on the second (dimension) level. Furthermore, while Mihajlov's framework allows raters to more precisely define the applicability of some of its criteria, this makes it more complex for raters to evaluate the system while it is unclear what the added value is of such precision on the end result.

In the end, we chose Renaud's framework since its use is more clearly defined and its output is more transparent compared to the frameworks provided by Bonneau et al. and Mihajlov et al.

Table 1
Modifiers of environmental factors and aspects in Renaud's mechanism. Environmental factors (marked in gray for clarity) influence all aspect values within their respective dimensions.

Dimension	Environmental factor/aspect	Value modifiers
Accessibility	Environmental factor: Control of Environment	Controlled (= 1.00), uncontrolled (= 1.50)
	Aspect: Special requirements	Hardware configuration (+0.33), software configuration (+0.33), technical expertise (+0.33)
	Aspect: Convenience	Enrollment time (+0.25), key replacement time (+0.25), authentication time (+0.50)
	Aspect: Inclusivity	Cognitive excluded (+0.33), mobility excluded (+0.33), sensory excluded (+0.33)
Memorability	Environmental factor: Frequency of use	Daily (= 0.50), weekly (= 1.00), monthly or less (= 1.50)
	Environmental factor: Forced Renewal	Not enforced (= 1.00), enforced (= 1.50)
	Aspect: Retrieval Strategy	Fully recognition-based (= 0.00), recall-based with cues support (= 0.50), recall-based (= 1.00)
	Aspect: Meaningfulness	Self-assigned & deducible through special scheme (= 0.00), self-assigned & meaningful to user (= 0.33), self-assigned but not necessarily meaningful or deducible (= 0.67), arbitrarily assigned (= 1.00)
	Aspect: Depth of Processing	No effort (= 0.00), particular level (= 0.33), visual mechanism (= 0.67), rehearsal-based (= 1.00)
Security	Environmental factor: Risk	No damage when compromised (= 0.50), damage to user (= 1.00), damage multiple users (= 1.50)
	Environmental factor: Security Motivation	Sanctions can be applied to irresponsible users (= 1.00), sanctions cannot be enforced (= 1.50)
	Aspect: Predictability	Authentication key is unpredictable (= 0.00), only by friends/family (= 0.50), widely predictable (= 1.00)
	Aspect: Abundance	Range of keys is $\geq 2^{64}$ (= 0.00), $\geq 2^{40}$ and $< 2^{64}$ (= 0.50), $< 2^{40}$ or unique and irreplaceable (= 1.00)
	Aspect: Disclosure	Impossible to disclose (= 0.00), possible by shoulder surfing (= 0.50), easily by user/attacker (= 1.00)
Vulnerability	Environmental factor: Auditing	System applies auditing (= 1.00), does not apply auditing (= 1.50)
	Confidentiality	Key is not revealed or cannot be reused (= 0.00), key is partly revealed (= 0.50), full key is revealed (= 1.00)
	Privacy	No personal details required (= 0.00), allowed to use (= 0.50), required to use (= 1.00)
	Break-/Crackability	Does not apply (= 0.00), vulnerable to research-based attacks (= 0.33), dictionary/brute-force attacks (= 0.67), keylogging (= 1.00)

4. Renaud's mechanism at a glance

An overview of the dimensions' aspects in Renaud's mechanism, their criteria and environmental factors is given in Table 1. We give a short description of the formulas used for aggregating the values of aspects to dimension deficiencies and from dimension deficiencies to the total quality coefficient. The same is done for applying the environmental factors.

Each dimension has three aspects (x, y and z). Each aspect has a minimum value of 0 (representing that the authentication method provides the highest quality or best fit for a particular aspect) and a maximum value of 1 (representing the lowest quality or worst fit). Each aspect is seen as equally important and therefore has an equal weight. The same is true for the different modifiers which define each aspect's value, with a single exception. For the convenience aspect in the accessibility dimension, the authentication time is seen as more important since users often authenticate, while both initial enrollment in the system and the replacement of lost security credentials happen less often.

The aspect values are used to calculate deficiency value d for each dimension using $d = \sqrt{x^2 + y^2 + z^2}$. d can be used to see the quality an authentication method has in a specific dimension, where a lower value is a higher quality. Based on the minimum and maximum values of the aspects, $\min(d) = \sqrt{0^2 + 0^2 + 0^2} = 0$ represents the highest quality while $\max(d) = \sqrt{1^2 + 1^2 + 1^2} = 1.732$ represents the lowest quality an authentication method can offer in each dimension.

In formulas, ad represents the deficiency value for the accessibility dimension, md does the same for the memorability dimension, etc. The total quality coefficient represents how well an authentication method fits all dimensions, and can be calculated by $\overline{eq} = \max(\overline{eq}) - (ad + md + sd + vd)$, where the maximum total quality coefficient $\max(\overline{eq}) = \max(d) * 4 = 6.93$, based on the summed maximum deficiency values of the four dimensions. A higher total quality coefficient value represents a higher overall quality.

Each dimension also has one or two environmental factors, representing the influence characteristics of the environment over the aspects within their respective dimensions. Environmental factors are represented in formulas by their shortened names. Whereas the total quality coefficient value is used to determine the overall quality an authentication method has on its own, the environmental quality coefficient represents the same under influence of environmental factors. To calculate the environmental quality coefficient value, first the total environmental deficiency has to be calculated: $\overline{d_{env}} = ad * control + md * freq * renewal + sd * risk * motivation + vd * auditing$. Then, the environmental quality coefficient can be calculated using: $\overline{eq_{env}} = \max(\overline{d_{env}}) - \overline{d_{env}}$, where $\max(\overline{d_{env}}) = \max(ad) * \max(control) + \cdots = 12.98$. Similar to the total quality coefficient, a higher value of the environmental quality coefficient represents a higher overall quality.

5. Expanding Renaud's mechanism with the feasibility dimension

Renaud notes that users are required to authenticate themselves to use computer systems and web sites securely [2]. Her evaluation mechanism targets user authentication methods in web environments, which correspond with entity authentication in online banking. Unfortunately, the mechanism misses some aspects which are vital to the secure use of an authentication method, especially transaction authentication methods. The four dimensions focus on aspects concerning usability (accessibility and memorability) and technical security (security and vulnerability). While the dimension memorability concerns usable security, it is limited to information in the authentication method which the user has to remember. There are other usable security aspects which are not part of Renaud's mechanism, but which are relevant to transaction authentication.

We introduce the new feasibility dimension. Its three aspects and environmental factor concern the feasibility of secure use of an evaluated authentication method. 'Secure use' is not simply a combined phrase to keep security and usability in mind as two

Table 2
The feasibility dimension's environmental factor (in gray), aspects and their modifiers.

Environmental factor and aspects	Modifiers
Environmental factor:	Users can correct mistakes within a reasonable time frame without repercussions. (= 1.00)
User correction	Users are not allowed to correct errors without repercussions. (= 1.50)
Aspect: Work flow expansion	User does not have to perform additional actions. (= 0.00) Some existing user actions are repeated as part of the authentication procedure. (= 0.50) New user actions are introduced to the user's work flow to support authentication. (= 1.00)
Aspect: Circumvention	The system's default state is insecure. (+0.33) The user interface does not support secure user behavior. (+0.33) User could subvert security due to inconvenience. (+0.33)
Aspect: Clarity	Interface gives a false impression of an ability or lacks the right information to ascertain its limits. (+0.33) Information necessary to make a good decision before an action is taken is inaccurate or missing. (+0.33) Input and output channels can be spoofed or are corruptible. (+0.33)

aspects. Instead, it relates to the challenge of having a security system which is feasible for users to use in a secure way. Herley notes that a user's capacity for effort (basically a combination of time and energy) is one of the most valuable and scarce resources available in the information security field. If a user is expected to spend his or her resources inefficiently or ineffectively on security, it can only be expected in return that security instructions will be ignored or circumvented [8].

The qualitative characteristics which are quantified for the aspects related to feasibility can be found in Table 2. As noted in Section 3, Renaud's mechanism does not have a deployability dimension. While we do recognize the added value of such a dimension, we decided that deployability does not fit the user-centric context of our scope. The cost of deployment does not necessarily improve the security or usability of authentication methods.

Note that the exact deficiency and coefficient values are not interesting. It is the relative weight of each dimension which allows comparisons to improve authentication methods. We can learn much by comparing the fulfillment of dimensions by each authentication method and observing where the low hanging fruit of improvements can be found, and which dimensions provide challenges.

As shown in Table 2, the different values of the work flow expansion aspect increase linearly, while the criteria of the circumvention and clarity aspects are proportionally equal. We consider the criteria for the circumvention and clarity aspects equal, which is why they have equal values. Similarly, the values of the work flow expansion aspect's criteria levels are based on the order in which each level taxes the aspect. Use of equal proportions reduces possible bias when applying the evaluation mechanism. Someone who applies the mechanism could decide that of an aspect one characteristic is more important than another and adjust the values accordingly. However, results of different evaluations are only comparable if the used mechanism to evaluate each authentication method is the same.

We describe the new aspects and environmental factor before we note the effects of the new dimension on the formulas of Renaud's mechanism.

5.1. Aspects of the feasibility dimension

As with the dimensions in Renaud's mechanism, the new feasibility dimension has three aspects.

5.1.1. Work flow expansion

Authentication can hardly be described as a desirable or enjoyable task. It is a mandatory procedure which distracts from other tasks the user needs or wants to conduct. Therefore, there is not much incentive for users to spend more time and cognitive capacity than strictly required when authenticating. The user's time should only be spend if the benefits outweigh the costs. Otherwise security advice is rejected [7,8], resulting in insecure use of authentication methods.

The cost in time for enrollment, recovery and authentication is represented in Renaud's mechanism by the accessibility dimension's convenience aspect. The question whether a user is cognitively capable of using an authentication method is also covered by the same dimension's inclusivity aspect. Work flow expansion focuses instead on the question if and how the user's normal work is expanded solely for the purpose of authentication. An action is defined as either something the user is expected to do physically or cognitively. We recognize three distinct levels:

- The user is not required to perform any actions aside from possibly remembering and entering something. Memorability is excluded since it is already rated in its own dimension. The installation and configuration of hard- and software is also excluded since these are quantified by the accessibility dimension's special requirements aspect.
- The user must perform some actions redundantly. No additional actions are introduced, but the user must perform some of the existing actions twice (or more). An action has to be executed multiple times, at least once as part of a regular work flow and at least once as part of the authentication procedure. An example would be the entry of the same transaction data in both the user's computer and in an authentication device provided by the bank.
- Additional actions are introduced in the authentication procedure which require cognitive processing, such as when a user has to verify transactions by comparing transaction data as entered and as received by the bank (and shown to the user on a secure device) on equality.

5.1.2. Circumvention

Another feasibility aspect is the user's (dis)ability to circumvent the security system. Yee describes several principles for user interaction design in secure systems [6].

- The principle of the path of least resistance notes that the natural way to use a system should be the secure way. Since users use their physical and mental effort sparingly, the path of least resistance is the natural path for a user to follow and should therefore be the secure way to use the system. The ultimate path of least resistance is for the user to do nothing. Not doing anything is classified as an action which a system should also securely handle. The system must be secure against attacks while it is not being used.
- User errors should not be accepted as a source of security problems [15]. It should not be possible for the user to subvert security unintentionally due to that the user interface does not support secure user behavior. An example is given by Yee, in which an icon of a lock can be clicked for security

information [6]. The associated action (examining security information) could be overlooked by the user if the icon does not look like it can be interacted with.
- It should not be possible for the user to subvert security intentionally (e.g. due to the negatively perceiving amount of required effort). Inconveniences for the user increase the probability that the system will be used insecurely. If we use the lock icon again as an example and make it a button (so it is clear that it can be interacted with), a user can still opt-out of examining security information by simply not clicking the button. It is a security risk if the user is expected to perform an unenforceable action (e.g. verifying information on correctness).

5.1.3. Clarity

The final feasibility aspect is the clarity of the system towards the user. This concerns clarity in both information and offered functionality.

- Yee notes the principle of clarity [6], which states that information should be accurate and available before a user action is taken, and also that the user interface does not present misleading, ambiguous or incomplete information. While the principle of clarity is defined from the perspective of a user who is granting security authorities, the same principle also applies when the act of authentication equals an act of authorization. Reliable information is required to make a good decision. The integrity of the information and its presentation should be protected. If they are not, the information (as interpreted by the user) is unreliable and unfit to make secure decisions on. An example of unreliable information would be aggregated transaction data, such as the number and sum of a set of transactions offered to the user on a secure device for verification. In this scenario, an adversary could change the destination account numbers of the transactions without the user being able to verify it. If non-aggregated information would be used, the user could check all the critical values (such as of each transaction the destination account number and the amount of money).
- As remarked by Yee's principles of identifiability and expected ability, the user interface itself should be identifiable and unambiguous regarding its abilities. If it is not, false user expectations can lead to wrong decisions with serious consequences. The use of ambiguous terms for functions and labels can obfuscate what the authentication method can and cannot be used for. Likewise, if functions are unidentifiable, the user is vulnerable to error through inadvertent collision or intentional masquerading, on which social engineering attacks thrive.
- Finally, Yee's principle of the trusted path describes that input and output channels should be secure against spoofing or corruption. An example of an insecure channel in authentication would be the use of SMS text messages to send critical decision information to a user, which can be spoofed [16]. Also, smartphones used for receiving text messages are vulnerable to malware, which compromises both the integrity and availability of any received text messages since they can be spoofed, changed and forwarded to another phone while being kept hidden from the user [17].

5.2. Environmental factor: user correction

Yee notes the principle of revocability [6]. Facilitating revocation is needed to accommodate users' ability to correct slip-ups and errors. If a correction can be made without repercussions within the system's environment, the available space for damage resulting from user errors is reduced. The stakes for additional authentication actions a user has to perform as part of an expanded work flow are much higher when mistakes cannot be corrected. In this case, a larger burden is placed on the user since there is less room for error. Likewise, consequences of circumventing the system are more serious if the user has no way to make amends. The same is true for unclear information, tasks, and in- and output channels, for which the lack of clarity will have a bigger impact if slip-ups are not correctable.

The ability of users to make corrections is noted as an environmental factor that influences all feasibility aspects. If users can make corrections, the environmental factor does not have any influence. When the ability is absent, all aspects of the feasibility dimension are weighted heavier.

5.3. Adapted formulas

Based on the formula used for other dimensions, the deficiency of the feasibility dimension (fd) can be calculated by: $fd = \sqrt{x^2 + y^2 + z^2}$ where x, y and z are the values of the dimension's aspects. Similar to the original four dimensions, the individual values can be used to compare authentication methods on specific aspects while the deficiency can be used to measure the quality an authentication method has in a specific dimension.

The total quality coefficient for all dimensions, including the feasibility dimension, is now calculated by $\overline{eq} = \max(d) - (ad + md + sd + vd + fd)$. Note that $\max(d)$ is 8.66 due to the inclusion of fd. The new total quality coefficient can be used to measure an authentication method's overall quality.

The new formula for the total environmental deficiency is: $\overline{d_{env}} = ad * control + md * freq * renewal + sd * risk * motive + vd * audit + fd * correction$ where *correction* is the value for the user correction environmental factor for the new dimension. The environmental quality coefficient is still calculated by $\overline{eq_{env}} = \max(\overline{eq_{env}}) - \overline{d_{env}}$, but the new $\max(\overline{eq_{env}})$ is 15.58 due to the added maximum values of the feasibility dimension's aspects and environmental factor.

5.4. Relative scoring formulas

We chose to improve the readability of the deficiency and coefficient values by converting them to relative values. This also makes it easier to read the effect the additional dimension has on the total quality coefficient. To compare the deficiency of each dimension against its minimum and maximum values in a way that makes a higher value represent a better fit, we use $dp = (1 - \frac{d}{\max(d)}) * 100\%$ to calculate a percentage (adp for the accessibility dimension, mdp for the memorability dimension, etc.). We also calculate an overall percentage for the total quality coefficient using $\overline{dp} = \frac{\overline{eq}}{\max(\overline{eq})} * 100\%$.

Calculating a percentage-based score for the total environmental quality coefficient would not have any added value. The resulting percentages would be the same as $\overline{dp}$ due to the use of the same fractions.

6. Multi-user evaluation

Based on Renaud's mechanism as described in Section 4, we propose an expansion in Section 5. To test whether both can give useful results, we applied the evaluation mechanism on 4 implemented and 8 proposed transaction authentication methods, which are described up ahead in Section 7.

With qualification mechanisms there always is some subjectivity involved. For example, Renaud's mechanism asks the rater whether technical expertise is required to apply the authentication method (as part of the accessibility dimension's special requirements aspect). Technical expertise is quite an ambiguous term.

Table 3
Example answers to questions related to a single authentication method. The majority defines the answer that will be used to evaluate an authentication method in Renaud's mechanism.

	Question			
	1	2	3	4
Rater 1	Yes	Yes	Yes	Yes
Rater 2	Yes	Yes	Yes	Yes
Rater 3	Yes	No	Yes	Unknown
Rater 4	Yes	No	Unknown	No
Rater 5	Yes	No	No	No
Majority	Yes	No	Yes	Unknown

Does installation of software on a home computer require technical expertise? Or the installation of an application on a smartphone? Another example would be the question of whether the method is time-consuming. A case can be made for that enrollment and replacement each take a large amount of time, since the user at least has to visit a bank's office or needs to wait until a new/replacement device or code arrives in the mail. However, for authentication it is up to the rater to decide whether something is time-consuming or not.

To compensate for this subjectivity, it is possible to apply the same evaluation mechanism on the same authentication methods multiple times by different raters. Renaud's mechanism (with and without our expansion) has the advantage that it is simple to translate the characteristics that define the aspect values to survey questions which can be answered with either 'yes', 'no' or 'I do not know'. It is not needed for raters to be familiar with Renaud's mechanism when they are provided with a description of an authentication method and a list of questions to answer. The average of an answer can be fed back into Renaud's mechanism to fill in the characteristics that define the aspect values. This is repeated for every authentication method to be evaluated. Of course, it would be recommended to choose experts to be raters to come to a meaningful answer.

Table 3 gives an example of how four questions about an authentication method are answered by five raters.[1] For question 1 and 2, it is simply the majority that defines what the average answer is. Although one rater did not know the answer for question 3, the answer would not have mattered since a majority had been reached by three other raters. question 4 shows an uncomfortable situation in which a majority could not be reached. This can happen when one or more raters do not know an answer (as in the example) or when an even amount of raters would provide equally distributed answers to the question. Since in this case the answer would be neither 'yes' or 'no', half of the relevant modifier's value (as shown in Tables 1 and 2) would be assigned. That would be 0.17 (or 1/6) for a value that is worth 0.33 (or 1/3) of an aspect's full value, 0.25 (or 1/4) for a value that is worth 0.50 (or 1/2) of an aspect's full value, etc.

In Section 8.4 starting at page 26 we describe how we let different raters apply Renaud's mechanism by itself and including our expansion. The results of the multi-user aspect of our experiment can be found in Section 9.4 starting at page 34.

7. Evaluated authentication methods

We applied Renaud's mechanism and our expansion on several used and proposed authentication methods. This section briefly describes the methods.

Each transaction authentication method applies an information scheme. We recognize three schemes [18]:

- Traditional transaction authentication (TTA). The method used for entity authentication is (re-)applied to authenticate transactions. User-recognizable transaction information is not used in this scheme.
- Customer verified transaction set authentication (CVTSA). A bank sends transaction information back to the authenticating user for verification.
- Entered single transaction authentication (ESTA). The integrity of transaction information is secured as soon as the information is created by the user.

The chosen identifiers used to refer to the authentication methods in the rest of this paper are based on the following format:

`<issuer> <characteristic> <user action and information type>`

`<issuer>` is the unique identifier of either a bank or a proposal's first author's last name.

`<characteristic>` is a short description (possibly abbreviated) of the method's main characteristic(s). These values can be:

- Entry. Applies to devices which require the user to enter transaction data on the device.
- hPIN/hTAN. A specific name for a proposal by Li et al. [12].
- Scan. Applies to devices which uses an optical sensor to scan data from a user's computer display.
- SMS (Short Message Service). Applies to methods which use SMS for transferring authentication information.
- USB (Universal Serial Bus). Applies to devices with which users interact and which are connected to a user's computer through USB.
- USB CR (Universal Serial Bus Card Reader). Applies to card readers without a user interface, connected through USB to the user's computer.
- ZTIC. A specific name for a proposal by Weigold and Hiltgen. [11].

`<user action and information type>` is an abbreviation that specifies which kind of action (**N**one, **V**erify or **E**nter) the user performs or is expected to perform for what kind of information (**N**one, **A**ggregated or **N**on-**A**ggregated) when using the method. With **N**one, no additional action is necessary. When **V**erify is specified, the user is expected to verify transaction information that the bank received and that was sent back to an authentication device in possession of the user. With **E**nter, the user has to enter critical transaction information on an authentication device. **V**erify and **E**nter relate to either **A**ggregated (such as the number of transactions and the total amount of money of a set of transactions) or **N**on-**A**ggregated transaction information (such as the destination account number and amount of each transaction).

The following combinations of abbreviations are used:

Abbreviation	User action on information	Transaction information processed by authentication device
NN	None	None
VA	Verify	Aggregated
VNA	Verify	Non-aggregated
ENA	Enter	Non-aggregated

What follows are identifiers and brief descriptions of the evaluated authentication methods. The first four are based on methods used by banks, each at least used by half a million customers on a regular base. The other eight are proposals by different authors.

Bank USB CR NN (Bank Universal Serial Bus Card Reader None None)

This method consists of a bank-issued USB smart card reader connected to the user's computer and supporting software. An

[1] The example uses five raters for clarity. The evaluations discussed in this paper are performed by seven raters.

Fig. 4. A USB smart card reader.

Fig. 5. A USB smart card reader with its own display and keypad.

Fig. 6. Prototype of the hPIN/hTAN.

example of such a reader is shown in Fig. 4. This device is used in combination with a user's bank card to login to the bank site and sign transactions shown on the user's computer. The bank card requires a PIN to unlock its functionality, which is entered on the user's computer. The user does nothing with any kind of transaction information in the authentication process (explaining the 'None None' or NN). Therefore, this method applies the TTA information scheme.

Bank SMS VA (Bank Short Message Service Verify Aggregated)

An SMS text message is sent to the user's mobile phone during transaction authentication when a set of transactions is ready to be authenticated. The message contains aggregated information (the total amount of money and the number of transactions) and a one-time password. The one-time password must only be used if the total transaction amount in the text message corresponds with the value shown on the user's computer. This method applies the CVTSA information scheme since users are expected to verify aggregated data (VA) of a set of transactions.

Bank USB VA (Bank USB Verify Aggregated)

Users are issued a device by their bank, of which Fig. 5 gives an example. The device is similar to Bank USB CR NN in that it features a card reader and a USB connection, but it also has a display and buttons for user interaction. This authentication method allows the user to verify aggregated transaction information of a transaction set (the number of transactions and the total amount of money) on the device during transaction authentication. Confidentiality and integrity of information between the bank and the device is protected. A browser plugin on the computer translates the USB commands to network commands to be sent to the bank site and vice versa. This method applies the CVTSA information scheme since users are expected to verify aggregated data (VA) of a set of transactions.

Bank Scan VNA (Bank Scan Verify Non-Aggregated)

This is another method which uses a bank-issued authentication device. The device is not connected to the user's computer. Interaction relies on a keypad, display, camera and smart card slot. In combination with a bank card and a PIN, the device is used to verify and sign transactions. During transaction authentication, non-aggregated information concerning individual transactions (destination account number and the amount of money) is projected on the display of a user's computer in a structured image and registered by the camera. The user enters a verification code shown by the device's display in his or her computer when confirming transactions. This method applies the CVTSA information scheme since users are expected to verify non-aggregated data (VNA) of a set of transactions.

Starnberger Scan VNA (Starnberger Scan Verify Non-Aggregated)

Starnberger et al. propose a transaction authentication method using an application on a user-owned mobile device [9]. The camera of the device is used to scan a QR code from a personal computer, which contains (confidentiality and integrity protected) non-aggregated transaction information and a verification code. The user can enter the code on his or her PC to verify the transactions shown on the device. This method applies the CVTSA information scheme since users are expected to verify non-aggregated data (VNA) of a set of transactions.

AlZomai Scan+SMS VNA (AlZomai Scan+SMS Verify Non-Aggregated)

AlZomai et al. propose something similar to Starnberger et al. Instead of scanning a QR code, they suggest to scan plain-text transaction details from a computer screen using the device's camera [10]. The scanned data is verified against SMS text messages received from the bank. If the data matches, a verification code is shown on the mobile device to enter on the user's computer. Users are still expected to verify that non-aggregated data (VNA) on their computer screen is correct, which is why this method also applies the CVTSA information scheme.

Li hPIN/hTAN VNA (Li hPIN/hTAN Verify Non-Aggregated)

A bank-supplied device is proposed by Li et al. [12]. The hPIN/hTAN consists of a USB connector, display and a single 'OK' button. A prototype of the device is shown in Fig. 6. Software on the user's computer is used to forward secure messages between the device and the bank. For entity authentication using hPIN, the bank sends a random digit (0–9) substitution table to the device for each new session, to be shown to the user. The user enters the required PIN in his or her computer using substituted digits. Only the bank and the device have access to the substitution table, which prevents the user's computer from eavesdropping the PIN.

With hTAN for transaction authentication, users enter critical transaction details on their keyboards, which is simultaneously sent to the authentication device's trusted display. During entry, the user verifies that the information is securely entered using the trusted display of the device. One press on the 'OK' button sends the information securely to the bank when it is deemed correct. Due to the verification of non-aggregated data (VNA), the device applies a CVTSA information scheme, although it must be noted

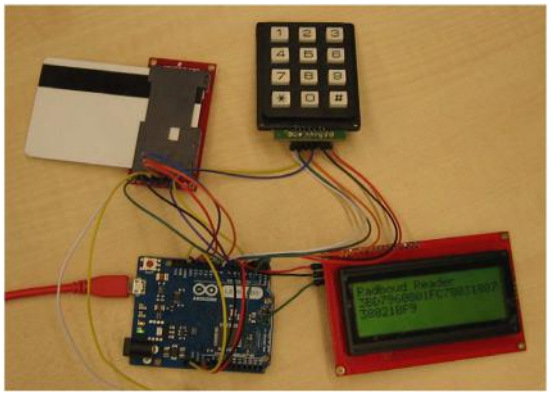

Fig. 7. A prototype USB smart card reader with a display and keypad.

Fig. 8. IBM's ZTIC used to verify a transaction.

that each transaction in a set submitted to the bank is processed individually by the user.

Weigold Entry ENA (Weigold Entry Enter Non-Aggregated)

Several solutions are proposed by Weigold and Hiltgen. Weigold Entry ENA consists of a disconnected, bank-supplied device on which the user enters critical transaction information [11]. A transaction-dependent authorization code (TAC) is created by the device, based on the entered transaction information. The same information and the TAC are entered by the user in his or her computer. The bank receives the information and checks whether it matches the TAC. If valid, the message is accepted. Due to that the user has to enter non-aggregated transaction information (ENA), this proposal applies an ESTA information scheme.

Weigold SMS VNA (Weigold SMS Verify Non-Aggregated)

Another proposal by Weigold and Hiltgen suggests the use of SMS text messages to send critical transaction information received by the bank back to the user for verification [11]. A verification code is also part of the message, which the user can enter on his or her computer to notify the bank that the received data is correct. This proposal also applies a CVTSA information scheme and is quite similar to the use of SMS text messages by Bank SMS VA, with the difference that this proposal presents non-aggregated transaction information (VNA) to the user.

Weigold Scan VNA (Weigold Scan Verify Non-Aggregated)

This is a variation of Weigold Entry ENA. A bank-issued device is used to verify entered transaction details [11]. Transaction data is not entered by the user, but scanned by an optical sensor through a flickering image on the user's computer. A verification code, shown on the display of the bank-issued device together with the critical transaction information, is entered by the user in his or her computer to indicate that information earlier received by the bank is correct. This proposal is similar to Bank Scan VNA and also applies a CVTSA information scheme to make the user verify non-aggregated data (VNA).

Weigold USB VNA (Weigold USB Verify Non-Aggregated)

This proposal is similar to Bank USB VA. It describes a device equipped with a display, keypad and smart card slot, connected with USB to a PC [11]. The device is used during transaction authentication to verify transaction data, so it also applies the CVTSA information scheme to make the user verify non-aggregated data (VNA). This solution also relies on a browser plugin to translate data from the bank server to USB commands. What makes this proposal different from Bank USB VA is that the former lets the user verify non-aggregated data (VNA) of each transaction instead of the total number of transactions and the total amount. An implementation has also been proposed by other authors [19], of which a prototype can be seen in Fig. 7.

Weigold ZTIC VNA (Weigold ZTIC Verify Non-Aggregated)

Finally, Weigold et al. mention the use of Zone Trusted Information Channel (ZTIC), as depicted in Fig. 8. Most banks use SSL/TLS for communication through a secure channel between a user's computer and a bank [1]. ZTIC uses a bank-issued device to put the client-side creation of the SSL/TLS channel outside of the untrusted domain of the user's computer [11]. The device provides a USB connection, a display, two buttons and a smart card slot. A smart card is used for cryptographic functions and storage. Its display and buttons are used by the user to confirm or reject login and transaction requests based on non-aggregated information (VNA), which is why this method also applies a CVTSA information scheme.

8. Applying the mechanism

In this section, we describe how Renaud's mechanism and the feasibility dimension were used to evaluate online banking authentication methods. We note an assumption concerning entity authentication we had to make for several proposals which only specify how transaction authentication is performed. After that, we describe how the new dimension's aspects apply to the evaluated transaction authentication methods. Finally, we provide environmental factor values to represent online banking to aid fellow researchers when comparing our results with their own.

8.1. Proposals which lack entity authentication information

We evaluate several proposals from literature. Some proposals focus exclusively on transaction authentication and not on entity authentication. The former complements the latter, which is why both should be evaluated. We make some assumptions about the use of entity authentication for proposals which do not describe this.

For proposals which only focus on transaction authentication and which do not rely on a bank-issued trusted device, we assume that a password is required for entity authentication and that it is entered in the user's computer. The initial password is chosen by the user. Methods for which we make this assumption are Starnberger Scan VNA, AlZomai Scan+SMS VNA and Weigold SMS VNA.

Weigold Entry ENA, Weigold Scan VNA and Weigold USB VNA rely on a bank-issued device. It is assumed that a user enters a PIN to unlock the device's functionality. The initial PIN is random and can be changed by the user.

Weigold ZTIC VNA and Li hPIN/hTAN VNA both describe entity and transaction authentication methods. ZTIC relies on PIN entry on the user's computer and not on the device itself. Li hPIN/hTAN VNA also relies on PIN entry on the user's computer, but the entered digits are manually substituted by the user using a table provided by the device. We assume for both methods that the initial PIN is randomly chosen and that a user can change it afterwards.

8.2. Applying feasibility aspects

In this section we note how we, as authors of this paper and one of the raters, apply the feasibility aspect on the evaluated authentication methods. The questions we asked the other raters are based on this. Note that the examples we give in this section are those we give from our own perspective. Other raters did not have to agree with these examples when answering the questions.

For each tested authentication method, the work flow expansion aspect is given a value based on whether a user needs to apply additional effort for transaction authentication. If the actions for authentication fit in a normal work flow, a value of 0.00 is given since no additional effort is required from the user aside from what is required for entity authentication. If the actions fit in the user's existing work flow but are redundant (e.g. the user has to perform a specific action twice instead of once), a value of 0.50 is given. Finally, if a work flow is expanded with one or more new kinds of additional actions, a value of 1.00 is given. Additional actions are those which are exclusive to transaction authentication and which are not considered by the other four dimensions. Examples of qualifying actions include comparing and substituting data values. Examples of actions which do not qualify are remembering and entering passwords or PINs, which are already covered by the memorability dimension and by the accessibility dimension's inclusivity aspect.

The circumvention aspect has three characteristics which can increase its value. If users are required to perform security actions which they can skip as part of their work flow, 0.33 is added to the aspect's value. The second characteristic concerns itself with whether the user interface supports secure user behavior. Banks have some control over the user interface with most transaction authentication methods, be it through a web interface, a mobile application interface or a separate user interface on a provided device. An exception is the use of text messages through a mobile phone, which relies completely on an existing user interface which is not tailored to support secure user behavior. Transaction authentication methods which rely on user interfaces which banks do not control get an additional 0.33. Finally, whenever a system's default state is insecure, another 0.33 is added. This last characteristic manifests itself if the user does nothing, yet an adversary can still launch an attack without (further) user action. An example is an adversary which has (remote) access to the user's password and to the user's smartphone. Even if a user does not initiate payments, an adversary can create a session with the bank (with the user's password) and verify transactions (through the user's smartphone).

For the clarity aspect, a minimum value of 0.33 is given to each transaction authentication method because the communication channel between the user's browser and the bank server is corruptible by malware. Some authentication methods rely on other corruptible communication channels between user and bank. The user cannot make an informed decision if all channels can provide inaccurate information. In this case, another 0.33 is added. A user interface which can present misleading, ambiguous or incomplete information is another characteristic which adds 0.33. For example, a browser can have a secure connection with a bank site and show this, whereas a mobile phone's interface for text messages does not.

8.3. Environmental factors

We do not apply environmental factors in our evaluation because we assume that all implemented and proposed authentication methods which we evaluate are in the same environment. Therefore, the values of the environmental factors are the same for each authentication method, and the relative score is the same for both $\overline{eq}$ and $\overline{eq_{env}}$. However, we do give the factor values for the online banking environment. This allows researchers to compare authentication solutions in the online banking environment with those in other environments with different environmental factors.

The control of environment factor in the accessibility dimension is deemed 'uncontrolled' with a value of 1.50. Online banking relies on the Internet, which cannot be exclusively managed by banks.

The frequency of use and forced renewal environmental factors of the memorability dimension concern how easy it is made for the user to remember required knowledge for authentication. We use a value of 1.00 for both. It is assumed that there will be periods in which online banking is used once a week or less (e.g. during a holiday), and that users are not required to renew their passwords or PINs.

There is financial damage in a successful attack. Who is affected depends on several factors. Banks can give reimbursements, but do not always have to do so. Since a single party is affected (the bank or the user), the risk environment factor of the security dimension gets a value of 1.00. While banks in some cases hold users liable for damage, it is not their role to give sanctions as a deterrent against insecure behavior. It can be assumed that sanctions will not be enforced to keep the public image of banks positive, which gives a value of 1.50 to the security motivation environmental factor of the security dimension.

Banks can apply pattern-based recognition of malicious transactions, giving a value of 1.00 to the vulnerability dimension's auditing environment factor.

Transactions are usually non-reversible by the end-user. Therefore, the feasibility dimension's user error tolerance environment factor gets a value of 1.50.

8.4. Performing a multi-user evaluation

As we described in Section 6, one way to decrease the amount of subjectivity when evaluating something is to let multiple raters evaluate the same subject with the same method. We do this with Renaud's mechanism itself (described in Section 4) and with our expansion (proposed in Section 5).

Based on the modifiers as shown in Table 1 on page 9, it can be expected that some parts of Renaud's mechanism will be more sensitive to subjectivity compared to others. To reduce the amount of time required to perform the evaluations we only prepared questions for the most subjective parts of Renaud's mechanism, and all dimensions and aspects of our expansion. Most aspects of Renaud's mechanism are quite objective. For example, whether extra hard or software is required (measured by the accessibility dimension's special requirements aspect) is not based on opinion but on clearly stated specifications of the authentication method. The subjective parts of Renaud's mechanism that we presented to the raters are the requirement for technical expertise (from the accessibility dimension's special requirements aspect) and whether much time is required to perform authentication (from the accessibility dimension's convenience aspect). The other required values for Renaud's mechanism can clearly be derived from the collected specifications of the authentication methods. This allows us to focus most of the raters' attention to our expansion, for which we will let them rate each aspect. Focusing on the subjective parts of Renaud's mechanism and on our added dimension allows us to get an insight in how sensitive the mechanism (original and with our expansion) is to subjectivity.

For each of the 12 authentication methods we asked the same questions in the same order. The order of the authentication methods to evaluate was also the same for all raters due to restraints in the system we used to perform the survey. This order is the same as the order of the authentication methods in the Appendix, from left to right.

The questions are shown in Table 4 in a condensed form. In the survey, the questions were a bit more extensive and each question had some background information that could help raters if they did not understand the context. In addition to the possible answers shown in Table 4 and as discussed in Section 6, each question could also be skipped by answering 'I do not know', and raters were encouraged to pick this option if they could not think of an answer.

Table 4
Summary of the questions asked to the raters for each authentication method.

#	Question	Answers	Relates to	
			Dimension	Aspect
1	Does the user require technical expertise to prepare or use the authentication method?	Yes/no	Accessibility	Special requirements
2	Does the user require much or little time to authorize a transaction?	Much/little	Accessibility	Convenience
3	When replacing password authentication, does the user now have to perform redundant actions?	Yes/no	Feasibility	User effort cost
4	When replacing password authentication, is the user now required to perform new actions?	Yes/no	Feasibility	User effort cost
5	Are one or more of the devices used for authorization protected against remote attacks?	Yes/no	Feasibility	Circumvention
6	Can the user know or check that he or she is using the authorization system of the bank?	Yes/no	Feasibility	Circumvention
7	Is it possible for the user to skip steps in the authorization process?	Yes/no	Feasibility	Circumvention
8	Are all communication channels between user and bank secure against adversaries?	Yes/no	Feasibility	Clarity
9	Of any information the user is required to verify, is the information complete?	Yes/no	Feasibility	Clarity
10	Of any information the user is required to verify, is the information accurate?	Yes/no	Feasibility	Clarity
11	Is the primary user interface capable of showing misleading, ambiguous or incomplete information?	Yes/no	Feasibility	Clarity

The raters were personally asked to participate in the survey and given a personal URL for participation, which they could do at their workplace or at home. Due to the length of the survey, they could pause and continue it at anytime they wanted at any location they wanted, so there was no time pressure. Before, during and after the survey the raters were given the continuous opportunity to ask questions. The raters did not communicate with each other while performing the survey.

For the experiment, we informed ourselves of the rules stated by the ethical review board of Open University of the Netherlands (known as Commissie Ethische Toetsing Onderzoek) whether a review would be required. The review board's main focus is medical examination, and the experiment did not require a review or approval. We were careful in our judgment on whether the experiment was safely performed, and also asked the raters (each of them a researcher and familiar with research ethics) whether they saw any ethical problems before the survey was conducted. The raters were not pressured to participate in the survey, and they were told several times explicitly that they can pause or cancel the survey at any time without stating a reason and without any repercussions. No personal information was asked or collected in the survey. A link between answered questions and personal information of the raters (name and email address) was only used for administrating the survey, and no further personal information was collected or processed. Any questions that we asked about the survey after it was finished were done in person. We reduced risks of personal damage as much as we could by only making raters evaluate the authentication methods from a theoretical perspective. For our experiment we were only interested in the opinions of the raters, not in how they perform actions themselves. Therefore, we did not request the raters to test or use any of the authentication methods (e.g. with their own bank accounts), neither did we make any implication that such an action would be necessary to partake in the survey, nor did we register any such actions.

7 experts rated the authentication methods defined in Section 7 using the questions we prepared. For our experiment, we considered an expert as someone whose research field and work (indirectly) relates to transaction authentication in online banking. 4 raters have a technical background and their research relates to technology that is used in, among others, online banking. The other 3 raters have backgrounds in social sciences, and their research focuses on combating online banking fraud from an organizational perspective (of law enforcement, banks and criminals), and improving the self-defense of users against external threats.

The raters were provided a summary of the workings of all authentication methods. Also, for the proposed methods they were given copies of the work which propose these [9–12].

Section 9.4 continues with the results of the evaluation as performed by multiple raters.

9. Resulting values

This section notes the results of our evaluation of implemented and proposed transaction authentication methods. The content of this section is based on our research data, which can be found in the Appendix.

9.1. Effects of the feasibility dimension

Fig. 9 visualizes the influence the feasibility dimension has on the overall percentages. Every authentication method has two bars. Each top bar illustrates the overall quality of the authentication method within Renaud's mechanism. Each bottom bar does the same, but also includes the feasibility dimension. A value of 100% for any bar represents the maximum value of the total quality coefficient $\overline{eq}$ based on the number of dimensions as noted in Sections 3 and 5 (6.93 for the original four dimensions, 8.66 for all five dimensions). The colors inside each bar show how each dimension's deficiency value d contributes to the overall quality. The first four colors in each top bar are also present in the corresponding bottom bar in a compressed form. This is because the addition of the feasibility dimension does not change the absolute d values of the original four dimensions. All it does is influence $\overline{eq}$ and its maximum possible value. When the feasibility dimension is added, the original dimension's retain their absolute d values but will relatively make up less of $\overline{eq}$, which is why their colors on the bottom bar take up less space. An effect of this is that the feasibility dimension can have a positive or negative effect on the relative $\overline{eq}$ value.

The feasibility dimension has the strongest effect on Bank USB CR NN. This is because it is the only authentication method that does not get the maximum penalty for work flow expansion while still scoring averagely for the circumvention and clarity aspects. For circumvention, the system is in a secure state by default and it is not possible for the user to circumvent security in any way. As for clarity, the only redeeming quality the card reader has is that its limited interface does not give the user a false impression of its functionality.

Weigold ZTIC VNA also shows a significant better quality due to the addition of the feasibility dimension. Although it shares the maximum work flow expansion with most other authentication methods because new user actions are introduced, it is quite favorably for the circumvention and clarity aspects. The system's default state is secure and unlike Bank USB CR NN, the user interface does support secure user behavior. The only negative modifier that applies to the circumvention aspect is that the user can subvert security due to inconvenience. As for clarity, the only penalty Weigold ZTIC VNA gets is that at least one in- and output channel (the communication channel between browser and bank) is corruptible.

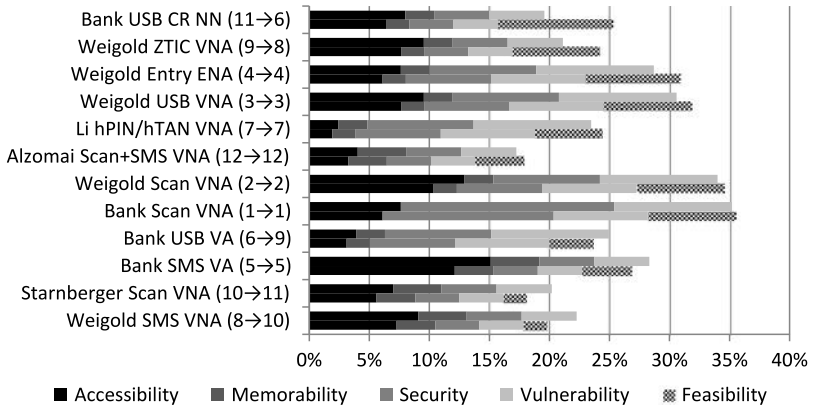

Fig. 9. Relative total quality coefficient values of the evaluated authentication methods of the original four dimensions, and the influence of the feasibility dimension. The list is sorted by the amount of influence the feasibility dimension has (from positive, to neutral, to negative). The maximum relative value of 100% represents the best fit in the original four dimensions of Renaud's mechanism (when the feasibility dimension is ignored) or all five dimensions (when the feasibility dimension is included).

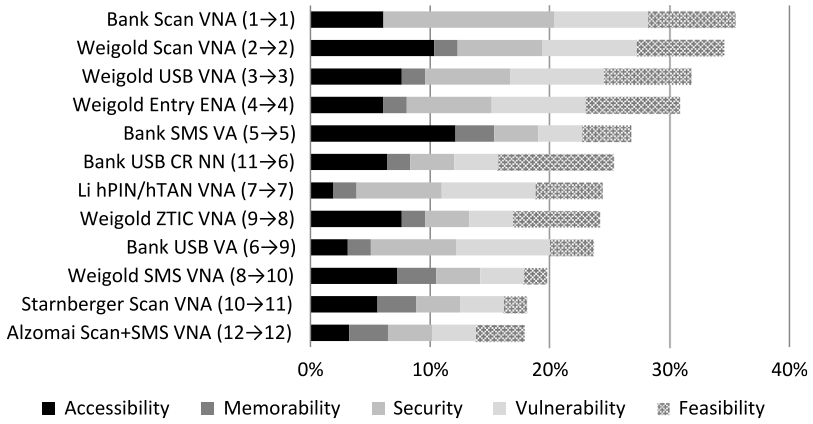

Fig. 10. Relative total quality coefficient values of the evaluated authentication methods in all five dimensions. The list is sorted by the relative total quality coefficient. The maximum value of 100% represents the best fit in all dimensions.

Weigold Entry ENA, Weigold USB VNA, Li hPIN/hTAN VNA, AlZomai Scan+SMS VNA, Weigold Scan VNA and Bank Scan VNA receive a minor increase in overall percentage but are barely affected due to the feasibility dimension's mediocre deficiency value.

The final four evaluated authentication methods are negatively affected by the feasibility dimension. Most notable is Weigold SMS VNA, which is fully penalized for both the work flow expansion and circumvention aspects. The only redeeming quality it has in this dimension is for the clarity aspect, for which the raters have stated that all information necessary to make a good decision is available during transaction authentication.

The added dimension changed 5 out of 12 ranks of the transaction authentication methods. The method with the highest overall percentage represents the best fit within the context of all dimensions, and therefore has the highest rank. With the original four dimensions, Bank USB CR NN had quite a bad overall quality, mostly due to its poor fit in the memorability dimension and also fitting quite poorly in the other dimensions. The feasibility dimension brings some of its commendable characteristics to the surface, boosting its overall percentage and giving it a six rank increase. Weigold ZTIC VNA received a smaller increase, but enough to make it rise one rank. As for rank decreases, Weigold SMS VNA loses a rank due to the earlier discussed bad fit in the feasibility dimension. The same is true for Starnberger Scan VNA, which drops a rank because it has the same poor fit. Bank USB VA drops three ranks, but this has less to do with its mediocre fit in the feasibility dimension. Instead, it drops three ranks due to the good fit Bank USB CR NN, Weigold ZTIC VNA and Li hPIN/hTAN VNA have.

9.2. Overall evaluation

We note the fit of the evaluated authentication methods within the dimensions of Renaud's mechanism and our added dimension. These results can only be used to rate the authentication methods on the criteria of the applied evaluation mechanism.

An overview based on the fulfillment of each dimension and based on the overall percentage is given in Fig. 10. The data represented in this Figure is also depicted by the bottom bars in Fig. 9, but Fig. 10 sorts the authentication methods by the relative total quality coefficient to make it easier to compare the qualitative fit of the methods with each other.

With the feasibility dimension included, Bank Scan VNA and Weigold Scan VNA have the highest overall percentage. They therefore have the best fit within the context of all five dimensions. This does not state that they have the best fit in each dimension. For example, while Bank Scan VNA has an exceptionally good fit in the security dimension, it has the worst fit in the memorability dimension. The latter is because the used authentication device relies on a bank card with a PIN code which cannot be changed (a bank-specific policy), which gives the most negative value to the memorability dimension's meaningless aspect. The other implemented methods allow the user to change PINs or passwords, and we assumed that this was also true for the proposed methods.

The evaluated methods are grouped by implementations and proposals for further comparisons. The proposal group has been split into two groups to compare the five proposals from Weigold and to improve the readability of the graphs. Radar charts are used to provide an overview of the different dimensions' fits. The center

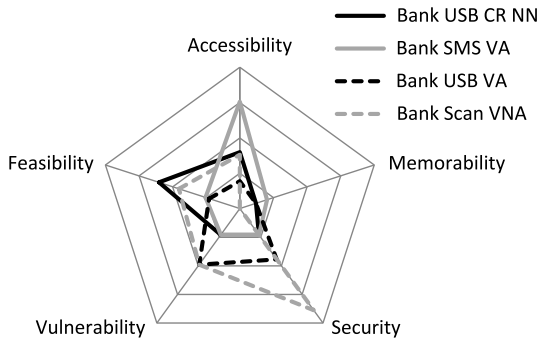

Fig. 11. Dimension fulfillment for implemented bank methods.

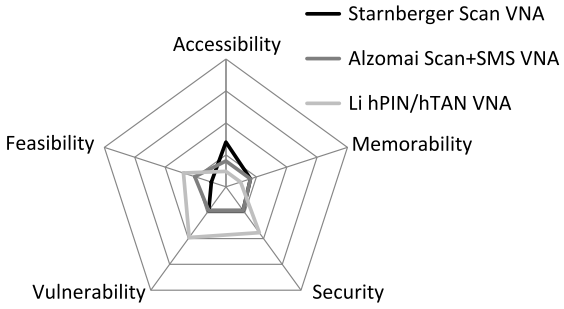

Fig. 12. Dimension fulfillment for various proposals.

of each radar chart represents 0%, while each line from inside to outside represents an additional 20%, making the total range 0% to 80%. Note that we do not rely on absolute numbers, but instead use the relative weights of the qualitative aspects to observe the fit within the dimensions of different authentication methods.

As shown in Fig. 11, it is quite easy to see where methods score favorably. Bank Scan VNA has a good fit in the security dimension because all secret key material used by this method (including the user's PIN) is distributed randomly and cannot be chosen by the user. This is different from Bank USB CR NN, Bank SMS VA and Bank USB VA, which do allow users to change their secret knowledge. The good fit of Bank SMS VA in the accessibility dimension can be explained that it does not require software or technical expertise to use, authentication does not take a lot of time and users with mobility or sensory disabilities are not excluded from using the authentication method.

The proposals made by Starnberger et al., AlZomai et al. and Li et al. do not fit particularly well, as depicted in Fig. 12. The line of Starnberger Scan VNA is hidden by the line of AlZomai Scan+SMS VNA in the memorability, security and vulnerability dimensions, which can be explained by that both methods are quite similar. Their differences are in the accessibility dimension (where it is thought that for Starnberger Scan VNA no technical expertise is required) and in the feasibility dimension (where it is thought that with AlZomai Scan+SMS VNA the user is unable to subvert security due to inconvenience). Li hPIN/hTAN VNA has a slightly better fit in the security and vulnerability dimensions since the required PIN to login to the bank's site is only entered in the user's computer with substitution digits, which the user received in a secure manner. As for the feasibility dimension, Li hPIN/hTAN VNA has a better fit compared to the others because it does not rely on a smartphone as an authentication device, and therefore it is less susceptible to remote intrusion by an adversary.

As shown in Fig. 13, all methods proposed by Weigold et al. fit quite poorly in the memorability dimension. However, the earlier discussed authentication methods have this poor fit as well. Syntactical passwords tax the memorability dimension heavily.

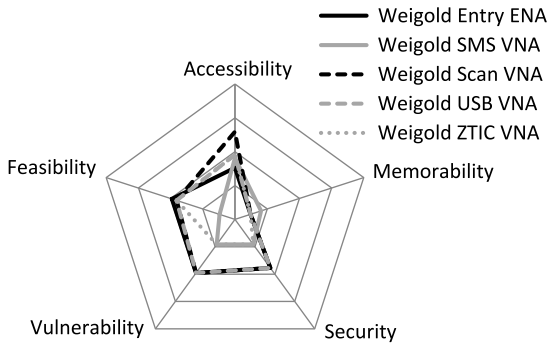

Fig. 13. Dimension fulfillment for Weigold's proposals.

Weigold Scan VNA does neither require technical expertise to use, nor does the user require a lot of time to use it. This makes it have the best fit in the accessibility dimension. Weigold SMS VNA fits the worst in the feasibility dimension, because it requires a lot of effort from the user to use (new actions are introduced in the user's work flow), it can be circumvented in every way that is evaluated by the mechanism and it relies on a smartphone, which is seen as an unreliable authentication platform.

9.3. Information scheme influence

We mentioned at the start of Section 7 three information schemes for transaction authentication methods. Bank USB CR NN uses TTA, Weigold Entry ENA uses ESTA and all other implemented methods and proposals apply CVTSA. Before we started our evaluation, we expected that TTA would rank low since it does not offer the user the option to securely verify transactions (like CVTSA) or the requirement to enter transactions in a secure device for automated verification (like ESTA), and therefore does not offer protection against malware attacks which change transaction information. We also expected that CVTSA would rank lower compared to ESTA, since a user can with the former (intentionally or not) perform the verification process incorrectly or skip it entirely.

Our first expectation was incorrect. Bank USB CR NN settles on a still admirable sixth position. It fits decently in the accessibility dimension due to that users do not need a lot of time to use it. The authentication method also has the highest fit in the feasibility dimension, which it mostly owes to that users are only required to perform redundant actions and not new actions during transaction authentication.

Our second expectation is also incorrect. Weigold Entry ENA (ESTA) has a high position, but is surpassed by Bank Scan VNA, Weigold Scan VNA and Weigold USB VNA (all CVTSA). Although Weigold Entry ENA scores high in the feasibility dimension due to that it cannot be circumvented in any way that the evaluation mechanism considers, it does not score exceptionally high in the other dimensions.

The evaluation we performed does not rule out any information scheme, which suggests that the evaluation mechanism can be used to compare authentication methods with different underlying schemes.

9.4. Variation between the raters

As we noted in Section 8.4, we did not perform the evaluation alone. For all aspects in the feasibility dimension and the most subjective characteristics of Renaud's original mechanism we used the average answer of seven raters.

It is unlikely that seven raters would always have the same opinion, especially since we expect that the aspects of Renaud's

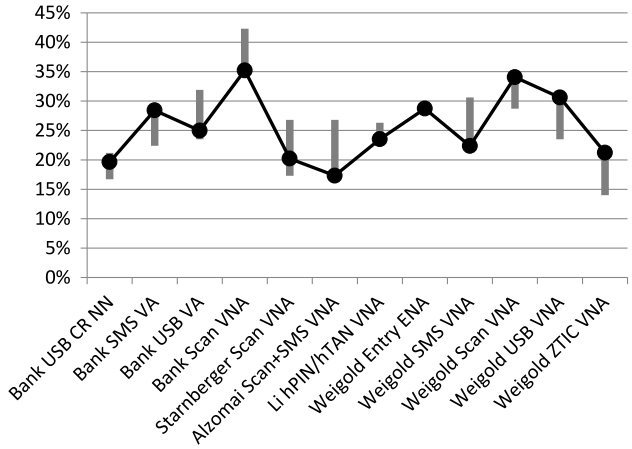

Fig. 14. Inter-rater variation of total quality coefficients with the dimensions of Renaud's mechanism. The black dots represent the total quality coefficients which comes forth from the average of the raters' answers (as noted in Table A.6), while the top and bottom of each gray bar represent respectively the highest and lowest total quality coefficients coming forth from the individual sets of answers.

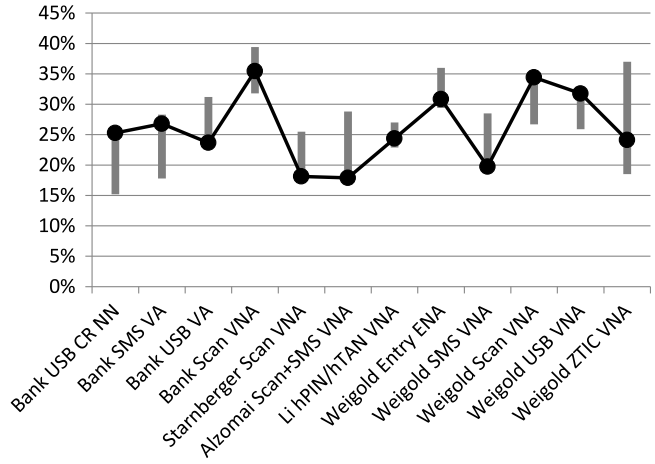

Fig. 15. Inter-rater variation of total quality coefficients with the dimensions of Renaud's mechanism and the proposed feasibility dimension.

mechanism and possibly our additional dimension are sensitive to subjectivity (as earlier discussed in Section 8.4). Indeed, only 32.6% of the questions were answered unanimously. For 64.4% of the questions an answer was provided by the majority. No majority was reached for the remaining 3.0% since for each of these questions three raters answered yes, three raters answered no and the final rater did not know the answer. One of seven raters did not know the answer for 8.33% of all questions.

Fig. 14 gives an overview of the total quality coefficient of each authentication method based on the four dimensions of Renaud's mechanism. It also shows the amount of variation there is between the sets of answers provided by the raters, giving an indication of how certain the raters are of the total fit within the mechanism.

Weigold Entry ENA has the similar total quality coefficient between all sets of answers, followed by Li hPIN/hTAN VNA and Bank USB CR NN. All other individual sets have more variation that either is more positive or negative compared to the average answer set. The one which stands out the most is Bank Scan VNA, of which even its minimal value (corresponding with the opinion of most raters) is higher compared to the others. At the opposite spectrum of the average answers is AlZomai Scan+SMS VNA, which fit the lowest. Still, there were some answer sets which could provide a higher outcome if the evaluation would solely rely on them.

The amount of variation can be explained by that the two tested aspect modifiers in the accessibility dimension (the need for technical expertise for the special requirements aspect, and a long authentication time for the convenience aspect) are quite ambiguous.

Fig. 15 shows the same graph for all five dimensions. The earlier discussed increase of the total quality coefficient for Bank USB CR NN is clearly visible, but it also shows that individual raters do not always come to such a final value.

Bank SMS VA, Bank USB VA, Li hPIN/hTAN VNA, Weigold Scan VNA and Weigold USB VNA stay more or less the same, both in the combined answer set and in variation of the individual answer sets. A significant increase in the amount of variation of Weigold ZTIC VNA can be seen, while the average value only climbs marginally. This can be explained by that two raters had the exact opposite values for the accessibility dimension. In addition, the rater with the more favorable values also gave the most favorable values to all aspects in the feasibility dimension, while the other rater was more critical.

As noted in Section 8.4, we also performed the role of a single rater. We expected that implemented and proposed authentication methods which used user-owned mobile devices would score lower compared to those which use a more secure environment offered by a bank-provided device. This is true for the proposed AlZomai Scan+SMS VNA, Starnberger Scan VNA and Weigold SMS VNA, but not for the implemented Bank SMS VA. The only effective difference between Bank SMS VA and Weigold SMS VNA is that the former uses text messages to make the user verify aggregated transaction data, whereas the latter proposes to do the same using non-aggregated transaction data. Therefore, it would be expected that Bank SMS VA would rank at least worse compared to Weigold SMS VNA, and not that it would rank so high as it does now. We asked one rater about this difference, who told us that he actually uses Bank SMS VA in daily life. The rater thinks that this method does not to take much of his time during authentication, explaining that he skips the verification check over the provided transaction data. He believes that the verification is not worth his time due to a low security risk perception.

The difference between the proposed and implemented text message-based methods combined with what the rater told us implies that there is a certain bias for methods which raters are familiar with. This presents a subjectivity that comes forth from the perspective in which raters answer questions. Combined with the earlier discussed ambiguous terms, it is suggested that the subjectivity came from two sources:

- Inherent subjectivity. This comes from the evaluation mechanism itself and the questions we asked based on it.
- Hidden subjectivity, which comes from the raters personal experience with the matter.

The former can possibly be reduced by providing more clear information. For the evaluation, we offered the original proposals for applicable authentication methods, which might have been too much information. The latter might be reduced by more explicitly asking questions. For example, we asked whether users would require much time to use the method. Instead, we could have asked whether the average user or most users would require much time to use the method. Different raters would still have different ideas about what an average or common user would be, but they might be more inclined to not think about the question from their own perspective.

As noted in Section 6, the order in which raters were required to rate the authentication methods was the same. This can be considered a bad practice due to that it can induce order bias. For example, a rater might be more positive about the first authentication method since he or she starts with a fresh look. By the time the rater is rating the last few authentication methods,

tiredness might make them rate more negatively. We tried to reduce the impact of any such bias this by telling the raters that they can pause and resume the survey at their leisure. The order in which the methods were rated is the same as shown in Figs. 14 and 15 (from left to right). As the Figures show, there does not seem to be any order bias. Methods were not rated overly negative or positive at the beginning or the end of the survey. It can be reasoned that the large variance in answers given for Weigold ZTIC VNA (the last rated authentication method) exists due to that this method differs quite a bit from the other methods. For example, this method is the only one using a device directly positioned in the communication flow between browser and bank. This increase in complexity might have confused the raters.

To conclude, Figs. 14 and 15 do indicate that it is useful to have multiple raters perform evaluations, either with Renaud's mechanism solely or with our expansion. There is quite a bit of subjectivity involved when answering the questions that need to be asked to use the mechanism, which is why it is unlikely that a single rater can state what is true and what is not.

10. Limitations, discussion and further research

The first step we performed was the examination of different evaluation mechanisms and frameworks. As noted in Section 3, we also examined Bonneau's framework [5]. For the scope of our work, we considered this framework wide in aspects and dimensions, but too shallow in its output since it does not quantify qualitative characteristics, nor does it provide indications of quality on multiple levels. For the work of others, Bonneau's framework could be modified to provide both quantified results on multiple levels while being less ambiguous on the values assigned to aspects. Considering the number of aspects this would mean a lot of work, but it has the potential to provide more detailed comparisons between authentication methods compared to our extension of Renaud. This partly comes from the inclusion of a deployability dimension, which also includes the potential cost to implement and maintain such a system.

Seven experts evaluated the subjective parts of Renaud's mechanism. We chose these parts since it would save the raters time which could be spend on rating our dimension instead. This presents a potential limitation in the multi-user evaluation since the choice of what is objective and what is subjective in Renaud's framework might be a subjective choice by itself. What we consider objective might be considered as subjective by others, in which case the relevant aspects should also be given a multi-user evaluation.

One of the raters told us that he uses one of the evaluated methods himself, and whenever he uses it he skips some of the security instructions. He justifies this by perceiving the risk of a security incident being low. This complies with Herley's vision [7,8]: the user's time is valuable and it will not be spend it on security if not strictly required and if a perceived need to do so is present. It would be useful to examine how users objectively work with and subjectively experience different authentication methods for online banking over a longer period of time, and to find out what motivates them to use the system in a more secure manner. This could be tested in a simulated online banking environment which needs to be flexible enough to support existing and new authentication methods.

We considered the differences in Renaud's mechanism when a new dimension is added. Further research can examine how the mechanism and dimensions can be changed or enhanced to take more aspects into account. Instead of modifying Bonneau's framework, a deployability dimension could also be considered as an extension for Renaud's and our work.

Table 5
Linked aspects.

Aspect	Source	
	Random	User
Meaningfulness	1	0–0.67
Predictability	0	0.5–1
Effective range	6.2%–96.5%	

We discovered that the memorability dimension's meaningfulness aspect and the security dimension's predictability aspect of Renaud's mechanism are linked with each other whenever a knowledge factor is present. The possible values depend on whether the knowledge is randomly or user chosen. This limitation of the model reduces the effective total coefficient range for methods which rely on knowledge. Table 5 shows the linked values and the effective relative range of the total quality coefficient. Further research could redefine the aspects in such a way that this and similar constraints are reduced or entirely removed from the mechanism.

11. Concluding remarks

We expanded Renaud's quantifying mechanism to accommodate aspects related to transaction authentication in online banking in a user-centric context. Several used and proposed transaction authentication methods for online banking were evaluated using the original four dimensions and our expansion by seven raters. The inclusion of an additional dimension changed the ranks of 5 out of the 12 evaluated authentication methods.

There is a large amount of subjectivity involved when applying Renaud's mechanism and our expansion. For almost a third of the asked questions did the (independent) raters come to an unanimous answer. This does not make the mechanism worthless, but it is advised that evaluations are performed by multiple raters, since it would be unwise to consider the opinion of a single expert as the truth.

The methods which have a good overall fit in both the original and the expanded mechanism include Bank Scan VNA, Weigold Scan VNA and Weigold USB VNA, closely followed by Weigold Entry ENA. The first three concern one implemented and three proposed authentication methods which use a Customer Verified Transaction Set Authentication information scheme, while the fourth uses Entered Single Transaction Authentication. This suggests that either information scheme can be applied to design an authentication method which can satisfy many aspects.

Trusted bank devices have a very good overall fit within the dimensions of the mechanism. User-owned mobile devices have a worse fit for online banking authentication purposes, except for the implemented Bank SMS VA. That this authentication method ranks so high is possibly due to personal bias among the raters who actually use this method in daily life, considering that the proposed Weigold SMS VNA is mostly the same but ranks much lower. When this outlier is ignored, it can be said that authentication methods which rely on user-owned devices tend to have an overall worse fit compared to those which rely on bank-issued devices.

Acknowledgments

We thank the raters. Your assistance gave us valuable insight in the usefulness of applying the evaluation mechanism multiple times by different people.

Also, we thank the anonymous reviewers who were asked by Elsevier's Future Generation Computer Systems to review our article. Your extensive, detailed and useful input was most valuable and appreciated.

Table A.6
Our research data. Each authentication method was first quantified using the original four dimensions of Renaud's mechanism. The results for the original four dimensions are noted on the rows for dimensions '1–4', and the results for the original four dimensions plus the feasibility dimension on the rows for dimensions '1–5'. Total quality coefficient is the resulting value of the mechanism and represents the fit of an authentication method within the specified dimensions. A higher value implies a better fit. Its maximum value can be calculated by $\max(\overline{eq}) = n * \max(d)$, where n is the number of dimensions and $\max(d)$ is the maximum value for a dimension's deficiency ($\max(d) = 1.73$). Overall percentages (given in **bold**) can be used to at a glance compare total quality coefficients for the original four dimensions with the same values for five dimensions.

Dimension(s)	Description	Formula	Authentication method											
			Bank USB CR	**Bank SMS VA**	**Bank USB VA**	**Bank Scan VNA**	**Starnberger Scan VNA**	**AlZomai Scan+ SMS VNA**	**Li hPIN /hTAN VNA**	**Weigold**				
										Entry ENA	**SMS VNA**	**Scan VNA**	**USB VNA**	**ZTIC VNA**
	Special req.	x	0.83	0.33	0.83	0	0.67	1	1	0	0.33	0	0.67	0.67
	Convenience	y	0.50	0.50	1	1	1	1	1	1	1	0.50	0.50	0.50
Accessibility	Inclusivity	z	0.67	0.33	0.67	0.67	0.33	0.33	0.67	0.67	0.33	0.67	0.67	0.67
	Deficiency	$ad = \sqrt{x^2 + y^2 + z^2}$	1.18	0.68	1.46	1.20	1.25	1.45	1.56	1.20	1.10	0.84	1.07	1.07
	Percentage	$adp = (1 - \frac{ad}{\max(d)}) * 100\%$	32%	61%	16%	31%	28%	16%	10%	31%	36%	52%	38%	38%
	Retrieval str.	x	1	1	1	1	1	1	1	1	1	1	1	1
	Meaningfulness	y	0.67	0.33	0.67	1	0.33	0.33	0.67	0.67	0.33	0.67	0.67	0.67
Memorability	Depth of proc.	z	1	1	1	1	1	1	1	1	1	1	1	1
	Deficiency	$md = \sqrt{x^2 + y^2 + z^2}$	1.56	1.45	1.56	1.73	1.45	1.45	1.56	1.56	1.45	1.56	1.56	1.56
	Percentage	$mdp = (1 - \frac{md}{\max(d)}) * 100\%$	10%	16%	10%	0%	16%	16%	10%	10%	16%	10%	10%	10%
	Predictability	x	1	1	1	0	1	1	1	1	1	1	1	1
	Abundance	y	0	0	0	0	0	0	0	0	0	0	0	0
Security	Disclosure	z	1	1	0.5	0.5	1	1	0.5	0.5	1	0.5	0.5	1
	Deficiency	$sd = \sqrt{x^2 + y^2 + z^2}$	1.41	1.41	1.12	0.50	1.41	1.41	1.12	1.12	1.41	1.12	1.12	1.41
	Percentage	$sdp = (1 - \frac{sd}{\max(d)}) * 100\%$	18%	18%	35%	71%	18%	18%	35%	35%	18%	35%	35%	18%
	Confidentiality	x	1	1	1	1	1	1	1	1	1	1	1	1
	Privacy	y	0	0	0	0	0	0	0	0	0	0	0	0
Vulnerability	Breakability	z	1	1	0.33	0.33	1	1	0.33	0.33	1	0.33	0.33	1
	Deficiency	$vd = \sqrt{x^2 + y^2 + z^2}$	1.41	1.41	1.05	1.05	1.41	1.41	1.05	1.05	1.41	1.05	1.05	1.41
	Percentage	$vdp = (1 - \frac{vd}{\max(d)}) * 100\%$	18%	18%	39%	39%	18%	18%	39%	39%	18%	39%	39%	18%
1–4	Total quality coeff.	$\overline{eq} = \max(\overline{d}) - \overline{d}$	1.36	1.97	1.73	2.44	1.40	1.20	1.63	1.99	1.55	2.36	2.12	1.47
	Overall percentage	$\overline{dp} = \frac{\overline{eq}}{\max(\overline{eq})} * 100\%$	**19.6**%	**28.4**%	**25.0**%	**35.2**%	**20.2**%	**17.3**%	**23.5**%	**28.7**%	**22.4**%	**34.1**%	**30.6**%	**21.2**%
	Ranking		11	5	6	1	10	12	7	4	8	2	3	9
	Work flow expansion	x	0.5	1	1	1	1	1	1	1	1	1	1	1
Feasibility	Circumvention	y	0.33	0.67	0	0.33	1	0.67	0.67	0	1	0.33	0.33	0.33
	Clarity	z	0.67	0.67	1	0.33	0.67	0.67	0.33	0.33	0.67	0.33	0.33	0.33
	Deficiency	$ed = \sqrt{x^2 + y^2 + z^2}$	0.90	1.38	1.41	1.10	1.56	1.38	1.25	1.05	1.56	1.10	1.10	1.10
	Percentage	$edp = (1 - \frac{ed}{\max(d)}) * 100\%$	48%	20%	18%	36%	10%	20%	28%	39%	10%	36%	36%	36%

(continued on next page)

Table A.6 (*continued*)

Dimension(s)	Description	Formula	Authentication method											
			Bank USB CR	**Bank SMS VA**	**Bank USB VA**	**Bank Scan VNA**	**Starnberger Scan VNA**	**AlZomai Scan+ SMS VNA**	**Li hPIN /hTAN VNA**	**Weigold**				
										Entry ENA	**SMS VNA**	**Scan VNA**	**USB VNA**	**ZTIC VNA**
1–5	Total quality coeff.	$\overline{eq} = \max(\overline{d}) - \overline{d}$	2.19	2.32	2.05	3.07	1.57	1.55	2.11	2.67	1.71	2.98	2.75	2.09
	Overall percentage	$\overline{dp} = \frac{\overline{eq}}{\max(\overline{eq})} * 100\%$	**25.3**%	**26.8**%	**23.7**%	**35.5**%	**18.1**%	**17.9**%	**24.4**%	**30.8**%	**19.7**%	**34.4**%	**31.8**%	**24.1**%
	Ranking		6	5	9	1	11	12	7	4	10	2	3	8
	Relative difference between dimensions 1–4 and 1–5		5.7%	−1.6%	−1.3%	0.2%	−2.1%	0.6%	0.8%	2.1%	−2.6%	0.4%	1.2%	2.9%
	Ranking difference		5	0	−3	0	−1	0	0	0	−2	0	0	1

This article is a product of the Dutch Research Program on Safety and Security of Online Banking. The research program is funded by the Dutch banking sector (represented by the Dutch Banking Association), the Police Academy, and the Dutch National Police.

Appendix. Evaluated authentication methods

See Table A.6.

References

[1] S. Kiljan, K. Simoens, D. De Cock, M. van Eekelen, H. Vranken, Security of online banking systems. Tech. Rep. TR-OU-INF-2014-01, (Open Universiteit), 2014, URL http://portal.ou.nl/documents/114964/523334/TR-OU-INF-2014-01.pdf.
[2] K. Renaud, Quantifying the quality of web authentication mechanisms: A usability perspective, J. Web Eng. 3 (2) (2004) 95–123. URL http://dl.acm.org/citation.cfm?id=2011143.2011146.
[3] M. Mihajlov, B. Blazic, S. Josimovski, Quantifying usability and security in authentication, in: 2011 IEEE 35th Annual Computer Software and Applications Conference, COMPSAC, July 2011, pp. 626–629.
[4] M. Mihajlov, B. Jerman-Blazic, S. Josimovski, A conceptual framework for evaluating usable security in authentication mechanisms—usability perspectives, in: 2011 5th International Conference on Network and System Security, NSS, Sept 2011, pp. 332–336.
[5] J. Bonneau, C. Herley, P.C. Van Oorschot, F. Stajano, The quest to replace passwords: A framework for comparative evaluation of web authentication schemes, in: 2012 IEEE Symposium on Security and Privacy, (SP), IEEE, 2012, pp. 553–567.
[6] K.-P. Yee, User interaction design for secure systems, in: R. Deng, F. Bao, J. Zhou, S. Qing (Eds.), Information and Communications Security, in: Lecture Notes in Computer Science, vol. 2513, Springer Berlin Heidelberg, 2002, pp. 278–290. URL http://dx.doi.org/10.1007/3-540-36159-6_24.
[7] C. Herley, So long, and no thanks for the externalities: The rational rejection of security advice by users, in: Proceedings of the 2009 Workshop on New Security Paradigms Workshop, NSPW'09, ACM, New York, NY, USA, 2009, pp. 133–144. URL http://doi.acm.org/10.1145/1719030.1719050.
[8] C. Herley, More is not the answer, IEEE Secur. Privacy 12 (1) (2014) 14–19.
[9] G. Starnberger, L. Froihofer, K. Goeschka, QR-TAN: Secure mobile transaction authentication, in: International Conference on Availability, Reliability and Security, 2009, ARES'09, March 2009, pp. 578–583.
[10] M. AlZomai, B. AlFayyadh, Jøsang, Display security for online transactions: SMS-based authentication scheme, in: 2010 International Conference for Internet Technology and Secured Transactions, ICITST, Nov 2010, pp. 1–7.
[11] T. Weigold, A. Hiltgen, Secure confirmation of sensitive transaction data in modern Internet banking services, in: 2011 World Congress on Internet Security, WorldCIS, Feb 2011, pp. 125–132.
[12] S. Li, A.-R. Sadeghi, S. Heisrath, R. Schmitz, J. Ahmad, hPIN/hTAN: A lightweight and low-cost E-banking solution against untrusted computers, in: G. Danezis (Ed.), Financial Cryptography and Data Security, in: Lecture Notes in Computer Science, vol. 7035, Springer Berlin Heidelberg, 2012, pp. 235–249. URL http://dx.doi.org/10.1007/978-3-642-27576-0_19.
[13] C.A. Mertler, Designing scoring rubrics for your classroom, Pract. Assess. Res. & Eval. 7 (25) (2001) 1–10.
[14] A.D. Trice, A Handbook of Classroom Assessment, Longman, 2000.
[15] M. Zurko, User-centered security: stepping up to the grand challenge, in: 21st Annual Computer Security Applications Conference, Dec 2005, pp. 14–202.
[16] W. Enck, P. Traynor, P. McDaniel, T. La Porta, Exploiting open functionality in SMS-capable cellular networks, in: Proceedings of the 12th ACM Conference on Computer and Communications Security, CCS'05, ACM, New York, NY, USA, 2005, pp. 393–404. URL http://doi.acm.org/10.1145/1102120.1102171.
[17] A.P. Felt, M. Finifter, E. Chin, S. Hanna, D. Wagner, A survey of mobile malware in the wild, in: Proceedings of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices, SPSM'11, ACM, New York, NY, USA, 2011, pp. 3–14. URL http://doi.acm.org/10.1145/2046614.2046618.
[18] S. Kiljan, H. Vranken, M. Van Eekelen, What you enter is what you sign: Input integrity in an online banking environment, in: 2014 Workshop on Socio-Technical Aspects in Security and Trust, STAST, July 2014, pp. 40–47.
[19] E. Poll, J. de Ruiter, The Radboud reader: A minimal trusted smartcard reader for securing online transactions, in: Policies and Research in Identity Management—Third IFIP WG 11.6 Working Conference, IDMAN 2013, London, UK, April 8–9, 2013. Proceedings. 2013, pp. 107–120. URL http://dx.doi.org/10.1007/978-3-642-37282-7_11.

Sven Kiljan has an interest in information security on the border of the theoretical and practical domains, and brings both together in his daily work. As a Ph.D. candidate, he participates in the Dutch Research Program on Safety and Security of Online Banking, funded by the Dutch banking sector (represented by the Dutch Banking Association), the Police Academy, and the Dutch National Police. In the substudy of technical security, he conducts his own research, and assists and receives assistance from the other substudies, which are related to cybercriminal organizations, customer–bank interaction and public–private partnership.

Harald Vranken is active in both research and education. His current projects include research in a distributed virtual computer security lab and the verification of security attributes. Some of the other topics of his past research include the applicability of cloud computing and server virtualization in existing organizations and processes.

Marko van Eekelen supervises the technical security substudy of the Kennisprogramma Veiligheid Digitaal Betalingsverkeer. Within the research program, he contributes his experience as a researcher in the field of security and as a coordinator of various educational programs. His most prominent field of research is security and correctness in modeling and programming.