Postprint Version of a Paper Accepted for IWPE '21
2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)

Privacy Considerations for Risk-Based Authentication Systems

Stephan Wiefling*, Jan Tolsdorf, and Luigi Lo Iacono
H-BRS University of Applied Sciences, Sankt Augustin, Germany
**Ruhr University Bochum, Bochum, Germany*
{*stephan.wiefling,jan.tolsdorf,luigi.lo_iacono*}*@h-brs.de*

***Abstract*—Risk-based authentication (RBA) extends authentication mechanisms to make them more robust against account takeover attacks, such as those using stolen passwords. RBA is recommended by NIST and NCSC to strengthen password-based authentication, and is already used by major online services. Also, users consider RBA to be more usable than two-factor authentication and just as secure. However, users currently obtain RBA's high security and usability benefits at the cost of exposing potentially sensitive personal data (e.g., IP address or browser information). This conflicts with user privacy and requires to consider user rights regarding the processing of personal data.**

We outline potential privacy challenges regarding different attacker models and propose improvements to balance privacy in RBA systems. To estimate the properties of the privacy-preserving RBA enhancements in practical environments, we evaluated a subset of them with long-term data from 780 users of a real-world online service. Our results show the potential to increase privacy in RBA solutions. However, it is limited to certain parameters that should guide RBA design to protect privacy. We outline research directions that need to be considered to achieve a widespread adoption of privacy preserving RBA with high user acceptance.

***Index Terms*—Password, Risk-based Authentication, Usable Security and Privacy, Big Data Analysis**

1. Introduction

Passwords are still predominant for authentication with online services [25], although new threats are constantly emerging. Credential stuffing and password spraying attacks [14] use leaked login credentials (username and password) sourced from data breaches, and try them in some way on (other) online services. These attacks are very popular today [2] since attackers can automate them with little effort. Major online services responded to this threat with implementing risk-based authentication (RBA) [36], aiming to strengthen password-based authentication with little impact on the user.

Risk-Based Authentication (RBA). RBA determines whether a login attempt is a legitimate one or an account takeover attempt. To do so, RBA monitors additional features when users submit their login credentials. Popular features range from network (e.g., IP address), device (e.g., smartphone model and operating system), or client (e.g., browser vendor and version), to (behavioral) biometric information (e.g., login time) [34], [36]. Based on the feature values and those of previous logins, RBA calculates a risk score. An access threshold typically classifies the score into low, medium, and high risk [12], [15], [21]. On a low risk (e.g., usual device and location), the RBA system grants access with no further intervention. On a medium or higher risk (e.g., unusual device and location), RBA requests additional information from the user, e.g., verifying the email address. After providing the correct proof, access is granted.

RBA is considered a scalable interim solution when passwords cannot simply be replaced by more secure authentication methods in many cases [34], [35]. The National Institute of Standards and Technology (NIST, USA) and National Cyber Security Centre (NCSC, UK) recommend RBA to mitigate attacks involving stolen passwords [13], [23]. Beyond that, users found RBA more usable than equivalent two-factor authentication (2FA) variants and comparably secure [35]. Also, in contrast to 2FA, RBA both offers good security and rarely requests additional authentication in practice [34], reducing the burden on users.

Research Questions. However, users obtain the security and usability gain of RBA at the cost of disclosing more potentially sensitive data with a personal reference, such as IP addresses and browser identifiers. Therefore, user privacy is at risk when RBA databases are forwarded or breached, as additional data besides usernames would potentially allow to identify individuals.

More and more data protection laws aim to protect users from massive data collection by online services. Considering that, we wondered whether and to what extent the integration of RBA systems complies with the principles of modern data protection. We also wondered which trade-offs are possible to balance security and privacy goals.

To further investigate RBA's privacy aspects, we formulated the following research questions:

RQ1: a) In what ways can RBA features be stored to increase the user privacy?
b) How can RBA features be stored to protect user privacy in terms of data breaches?
RQ2: To what extent can a RBA feature maintain good security while preserving privacy in practice?

Contributions. We propose and discuss five privacy enhancements that can be used by RBA models used by the

This research was supported by the research training group "Human Centered Systems Security" (NERD.NRW) sponsored by the state of North Rhine-Westphalia.

DOI: 10.1109/EuroSPW54576.2021.00040

majority of deployments found in practice. To estimate their usefulness in practice, we evaluated a subset of these enhancements on a RBA feature that is highly relevant in terms of security and privacy, i.e., the IP address. We evaluated with a data set containing the login history of 780 users on a real-world online service for over 1.8 years.

Our results show for the first time that it is possible to increase feature privacy while maintaining RBA's security and usability properties. However, increasing privacy is limited to certain conditions that need to be considered while designing the RBA system. We also identified future challenges and research directions that might arise with a widespread RBA adoption in the future.

The results support service owners to provide data protection compliant RBA solutions. They assist developers in designing RBA implementations with increased privacy. Researchers gain insights on how RBA can become more privacy friendly, and further research directions.

2. Background

In the following section, we provide a brief introduction to RBA and explain how the use of RBA correlates with the several privacy principles defined by industry standards and legislation.

2.1. RBA Model

Since RBA is not a standardized procedure, multiple solutions exist in practice. We focus on the implementation by Freeman et al. [12], since it performed best in a previous study [34]. Also, this RBA model is known to be widely used, e.g., by popular online services like Amazon, Google, and LinkedIn [34], [36].

The model calculates the risk score S for a user u and a set of feature values $(FV^1, ..., FV^d)$ with d features as:

$$S_u(FV) = \left(\prod_{k=1}^{d} \frac{p(FV^k)}{p(FV^k|u, legit)} \right) \frac{p(u|attack)}{p(u|legit)} \quad (1)$$

S has the probabilities $p(FV^k)$ that a feature value appears in the global login history of all users, and $p(FV^k|u, legit)$ that a legitimate user has this feature value in its own login history. The probability $p(u|attack)$ describes how likely the user is being attacked, and $p(u|legit)$ describes how likely the legitimate user is logging in.

2.2. Regulatory Foundations

In the past few years, the introduction of new data protection laws, such as the General Data Protection Regulation (GDPR) [8] and the California Consumer Privacy Act (CCPA) [30], dramatically changed the way online services (i.e., data controllers) process their users' data. Formerly loose recommendations on handling user data have been replaced by clear and binding data protection principles, which data controllers must adhere to. However, the details and scope of the principles vary between jurisdictions. For internationally operating data controllers, this poses the problem that their data processing operations must be designed to be compatible with different requirements. Fortunately, the privacy framework specified in ISO 29100:2011 [16] already compiles an intersection of privacy principles from data protection laws worldwide. Thus, it provides data controllers a solid basis for designing legally compliant data processing operations that can be tailored to the details of different jurisdictions. We outline the requirements for the design of RBA systems based on the privacy principles defined in ISO 29100:2011, aiming at compatibility with different jurisdictions.

Applicability of Privacy Principles. Generally speaking, the privacy principles defined in established privacy laws and frameworks aim to protect the privacy of individuals. Thus, they only apply to data with a personal reference. Such data are called, e.g., *"personal data"* (GDPR [8]), *"personal information"* (CCPA [30]), or *"personally identifiable information"* (PII) (ISO [16]). The definitions are very similar and usually refer to *"any information that (a) can be used to identify [an individual] to whom such information relates, or (b) is or might be directly or indirectly linked to [an individual]"* [16].

The data processed by RBA certainly fall within this definition, since implementations rely on features that already serve as (unique) identifiers by themselves (e.g., IP address) [36]. Also, the risk score calculated by RBA represents an identifier by itself, as it constitutes a set of characteristics that uniquely identifies an individual. Therefore, RBA has to comply with ISO 29100:2011's privacy principles discussed below.

Consent and Choice. In general, data controllers must ensure the lawfulness of data processing. While most jurisdictions recognize user consent as a lawful basis, applicable laws may allow processing without consent. Depending on the assets associated with a user account, data controllers may argue that RBA use is required to comply with the obligation to implement appropriate technical safeguards against unauthorized access. Nonetheless, to ensure compliance, providers should design RBA mechanisms with consent in mind and provide their users with clear and easy-to-understand explanations.

Collection Limitation and Data Minimization. Data controllers must limit the PII collection and processing to what is necessary for the specified purposes. RBA feature sets should therefore be reviewed for suitability with redundant or inappropriate features removed [34]. This includes considering using pseudonymized data for RBA and disposing of the feature values when they are no longer useful for the purpose of RBA. In practice, this creates the challenge to not reduce a risk score's reliability.

Use, Retention, and Disclosure Limitation. The data processing must be limited to purposes specified by the data controller, and data must not be disclosed to recipients other than specified. RBA should ensure that features cannot be used for purposes other than the calculation of risk scores. Moreover, after a feature value becomes outdated, it should be securely destroyed or anonymized. We would point out that privacy laws do not apply to anonymized data and could therefore serve data controllers for developing and testing purposes beyond the retention period specified in their privacy statements.

Accuracy and Quality. Data controllers must ensure that

the processed data are accurate and of quality. This is not only due to their own business interests, but also because data subjects have a right to expect their data being correct. This directly affects RBA, since it has the power to deny a significant benefit to users (i.e., access to their user account) with potentially significant harm. Data controllers must hence ensure by appropriate means that the stored feature values are correct and valid.

Individual Participation and Access. Data controllers must allow data subjects to access and review their PII. For RBA, this means that users should be allowed to be provided with a copy of the feature values used.

Information Security. Data controllers are obliged to protect PII with appropriate controls at the operational, functional, and strategic level against risks. These include, but are not limited to, risks associated with unauthorized access or processing and denial of service. Privacy laws demand extensive protections in this regard, *"taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons"* (Art. 32 (1) GDPR). Since RBA risk scores do not necessarily rely on evaluating plain text feature values [34], the collected data should be stored in an appropriate pseudonymized, masked, truncated, or encrypted form, depending on the RBA implementation. Moreover, data controllers should implement additional technical and organizational measures as needed, and be able to ensure the integrity, availability, and resilience of RBA.

Accountability and Privacy Compliance. Data controllers should inform data subjects about privacy-related policies, transfers of PII to other countries, and data breaches. Data controllers should also implement organizational measures to help them verify and demonstrate legal compliance. These include, but are not limited to, risk assessments and recovery procedures. RBA implementations should therefore consider the worth of RBA features to both attackers and data subjects, and the recovery from data breaches. This is crucial in order not to undermine the security of user accounts and their associated assets.

3. Privacy Enhancements (RQ1)

To comply with the privacy principles and derived data protection requirements, service owners should consider mechanisms to increase privacy in their RBA implementations. In the following, we introduce threats and their mitigation to increase privacy properties of RBA features.

3.1. Feature Sensitivity and Impact Level

RBA feature sets always intend to distinguish attackers from legitimate users. In doing so, the features may contain sensitive PII. However, not only do users perceive such PII differently regarding their sensitivity [28]. Their (unintended) disclosure could also have far-reaching negative consequences for user privacy. Developers and providers should therefore determine the impact from a loss of confidentiality of the RBA feature values. Specifically, the following aspects need consideration [20]:

Identifiability and Linkability. RBA feature sets should be evaluated regarding their ability to identify natural persons behind them. In particular, RBA systems that rely on intrusive online tracking methods, such as browser fingerprinting, store sensitive browser-specific information that form a linked identifier. In the event of losing confidentiality, the features would allow clear linkage between profiles at different online services, despite users using different login credentials or pseudonyms. Depending on the service, this could result in negative social or legal consequences for individuals. It could also enable more extensive and unintended activity tracking, and deanonymizing information associated with user accounts. Previous work found that powerful RBA feature sets do not require to uniquely identify users when focusing on the detection of account takeover attempts [34]. Also, users are more willing to accept the processing of sensitive information when they are certain that it is anonymous and does not allow them to be identified [18], [29]. Thus, the use of non-intrusive features may increase user trust in online services, too.

Feature Values Sensitivity. Aside from identifying individuals by RBA feature sets, the individual feature values may already contain sensitive PII. Sensitive PII in the scope of RBA may be feature values that are easily spoofable and can be misused to attack other online services in the event of a data breach. Sensitive PII may also refer to data perceived as sensitive by online users. For example, the most important feature of current RBA methods, namely the IP address [12], [15], [31], [34], [36], is perceived as highly sensitive by online users of diverse cultural backgrounds [3], [19], [28]. Since users are generally less willing to share data with increased sensitivity, RBA feature sets should limit the use of sensitive data if possible, in order to meet user interests.

3.2. Threats

RBA features may contain personal sensitive data, which has to be protected against attackers. To support online services in their protection efforts, we introduce three privacy threat types. We based the threats on those found in literature and our own observations in practice.

Data Misuse. Online services could misuse their own RBA feature data for unintended purposes, such as user tracking, profiling, or advertising [5]. This type of misuse previously happened with phone numbers stored for 2FA purposes [33]. While users have to trust online services to not misuse their data, responsible online services should also take precautions to minimize chances for miuse scenarios or unintended processing, e.g., by internal misconduct or after the company changed the ownership.

Data Forwarding. Online services can be requested or forced to hand out stored feature data, e.g., to state actors, advertising networks, or other third parties. Especially IP addresses are commonly requested [9]. When such data are forwarded to third parties, the users' privacy is breached. For instance, the IP address could be used to reveal the user's geolocation or even their identity.

Data Breach. Attackers obtained the database containing the feature values, e.g., by hacking the online service. As a

result, online services lost control over their data. Attackers can try to re-identify users based on the feature values, e.g., by combining them with other data sets. They can further try to reproduce the feature values and try account takeover attacks on a large scale, similar to credential stuffing. On success, they could access sensitive user data stored on the online service, e.g., private messages.

3.3. Mitigation

Online services can implement several measures to mitigate the outlined privacy threats. We propose five measures that are based on methods found in related research fields, as well as privacy regulations and our own observations with the selected RBA model (see Section 2). Based on the introduced RBA model, we considered all feature values as categorical data, in order to calculate the probabilities. When this condition is met, the proposed measures are also applicable to other RBA models [34].

As an example for practical solutions, we describe how the considerations can be applied to the *IP address* feature, with regard to the IPv4 address. We chose this feature since it is both considered the most important RBA feature in terms of security to date and sensitive re-linkable data in terms of privacy (see Section 3.1).

3.3.1. Aggregating. The RBA model only depends on feature value frequencies. To minimize data and limit misuse [16], we can aggregate or reorder feature data in the login history without affecting the results. The data set would then reveal how often a feature combination occurred, but not its chronological order. Removing this information can mitigate re-identification in login sequences.

3.3.2. Hashing. A cryptographic hash function, such as SHA-256, transforms a data input of arbitrary value to an output of fixed length. As inverting a hash function is not possible in theory, attackers need to recalculate all possible hashing values to restore the input values [17]. Assuming that the hashes are practically collision-free, using hashed feature values will produce the same RBA results as with the original values. This is the case, because the feature values are only transformed into a different representation. Therefore, this could be a solution to protect feature data in terms of the outlined threats.

However, the IPv4 address has 32 bit limited input values, where some addresses have a specific semantic and purpose, and cannot be assigned to devices. Thus, attackers can simply hash all $2^{32}-1$ values to restore the correct IP address. To counteract this problem, we can append a large random string (salt) to the input value:

$$H(192.168.1.166 \mid\mid salt) = 243916...aad132 \quad (2)$$

Attackers need to guess the salt correctly, which is high effort when the salt is large. Thus, this mitigation strategy increases the required guessing time for each feature value. Taking it a step further, we can even hash the results multiple times to increase the computation time:

$$H(H(...H(192.168.1.166 \mid\mid salt))) = [hash] \quad (3)$$

This is similar to key derivation strategies used in password databases [22]. However, we can only use a global salt for all database entries, as RBA mechanisms need to be able to identify identical feature values across users in the database. By increasing the computational cost, attackers cannot scale attacks as they would have with the unhashed feature values.

3.3.3. Truncation. A more destructive approach to increase privacy for RBA features is to change or remove details from their data values. This can reduce the number of records with unique quasi identifiers. Since the feature data then becomes less useful for other use cases like tracking or re-identification, we consider it a measure to mitigate the privacy threats. Regarding the IP address, we could set the last bits to zero. For truncating the last eight bits, for example, this would result in:

$$Truncate(192.168.1.166, 8\ Bit) = 192.168.1.0 \quad (4)$$

This mechanism is known from IP address anonymization strategies [6], [7]. However, we can also apply it on other features, e.g., reducing timing precision or coarse-graining browser version number in the user agent string [24]. Since we remove information that could potentially identify an individual, e.g., the device's internet connection, this can potentially increase privacy. However, this can also influence the RBA results, as there are fewer feature values for attackers to guess.

3.3.4. K-Anonymity. The k-anonymity privacy concept [32] ensures that at least k entries in a data set have the same quasi identifier values. If attackers obtained the data set and know a victim's IP address, they would not be able to distinguish the person from k other users. This makes it an effective countermeasure against re-identification in case of data forwarding and data breaches.

To achieve k-anonymity for RBA, at least k users need to have the same feature value. To ensure this, we added potentially missing entries to the RBA login history after each successful login. We added these entries to random users to only affect the global login history probabilities in order to keep a high security level. We created these users just for this purpose. To retain the global data set properties, the user count increased gradually to have the same mean number of login attempts per user.

3.3.5. Login History Minimization. Another approach is to limit the login history, in terms of the amount of features and entries, for a number of entries or a constant time period [16]. A study already showed that few entries are sufficient to achieve a high RBA protection [34]. In so doing, we mitigate tracking users for an extended period of time. However, this can affect the RBA performance based on the usage pattern of the corresponding online service. Especially when it is a less-than-monthly-use online service, we assume that features need to be stored for a longer period than for daily use websites to achieve a comparable RBA performance.

4. Case Study Evaluation (RQ2)

Aggregating and hashing, when collision-free, does not affect the RBA results, as they only change the data representation for the RBA model. The other approaches, however, potentially could. To assess their impact on

RBA behavior in practice, we studied truncation and k-anonymity using real-world login data. The properties and limited size of our data set did not allow to reliably test the login history minimization approach, so we left it for future work. Nevertheless, we outlined this relevant privacy consideration for the sake of completeness. We used the IP address feature as in the other examples.

4.1. Data Set

For the evaluation, we used our long-term RBA data set, including features of 780 users collected on a real-world online service [34]. The online service collected the users' features after each successful login. The users signed in 9555 times in total between August 2018 to June 2020. They mostly logged in daily (44.3%) or several times a week (39.2%), with a mean of 12.25 times in total. To improve data quality and validity, we removed all users who noticed an illegitimate login in their account. The online service was an e-learning website, which students used to exercise for study courses and exams. As the users were mostly located in the same city, it is a very challenging data set for RBA. They could get similar IP addresses with higher probability. Therefore, it is important to evaluate how the RBA protection changes in such a challenging scenario.

4.1.1. Legal and Ethical Considerations. The study participants [34] signed a consent form agreeing to the data collection and use for study purposes. They were always able to view a copy of their data and delete it on request. The collected data were stored on encrypted hard drives and only the researchers had access to it.

We do not have a formal IRB process at our university. Still, we made sure to minimize potential harm by complying with the ethics code of the German Sociological Association (DGS) and the standards of good scientific practice of the German Research Foundation (DFG). We also made sure to comply with the GDPR.

4.1.2. Limitations. Our results are limited to the data set and the users who participated in the study. They are limited to the population of a certain region of a certain country. They are not representative for large-scale online services, but show a typical use case scenario of a daily to weekly use website. As in similar studies, we can never fully exclude that intelligent attackers targeted the website. However, multiple countermeasures minimized the possibility that the website was infiltrated [34].

4.2. Attacker Models

We evaluated the privacy enhancements using three RBA attacker models found in related literature [12], [34]. All attackers possess the login credentials of the target.

Naive attackers try to log in from a random Internet Service Providers (ISP) from somewhere in the world. We simulated these attackers by using IP addresses sourced from real-world attacks on online services [11].

VPN attackers know the country of the victim. Therefore, we simulated these attackers with IP addresses from real-world attackers located in the victim's country [11].

Targeted attackers know the city, browser, and device of the victim. Therefore, they choose similar feature values, including similar ISPs. We simulated these attackers with our data set, with the unique feature combinations from all users except the victim. Since the IP addresses of our data set were in close proximity to each other, our simulated attacker was aware of these circumstances and chose them in a similar way.

4.3. Methodology

In order to test our privacy enhancements in terms of practical RBA solutions, we defined a set of desired properties. Our enhancements need to: (A) **Keep the percentage of blocked attackers**: The ability to block a high number of attackers should not decrease when using the privacy enhancements. This is necessary to keep the security properties of the RBA system. (B) **Retain differentiation between legitimate users and attackers**: When applied, the risk score differences between legitimate users and attackers should only change within a very small range. Otherwise, the usability and security properties of the RBA system would decrease.

We outline the tests to evaluate the privacy enhancements below. Based on the method in Wiefling et al. [34], we reproduced the login behavior for attackers and legitimate users by replaying the user sessions. We integrated truncation and k-anonymity in the reproduction process, to test the countermeasures.

The RBA model used the IP address and user agent string as features, since this can be considered the RBA state of practice [34], [36]. We truncated the IP addresses in ranges from 0 to 24 bits, to observe the effects on the RBA performance. We assume that cutting more than 25 bits will not allow to reliably detect attackers. We also tested k-anonymity with the IP address feature until $k = 6$. As US government agencies consider less than five entries to be sensitive [10], we chose to cover this threshold.

4.3.1. Test A: Percentage of Blocked Attackers. To compare the RBA performance regarding all three attacker models, we calculated the percentage of how many attackers would be blocked. We call this percentage the *true positive rate* (TPR), as previous work did [12], [34]. For a fair comparison, we observed how the TPR changed when aiming to block 99.5% of attackers. We chose this TPR baseline since it showed good performance regarding usability and security properties in a previous study [34].

To ease comparison, we adjusted the TPR for each truncation or k-anonymity step x_i as percentage differences to the baseline without modifications (relative TPR):

$$TPR_{relative_{x_i}} = \frac{TPR_{x_i} - TPR_{baseline}}{TPR_{baseline}} \tag{5}$$

Following that, $TPR_{relative_{x_i}} < 0.0$ means that the TPR decreased compared to the baseline.

4.3.2. Test B: Risk Score Changes. To determine the degree that attackers and legitimate users can be differentiated in the RBA model, we calculated the *risk score relation* (RSR) [34]. It is the relation between the mean risk scores for attackers and legitimate users:

$$RSR_{basic} = \frac{mean\ attacker\ risk\ score}{mean\ legitimate\ risk\ score} \tag{6}$$

To ease comparison, we normalized each RSR for every truncation or k-anonymity step x_i as percentage differences to the baseline (relative RSR). The baseline is the IP address without modifications:

$$RSR_{relative_{x_i}} = \frac{RSR_{basic_{x_i}} - RSR_{baseline}}{RSR_{baseline}} \quad (7)$$

As a result, $RSR_{relative_{x_i}} < 0.0$ signals that attackers and legitimate users can no longer be distinguished as good as they were before introducing the privacy enhancing measures.

4.3.3. Limit Extraction. For each test, we defined the following thresholds to extract limits that do not degrade RBA performance to an acceptable extent.

(Test A) We require the RBA performance to remain constant. Thus, we selected the reasonable limit as the point at which the relative TPR decreases compared to the baseline, i.e., attackers cannot be blocked as good as before any more. (Test B) Unlike tracking, RBA uses the feature information in addition to an already verified identifier, e.g., passwords. Thus, we consider it feasible to reduce the RSR slightly for the sake of privacy. Based on our observations, RSR changes below 0.01 can be tolerable for our case study evaluation. Thus, we chose the reasonable limit as the point at which the relative RSR is lower than 0.01.

4.4. Results

In the following, we present the results for all attacker models. We discuss the results after this section. We used a high performance computing cluster using more than 2000 cores for the evaluation. This was necessary since calculating the results with the simulated attackers was computationally intensive.

For statistical testing, we used Kruskal-Wallis tests for the omnibus cases and Dunn's multiple comparison test with Bonferroni correction for post-hoc analysis. We considered p-values less than 0.05 to be significant.

4.4.1. Truncation. Figure 1 shows the truncation test results for all attackers. The TPR differences between the targeted attacker and both remaining attackers were significant (Targeted/Naive: p=0.0151, Targeted/VPN: p<0.0001). The TPRs exceeded the limit after 20 bits for naive, 3 bits for VPN, and 14 bits for targeted attackers.

Regarding the relative RSRs, there are significant differences between VPN and both remaining attackers (p<0.0001). The RSRs exceeded the limit after 3 bits for naive, 21 bits for VPN, and 3 bits for targeted attackers.

Combining both results, the accepted truncation limits based on our criteria were 3 bits for all attacker models.

4.4.2. K-Anonymity. Figure 2 shows the combined k-anonymity test results for the three attacker models. The relative TPR decreased after $k = 1$ for targeted attackers, $k = 2$ for naive attackers, and not at all for VPN attackers until at least $k = 6$. There were significant TPR differences between naive and VPN attackers (p=0.0066).

The relative RSR did not decrease for all attacker types and there were no significant differences.

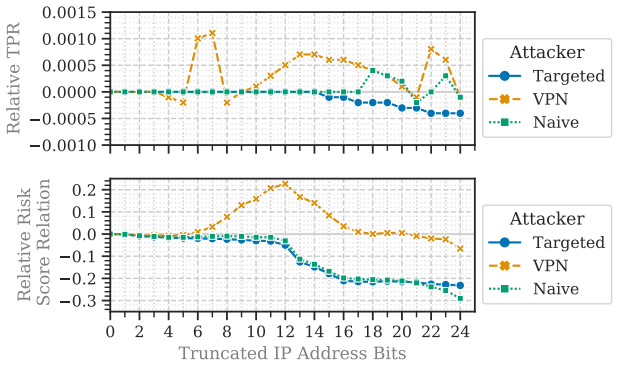

Figure 1. Results for truncating the IP address. Top: Relative TPR (Test A). There were significant differences between targeted and both VPN and naive attackers. Bottom: Relative RSR (Test B). The differences between VPN and both targeted and naive attackers were significant.

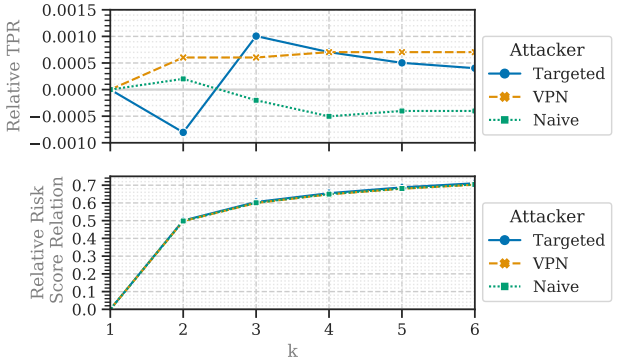

Figure 2. Results for k-anonymity regarding the IP address. Top: Relative TPR (Test A). Differences between naive and VPN attackers were significant. Bottom: Relative RSR (Test B). There were no significant differences.

Combining the results, the acceptable k levels based on our criteria were $k = 1$ for targeted attackers, $k = 2$ for naive attackers, and at least $k = 6$ for VPN attackers.

5. Discussion

Our results show that IP address truncation significantly affects the RBA risk score and reduces the probability of attack detection. The truncation for VPN attackers resulted in a local maximum of the RSR at 12 bits, and thus apparently improved detection. However, this was due to the fact that the VPN attacker only had an IP address range limited to the VPN service's server locations. Since the first IP address bits correspond to a node's geolocation, they were mostly distinct from legitimate users residing in different areas. Thus, truncating increased the risk scores for VPN attackers until 12 bit, as the probability for the global login history $p(FV^k)$ decreased but the one for the local history $p(FV^k|u, legit)$ remained constant. In contrast to that, targeted attackers also had a limited IP address range, but they were located in the same region as the legitimate users. Also, naive attackers had a large IP address range. Thus, in both cases, the differences between $p(FV^k)$ and $p(FV^k|u, legit)$ remained constant to similar levels until 12 bits.

Following that, and what our evaluation indicates, we

TABLE 1. OVERHEAD CREATED BY ADDITIONAL LOGIN ENTRIES TO ACHIEVE K-ANONYMITY

k	Additional Entries	Increase to Baseline
1	0	0.0
2	3928	0.41
3	7965	0.83
4	12013	1.26
5	16065	1.68
6	20120	2.11

do not recommend truncating more than three bits for a stable RBA performance in our case study scenario.

K-anonymity increased the distinguishability between legitimate users and attackers, i.e., the RSR. This was due to the fact that this mechanism added new entries to the global login history. As a result, the overall probability for unknown feature values in the global login history $p(FV^k)$ decreased, making it harder for attackers to achieve a low risk score. However, this also decreased the detection of attackers, i.e., the TPR, in most cases, since k more users had similar feature values in the data set. As a side effect of these results, unique feature values got less unique in total. Thus, due to the determined limit of $k = 1$, k-anonymity for targeted attackers can only be achieved with degraded RBA performance.

The overhead produced by the additional entries increased with each k (see Table 1). It was even more than the data set itself at k>3, which makes the current mechanism impractical for very large online services. To mitigate this issue, mechanisms could be introduced which remove some additional login entries when k-anonymity can be fulfilled after some time.

K-anonymity is not scalable with an increasing number of features [1], while the other approaches are. Thus, sensible RBA privacy enhancements might be a combination of all outlined countermeasures, to ensure scalability.

Based on our results, we discuss privacy challenges and further research directions in the following.

5.1. Privacy Challenges

When integrating privacy into RBA systems, there are several challenges that should be considered in practice. We describe them below.

Role of the IP Address Feature. Using a combination of privacy enhancements for the IP address might be sufficient for some applications. However, this feature is still sensitive information. Thus, the question arises whether online services should consider privacy enhancing alternatives instead of storing the IP address. One alternative could be to derive only the region and ASN from the IP address, and discard the rest. Other approaches even enable identifying network anomalies, e.g., IP spoofing using a VPN connection, without having to rely on the IP address at all. For example, the server-originated round-trip time (RTT) [34] can be used to estimate the distance between the user's device and the server location and may replace IP addresses as RBA features. As the RTTs vary based on the server location, they become useless for most re-identification attacks using leaked databases, as server locations are distributed in practice. They can even be enriched with random noise to further enhance privacy.

Risk of Feature Stuffing. Such considerations can be more and more important with widespread RBA adoption in the future. We assume that when databases with RBA feature values got stolen, this might have serious consequences for other services using RBA. In contrast to passwords, behavioral RBA feature values cannot be changed after compromise. Attackers can attempt to automatically reproduce these feature values on other websites. Thus, more privacy preserving alternatives that are hard to spoof for attackers might be crucial to mitigate largely scalable "feature stuffing" attacks.

Handling Data Deletion Requests. Further conflicts could arise with data protection regulations. Users are legally permitted to request data deletion. So when they request online services to delete their RBA feature data, they might lose RBA protection on their user accounts.

5.2. Research Directions

Our case study evaluation provided first insights on truncating feature values to increase privacy. As the results showed that this is possible to a certain degree while maintaining RBA performance, further work can investigate it for other types of features, e.g., the user agent string.

The proposed k-anonymity mechanism can increase privacy regarding unique entries in the data set. However, users might still be identifiable when they have a combination of typical feature values, e.g., a home and a work IP address. This non-trivial task had been addressed in dynamic databases [27], [37]. Future work may investigate whether such mechanisms are also applicable to RBA.

As we could not reliably test the login history minimization approach with our data set, future work should investigate this on a medium to large-scale online service with regular use.

6. Related Work

Burkhard et al. [6] investigated truncating IP addresses in anomaly detection systems. They found that truncating more than four bits degraded the performance of these systems. Chew et al. [7] further evaluated IP truncation in intrusion detection systems. Their results showed that the detection accuracy in many of the tested classifiers decreased after removing more than 8 bits. Our study showed that three bits could be removed from the IP address to maintain RBA performance at the same time.

Both Safa et al. [26], and Blanco-Justicia and Domingo-Ferrer [4] proposed privacy-preserving authentication models for implicit authentication using mobile devices. Their models relied on client-originated features, and the former also calculated risk scores on the client's device. However, this is not applicable to our RBA use case, as it relies on server-originated features and risk scores to prevent client-side spoofing.

To the best of our knowledge, there were no studies investigating privacy enhancements in RBA systems. However, some literature touched on privacy aspects related to RBA. Bonneau et al. [5] discussed privacy concerns of using additional features for authentication. They found that privacy preserving techniques might mitigate these concerns, but these had not been deployed in practice. We

proposed and tested some techniques for the first time in our case study. Wiefling et al. [35] investigated RBA's usability and security perceptions. The results showed that users tended to reject providing phone numbers to online services for privacy reasons. They further studied RBA characteristics on a real-world online service [34], showing that the feature set can be very small to achieve good RBA performance. We demonstrated that the privacy can be further enhanced through different mechanisms.

7. Conclusion

With a widespread use of RBA to protect users against attacks involving stolen credentials, more and more online services will potentially store sensitive feature data of their users, like IP addresses and browser identifiers, for long periods of time. Whenever such information is forwarded or leaked, it poses a potential threat to user privacy. To mitigate such threats, the design of RBA systems must balance security and privacy.

Our study results provide a first indication that RBA implementations used in current practice can be designed to become more privacy friendly. However, there are still challenges that have not been resolved in research to date. An important question is, e.g., how the IP address feature can be replaced with more privacy preserving alternatives. On the one hand, we assume that the IP address is very relevant for re-identification attacks [9]. Discarding it from the RBA login history can therefore increase privacy protection. On the other hand, the IP address is a feature providing strong security [34]. Future research must carefully identify and analyze such trade-offs, so that RBA's user acceptance does not drop with the first data breach.

References

[1] C. C. Aggarwal, "On k-Anonymity and the Curse of Dimensionality," in *VLDB '05*. VLDB Endowment, Aug. 2005.

[2] Akamai, "Loyalty for Sale – Retail and Hospitality Fraud," *[state of the internet] / security*, vol. 6, no. 3, Oct. 2020.

[3] K. Almotairi and B. Bataineh, "Perception of Information Sensitivity for Internet Users in Saudi Arabia," *AIP*, vol. 9, no. 2, 2020.

[4] A. Blanco-Justicia and J. Domingo-Ferrer, "Efficient privacy-preserving implicit authentication," *Computer Communications*, vol. 125, Jul. 2018.

[5] J. Bonneau, E. W. Felten, P. Mittal, and A. Narayanan, "Privacy concerns of implicit secondary factors for web authentication," in *WAY '14*, Jul. 2014.

[6] M. Burkhart, D. Brauckhoff, M. May, and E. Boschi, "The risk-utility tradeoff for IP address truncation," in *NDA '08*. ACM, 2008.

[7] Y. J. Chew, S. Y. Ooi, K.-S. Wong, and Y. H. Pang, "Privacy Preserving of IP Address through Truncation Method in Network-based Intrusion Detection System," in *ICSCA '19*. ACM, 2019.

[8] European Union, "General Data Protection Regulation," May 2016, Regulation (EU) 2016/679.

[9] Europol, "SIRIUS EU Digital Evidence Situation Report 2019," Dec. 2019.

[10] Federal Committee on Statistical Methodology, "Report on Statistical Disclosure," Dec. 2005.

[11] FireHOL, "All cybercrime ip feeds," Aug. 2020. [Online]. Available: http://iplists.firehol.org/?ipset=firehol_level4

[12] D. Freeman, S. Jain, M. Dürmuth, B. Biggio, and G. Giacinto, "Who Are You? A Statistical Approach to Measuring User Authenticity," in *NDSS '16*. Internet Society, Feb. 2016.

[13] P. A. Grassi *et al.*, "Digital identity guidelines: authentication and lifecycle management," National Institute of Standards and Technology, Tech. Rep. NIST SP 800-63b, Jun. 2017.

[14] M. J. Haber, "Attack Vectors," in *Privileged Attack Vectors: Building Effective Cyber-Defense Strategies to Protect Organizations*. Apress, 2020.

[15] A. Hurkała and J. Hurkała, "Architecture of context-risk-aware authentication system for web environments," in *ICIEIS '14*, 2014.

[16] ISO, *ISO/IEC 29100:2011(E): Information Technology — Security Techniques — Privacy Framework*. ISO/IEC, 2011.

[17] D. Llewellyn-Jones and G. Rymer, "Cracking PwdHash: A Brute-force Attack on Client-side Password Hashing." Apollo, 2017.

[18] E. Markos, L. I. Labrecque, and G. R. Milne, "A New Information Lens: The Self-concept and Exchange Context as a Means to Understand Information Sensitivity of Anonymous and Personal Identifying Information," *JIM*, vol. 42, 2018.

[19] E. Markos, G. R. Milne, and J. W. Peltier, "Information Sensitivity and Willingness to Provide Continua: A Comparative Privacy Study of the United States and Brazil," *JPP&M*, 2017.

[20] E. McCallister, T. Grance, and K. A. Scarfone, "Guide to protecting the confidentiality of Personally Identifiable Information (PII)," Tech. Rep. NIST SP 800-122, 2010.

[21] I. Molloy, L. Dickens, C. Morisset, P.-C. Cheng, J. Lobo, and A. Russo, "Risk-based Security Decisions Under Uncertainty," in *CODASPY '12*. ACM, Feb. 2012.

[22] K. Moriarty, B. Kaliski, and A. Rusch, "Pkcs #5: Password-based cryptography specification version 2.1," RFC 8018, January 2017.

[23] National Cyber Security Centre, "Cloud security guidance: 10, Identity and authentication," Tech. Rep., Nov. 2018.

[24] G. Pugliese, C. Riess, F. Gassmann, and Z. Benenson, "Long-Term Observation on Browser Fingerprinting: Users' Trackability and Perspective," *PoPETS*, vol. 2020, no. 2, Apr. 2020.

[25] N. Quermann, M. Harbach, and M. Dürmuth, "The State of User Authentication in the Wild," in *WAY '18*, Aug. 2018.

[26] N. A. Safa, R. Safavi-Naini, and S. F. Shahandashti, "Privacy-Preserving Implicit Authentication," in *IFIP SEC '14*. Springer, 2014.

[27] J. Salas and V. Torra, "A General Algorithm for k-anonymity on Dynamic Databases," in *DPM '18*. Springer, 2018.

[28] E.-M. Schomakers, C. Lidynia, D. Müllmann, and M. Ziefle, "Internet users' perceptions of information sensitivity – insights from Germany," *IJIM*, vol. 46, Jun. 2019.

[29] E.-M. Schomakers, C. Lidynia, and M. Ziefle, "All of me? Users' preferences for privacy-preserving data markets and the importance of anonymity," *Electronic Markets*, vol. 30, no. 3, Feb. 2020.

[30] State of California, "California Consumer Privacy Act," Jun. 2018, Assembly Bill No. 375.

[31] R. H. Steinegger, D. Deckers, P. Giessler, and S. Abeck, "Risk-based authenticator for web applications," in *EuroPlop '16*. ACM, Jun. 2016.

[32] L. Sweeney, "k-anonymity: A model for protecting privacy," *IJUFKS*, vol. 10, no. 05, Oct. 2002.

[33] G. Venkatadri, E. Lucherini, P. Sapiezynski, and A. Mislove, "Investigating sources of PII used in Facebook's targeted advertising," *PoPETS*, vol. 2019, Jan. 2019.

[34] S. Wiefling, M. Dürmuth, and L. Lo Iacono, "What's in Score for Website Users: A Data-driven Long-term Study on Risk-based Authentication Characteristics," in *FC '21*. Springer, Mar. 2021.

[35] S. Wiefling, M. Dürmuth, and L. Lo Iacono, "More Than Just Good Passwords? A Study on Usability and Security Perceptions of Risk-based Authentication," in *ACSAC '20*. ACM, Dec. 2020.

[36] S. Wiefling, L. Lo Iacono, and M. Dürmuth, "Is This Really You? An Empirical Study on Risk-Based Authentication Applied in the Wild," in *IFIP SEC '19*. Springer, Jun. 2019.

[37] X. Xiao and Y. Tao, "M-invariance: towards privacy preserving re-publication of dynamic datasets," in *SIGMOD '07*. ACM, 2007.