基于风险的动态身份认证方案研究

魏同心[1),2),3)] 汪定[1),2),3)]

[1)](南开大学 密码与网络空间安全学院, 天津 300350)

[2)](数据与智能系统安全教育部重点实验室 (南开大学), 天津 300350)

[3)](天津市网络与数据安全技术重点实验室 (南开大学), 天津 300350)

摘 要 基于风险的身份认证（Risk-Based Authentication，RBA）通过综合评估用户登录环境与行为特征来识别账户访问风险，并根据风险水平触发附加验证，从而在账户安全性与认证可用性之间实现动态平衡。然而，现有 RBA 方法在大规模在线服务场景中仍面临特征权重相对固定、风险阈值缺乏个体适应性以及对长期历史日志依赖较强等问题，容易造成合法用户误报、附加验证频繁以及用户行为数据长期留存等问题。针对上述局限，本文提出一种基于风险的动态身份认证框架（Risk-Based Dynamic Authentication，RBDA）。该框架包括三个核心模块：首先，设计群体—个体特征重标定机制，通过联合建模同一时区或业务区域内用户群体的典型登录分布与用户近期行为特征，动态调整 IP 地址、浏览器、屏幕分辨率、登录时间和地理位置等风险特征的权重，以降低正常行为变化引起的误报；其次，引入短期行为一致性建模机制，利用一周尺度的滑动行为窗口度量当前登录行为与用户近期访问规律之间的一致性，在降低长期历史数据依赖的同时保留用户行为的时间节律特征；最后，构建动态多级阈值决策机制，根据用户历史行为稳定度生成个性化风险阈值，并结合风险等级自适应选择提示、验证码、人机识别、短信验证或多因素认证等不同强度的验证方式，实现细粒度风险处置。基于 112 个主流网站公开可观测的登录与认证机制，并结合脱敏交互日志，本文构建模拟登录环境开展非侵入式风险评估重放实验。实验结果表明，在保持与固定阈值 RBA 相当或更高攻击拦截能力的同时，RBDA 将日常活跃用户的合法登录二次验证触发率降低至 18.9%，较 Linear 相对下降约 22%，差异具有统计学显著性（$p<0.01$）。此外，RBDA 仅依赖一周的短期行为窗口即可维持较稳定的风险识别能力，表明其能够在减少长期用户行为数据依赖的同时兼顾认证安全性与用户可用性。实验结果验证了 RBDA 在动态特征建模、误报控制和自适应验证决策方面的有效性，并体现出面向大规模在线身份认证场景的实际部署潜力。

关键词 基于风险的身份认证；可用安全性；多因素身份认证；风险特征；二次验证

中图法分类号 TP391 **DOI 号** *投稿时不提供 DOI 号*

Risk-Based Dynamic Authentication

WEI Tong-Xin[1),2),3)] WANG Ding[1),2),3)]

[1)](College of Cryptology and Cyber Science, Nankai University, Tianjin 300350)

[2)](Key Laboratory of Data and Intelligent System Security, Nankai University, Tianjin 300350)

[3)](Tianjin Key Laboratory of Network and Data Security Technology, Nankai University, Tianjin 300350)

Abstract Risk-Based Authentication (RBA) evaluates account risk by analyzing users' login environments and behavioral characteristics and triggers additional authentication when necessary, thereby seeking to balance account security and authentication usability. However, existing RBA approaches still exhibit several limitations in large-scale online services. In particular, relatively static feature weights, insufficiently adaptive risk thresholds, and strong dependence on long-term historical logs may lead to elevated false-positive rates for legitimate users, frequent secondary

收稿日期：2025 年 10 月 28 日；最终修改稿收到日期：2026 年 10 月 8 日。本课题得到国家自然科学基金项目（No. 62222208），天津市自然科学基金（No. 25JCJQJC00230）资助。魏同心，博士，主要研究领域为身份认证可用安全与隐私。E-mail: weitongxin@mail.nankai.edu.cn。汪定，博士，教授，计算机学会（CCF）杰出会员，CCF 优秀博士论文奖获得者，主要研究领域为密钥安全、身份认证。E-mail: wangding@nankai.edu.cn。
通讯作者：汪定，手机号：18511345776，E-mail: wangding@nankai.edu.cn。

authentication, and prolonged reliance on user behavioral data. These limitations can reduce authentication usability and deployment flexibility while increasing concerns associated with long-term behavioral-data retention.

To address these issues, this paper proposes a Risk-Based Dynamic Authentication framework (RBDA) consisting of three core modules. First, RBDA introduces a group–individual feature recalibration mechanism. The mechanism jointly models typical login distributions among users within the same time zone or service region and the recent behavioral characteristics of individual users, and dynamically adjusts the weights of authentication-related features, including IP address, browser, screen resolution, login time, and geographic location. This design allows RBDA to better accommodate legitimate variations in user behavior and reduce false alarms caused by rigid feature weighting. Second, RBDA employs a short-term behavioral consistency modeling mechanism based on a one-week behavioral window. It evaluates the consistency between the current login attempt and the user's recent behavioral patterns, thereby retaining useful temporal regularities while reducing dependence on extensive historical logs. Compared with approaches requiring long-term behavioral histories, this mechanism focuses on recent access behavior and provides a more lightweight basis for risk assessment. Third, RBDA incorporates a dynamic multi-level threshold decision mechanism. Rather than applying a single fixed threshold across users, RBDA generates personalized access thresholds according to the stability of each user's historical behavior. It then assigns different authentication intensities according to the estimated risk level, including security prompts, CAPTCHAs, human–machine verification, SMS-based verification, and multi-factor authentication. This enables more fine-grained and adaptive security decisions for login attempts with different risk levels.

To evaluate RBDA, this study considers 112 mainstream websites covering social networking, government services, transaction-oriented platforms, travel services, and other categories. Based on publicly observable login and authentication mechanisms from these websites, together with anonymized interaction logs, we construct a simulated login environment and conduct non-intrusive replay-based risk assessment experiments. The experimental results show that RBDA maintains attack-blocking performance comparable to or higher than that of the RBA baseline. For daily active users, defined as users who log in every day or almost every day, RBDA reduces the secondary-authentication trigger rate for legitimate login attempts to 18.9%, representing an approximately 22% relative reduction compared with Linear, with statistical significance ($p < 0.01$). In addition, RBDA achieves stable risk-identification performance using only a one-week short-term behavioral window. These results demonstrate that RBDA can reduce unnecessary authentication challenges and dependence on long-term behavioral data while maintaining effective risk control. Overall, the results validate the effectiveness of RBDA in dynamic feature modeling, false-positive control, and adaptive authentication decision-making, and indicate its practical potential for large-scale online authentication scenarios.

Key words Risk-based authentication; usable security; multi-factor authentication; risk feature; secondary verification

1 引言

口令本质上是人类可记忆的短密钥，是人类可感知、直接接触的主要密码学组件，也是全球 60 亿网民最主要的身份认证手段[1]。然而，随着口令猜测算法的进步，如定向口令猜测[2]，攻击者可以利用用户的个人信息及用户在其它网站泄露的口令，极大的提高猜测攻击的成功率[3,4]。特别地，调研显示用户需要维护的口令账号数量不断增多，当前平均达到 187 个[5]。这使得用户跨站复用或轻微修改已有口令的脆弱行为会更频繁，进一步放大口令凭证泄露造成的账户风险[6]。这些导致仅依赖口令的身份认证机制面临越来越大的安全挑战。因此，迫切需要增强单一口令身份认证的安全性。

鉴于口令在可预见未来仍不可替代[1,5]，学术界提出了大量的口令认证增强方案。根据所采用技术路线的不同，这些口令认证增强方案主要可以分为两类。第一类是增强口令认证因子本身，包括强口令规则[7,8]、口令强度计[9]、强口令记忆术[10]等。然而，强口令往往导致用户记忆和管理口令困难，进而降低了用户体验[7,11]。

第二类身份认证方案通过显式引入额外的认证因素，使得单因素口令认证变为多因素身份认证（Multi-factor Authentication，MFA）[12,13]和基于风险的身份认证（Risk-based authentication，RBA）[14,15]等。该类方案要求用户在输入口令的基础上，提供额外的身份验证凭证或权限依据[16]，以进一步提高账户的安

全性。现有的多因素身份认证方式通常涵盖三类因素：知识因素（如口令）、所有权因素（如手机或安全密钥[13]）、生物特征因素（如指纹或面部识别），这些因素可以两两或三个共同验证用户的身份。通过结合这些不同的认证因素，MFA 相比单因素认证提供了更强的安全性，使攻击者很难通过多重认证步骤以突破系统。特别是，攻击者不仅需要获取用户的口令，还需要掌握用户的其他认证因素，这大大增加了攻击的难度。尽管多因素身份认证在提升安全性方面取得了显著进展，并且受到了越来越多的关注，但其实施也带来了一些挑战。首先，MFA 的实施增加了系统的复杂性，并引入了对外部设备（如手机或安全密钥）和生物特征传感器的依赖。其次，多因素认证可能导致用户体验的下降。此外，在很多场景下一成不变地要求用户进行多重验证，可能会非必要地增加登录流程的复杂性和时间成本。特别是用户在常用环境下（如相同设备、地点和时间）身份验证时固定地要求多因素认证，会带来不必要的可用性下降。

此外，多因素认证还可能引发关于用户隐私的担忧。生物特征和位置数据等敏感信息的收集和存储需要高度安全的保护措施，否则可能会引发隐私泄露。用户在享受更高安全性的同时，也必须信任系统能够有效保护其敏感数据不受攻击或泄露。静态口令认证仍面临弱口令、口令泄露等安全风险[17]。MFA 通过引入额外认证因素提高了账户安全性，但额外验证步骤也会增加认证成本并降低用户体验[13]。

第三类身份认证方案通过引入隐式认证因子方式，来增强口令认证安全性，同时不降低可用性。常见的隐式认证因子包括设备指纹和用户行为特征。该类方案依据风险的大小来判断是否启用这些隐式的认证因子。这类方案也称为基于风险的身份认证方案

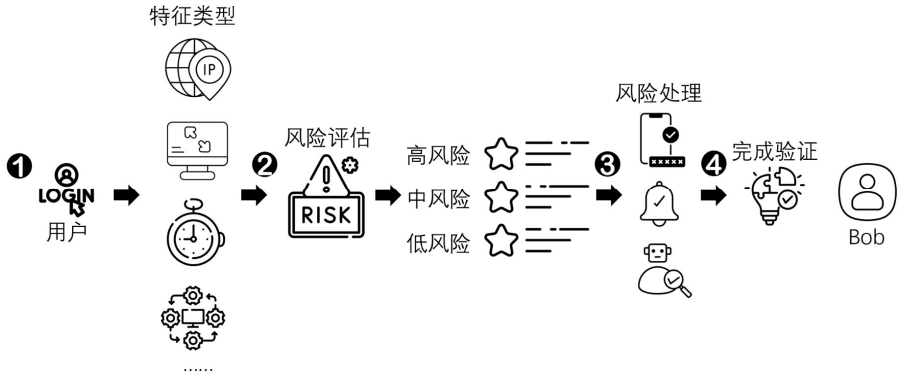


图 1 用户风险身份认证流程图

（Risk-based Authentication，RBA）[14]它通过评估用户行为模式和环境因素来评估认证风险，并根据风险是否存在来调整身份认证方式（如图 1 所示），在保证账户身份认证安全性的前提下优化了用户体验。尽管其拥有更灵活的身份认证方式，减少了不必要的身份认证和账户登录环境安全威胁的精确评估，但也面临复杂性、准确性和隐私泄露问题。

2019 年，Wiefling 等人[14]细分析了影响网站风险身份认证触发的环境特征，如 IP 地址、操作系统、屏幕分辨率、语言以及登录时间。2020 年，Wiefling 等人[15]对 592 名用户进行了调查，比较了用户对二次认证中“code”和“magic link”身份认证方式的接受度，发现用户更倾向于使用 code 身份认证方式。2020 年，Wiefling 等人[18]评估了多因素认证和基于风险的身份认证方式在用户身份认证中的可行性和可靠性，并指出在进行风险身份认证时有必要告知用户潜在的安全风险。

2021 年，Wiefling 等人[19]在一个小型在线服务上对 RBA 方案进行了调查，但该研究仅限于用户主要位于同一个城市的小型在线服务系统。为了分析用户行为的结果是否适用于全球分布的大型用户群，2022 年 Wiefling 等人[20]指出网站的安全隐私策略会影响基于风险的身份认证方式采集用户的历史环境信息。然而，在现实世界的大规模在线服务中，基于风险的身份认证迄今仍未被广泛使用[21,22,23]，因为即使是很小的配置错误也可能在很大程度上影响大批用户的可用性和安全性。为了填补实验数据的空白，Wiefling 等人[20]在一年中采集了 330 万用户和 3130 万次的登录尝试特征数据，并提出一种基于机器学习的风险评估参数优化方法。然而，其研究结果仅对挪威的在线服务类型具有代表性，并不代表全球所有的在线服务且每次登录只测量一次的登录时间，这导致实验的测试结果缺少普适性。

综上所述，当前基于风险的身份认证研究中的主要挑战包括：（1）缺乏关于触发风险认证的账户登录环境特性的标准化和一致性；（2）不同风险特征触发额外身份认证时的安全权重比例不确定，重要特征和次要特征的划分模糊；（3）针对不同文化背景和生活区域的用户数据集而言，由于用户行为习惯的客观差异，比如：不同时区的用户使用账户的登录时间也不

一致，因此难以设置完全一致的特征分类权重标准。

1.1 本文贡献

针对上述挑战，本文提出了一种基于风险的动态身份认证方案 RBDA（Risk-based dynamic authentication），可以将相同时区内用户登录习惯平均值与当前用户个体的登录行为日志映射到一个共同时区的历史登录特征空间中，方案框架图如图 2 所示。本文的贡献如下。

(1) 提出一个面向大规模在线服务的风险自适应认证框架 RBDA（见图 2）。RBDA 通过“群体-个体时区对齐”的特征重标定，将同一时区内用户群体的典型登录行为模式与目标用户的个人短期行为进行联合建模，动态估计安全相关特征（如 IP、浏览器、屏幕分辨率、登录时间、地理位置等）的权重。这解决了固定阈值 RBA 中“特征权重固定、主次特征划分模糊”的问题。

(2) 提出“周期化短期行为一致性建模”机制：使用以一周为周期的短窗口行为轨迹来度量当前登录与“近期自我”的一致性，而非依赖全年历史。该机制一方面捕捉用户稳定的日常/每周使用节律，另一方面显著降低了对长期行为留存的依赖，从而减少对长期行为日志的持续留存需求，支持 GDPR、CCPA 等数据保护法规所强调的数据最小化原则。

(3) 提出“时间周期感知的动态多级安全阈值决策”机制：区别于固定阈值RBA将风险分数与单一静态阈值比较后统一触发相同的附加验证，本文根据用户累积的登录历史稳定度动态回归个体化阈值，并将风险等级映射到一组分级验证策略（风险提示/人机验证/短信验证/MFA等）。该机制在保持高True positive rate（TPR）的同时，显著降低了合法用户遭遇高强度二次验证的频率，从而提升可用性。

(4) 对112个中国主流网站开展了RBA触发行为的系统性测量，量化了其现实部署现状（例如，在全部112个受测站点中，有41个站点在本文设定的正常登录测试条件下触发了额外验证），并基于这类真实触发特征分布对RBDA进行了评估。在对比固定阈值RBA、机器学习式RBA阈值优化和多因素认证时，RBDA在高针对性攻击（包括模拟“超级定向攻击者”）场景下展现出更强的防御韧性，同时显著减少合法用户的重复认证挑战次数。

1.2 组织架构

本文第 2 节介绍基于风险的身份认证的相关方法和研究现状。第 3 节涵盖了本文所需的基础知识，包括身份认证特征值的选择以及安全阈值的计算和比较方法，相关的攻击者模型。第 4 节详细描述了本文构建的基于风险的动态身份认证方案。第 5 节通过对比实验验证了所提方案的有效性。最后总结全文。

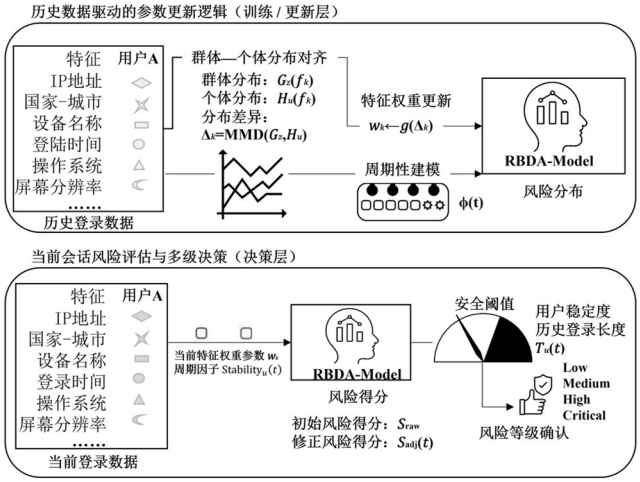


图 2 RBDA 总体框架及动态风险决策流程

2 基于风险的身份认证相关工作

现有的基于风险的身份认证测试模型大多是基于手工或自动化工具的方法且集中于触发收到通知或额外的身份认证[14,20,23]。然而对于基于风险的身份认证触发的结果和触发条件而言，其训练和实验数据是非常稀缺的，并且受语言和网站使用地址的影响，如果仅只使用某个地区或国家的实验数据用于训练则无法覆盖所有的用户数据，往往很难构建有效且实用的安全阈值和特征权重模型，以用于用户身份认证风险触发过程。研究人员提出了针对某一特定区域的风险身份认证特征权重的预测，大多数的基于风险的身份认证预测方法有一个重要的前提，即网站针对相关风险会触发额外的通知或二次身份认证。它们需要用户的历史登录数据和当前用户登录环境进行比对和度量，这对用户的行为习惯和历史登录数据有很强的依赖性，且由于每个特征权重和触发风险的验证是黑盒，近年来引起了大量的研究兴趣。

基于风险的身份认证是指通过评估用户的登陆环境风险特征和行为模式来验证其身份的一种认证方式。在这种方法中，系统会根据用户登录的环境，设备信息，历史行为等因素来确定认证的安全水平。通过对风险因素进行分析和评估，系统可以实时调整认证流程和安全措施，从而提高系统的安全性和防范潜在的身份盗窃或欺诈行为。基于风险的身份认证中，触发风险的特征是通过判断用户是否接收额外的风险通知或进行额外的身份验证来确定的。使用用户的历史登录数据，如：IP 地址和使用设备等。利用从

其他登录环境中收集的环境数据来比对当前账户登录环境的一致性和不一致性，为多因素身份认证提供了一个新的视角。当用户的身份认证历史环境与当前环境不匹配时，会触发额外的身份认证，否则单因素即可满足身份认证要求。最近，有相关工作提出了几种基于风险的身份认证网站调研[14,20,23]，用于评估当前的网站对风险身份认证使用和部署情况（即基于风险的身份认证使用率和分布）。除了通过通用的手工触发方式进行网站的风险身份认证外，使用自动化工具进行数据采集可能具有更高的效率，尤其适于大规模网站。

最初 Freeman 等人[21]提出了风险得分计算的方法，该方法通过用户登录环境的特征计算风险评分，通常分为三个级别（低、中、高风险）。根据评分和级别，服务可以执行各种操作。当评分超过低阈值且属于中等风险时，服务通常会要求用户提供额外的身份验证要素，如验证邮箱或电话号码，输入验证码，或通知用户可疑活动。在风险评分被视为高风险时，服务可能会考虑完全阻止访问，尽管这种情况很罕见，因为这可能会导致被误分类为高风险的合法用户受阻。对于每个具体的风险基于认证场景，需要仔细选择用户受到怀疑的评分阈值。

为了计算风险评分及触发风险特征的影响，Wiefling 等人[14]对八个高流量在线服务（包括亚马逊、Facebook、GOG.com、谷歌、iCloud、LinkedIn、Steam 和 Twitch）上的 RBA 使用情况进行了调查。研究团队建立了 28 个虚拟身份和 224 个用户账户。在为期 3.5 个月的研究中，他们确定一组特征还捕获和分析已部署的其他身份验证因素。研究团队还开发了一个自动化的浏览器测试框架，并在每个在线服务上模拟了类似人类用户的行为。这个框架包含增强的技术伪装措施，使其难以被区分为自动化工具。他们开发的测试框架可用于分析 RBA 功能的黑盒服务。

Wiefling 等人[18]对基于风险的身份认证，多因素认证，单因素认证的安全性和可用性展开研究，结果表明基于风险的身份认证比多因素身份认证更可用，比仅口令的身份认证更安全。并指出用户对基于风险的身份认证接受程度更多的取决于网站的类型和主要使用的设备。在处理敏感数据的网站上，基于电子邮件地址确认的 RBA 获得认可，而使用手机号码或身份验证器应用程序的重新身份验证 RBA 则较不受欢迎。因此，在 RBA 的部署中，必须谨慎考虑每个使用场景。研究结果显示，RBA 在各种网站上被认为与多因素认证同样安全，但对于高度安全要求的网站，如网上银行，MFA 更合适。用户对网站安全性的需求很高，尤其是涉及敏感数据，几乎所有处理敏感数据的网站都应考虑采用 RBA 来保护用户。

2022 年，Wiefling 等人[20]在采集了 330 万用户和 3130 万次登录尝试的特征数据，基于这些数据，提出一种基于机器学习的 RBA 参数化优化方法，来支持网站管理员为用户配置最佳的用例场景。同时通过评估网站与服务器响应的时间特征来代替 IP 地址以增强用户的隐私保护。Gavazzi 等人[23]通过调研 208 个热门网站，并引入单点登录身份验证方式来比较其与多因素认证、基于风险的身份认证、单一口令身份认证之间的安全性和可用性，并指出单点登录中的身份认证可以平衡多因素和基于风险的身份认证对用户的账户带来的隐私保护问题。RBA 已形成包括设备、位置[24]和行为上下文等风险信号的产业化认证方案[25]。

3 预备知识

本文的方法主要基于用户登录习惯的特征数据，以及这些特性数据与相关权重分布和安全阈值之间的关系。同时本文引入相关的攻击者模型。本节旨在介绍相关概念和基本知识，方便理解这一方法。

3.1 基于风险的身份认证用户登录特征

获取触发风险身份认证的特征数据的基本方法是通过在被测试账户的在线服务上创建账户，并在触发风险身份认证的条件场景下尽可能地控制相关的特征条件来观察访问账户时的结果。关于特征的提取和验证可能受到下列因素的影响。

（1）当账户的登录次数少于 5，系统无法稳定采集用户的历史行为数据，这影响账户登录时的风险评分，以及后续的风险身份认证触发结果。

（2）使用自动化工具测试可能会遇到人机身份验证，导致进一步的数据采集失败或无效数据。

（3）由于触发身份风险的特征有非常多的可选项，无法完全控制变量或简单测试所有的特征组合结果，只能大概判断哪些特征可能会触发账户的风险通知。

（4）由于 RBA 的触发结果反馈在不同的网站上的表现形式是不一致的，无法从某个固定的触发特征来判断是导致了风险身份认证，因此大多数的研究仅归类为触发或未触发两种情况。

综上所述，触发网站的风险身份认证的概率和分布是由网站部署的安全机制和用户行为来决定的。其中，可能引发风险身份认证的特征如下表 1 所示。本文在公开特征数据集上进行实验，特征包括 IP 地址、城市、使用的设备和服务器响应时间等。在这些数据集中，用户的 IP 地址、使用语言、屏幕分辨率、cookies 为引发风险通知的重要权重特征。

表 1 特征数据集

登录环境特征	参考文献来源	特征权重	用户行为特征	参考文献来源	特征权重
IP 地址	[14, 23, 24, 26, 27]	高	点击模式	[14, 23]	高
国家	[14, 20, 21, 23,27, 28, 29]	高	页面停留时间	[18, 19, 20]	中
用户代理字符	[21, 27,28]	高+	认证失败次数	[2, 14, 15, 21, 23]	高
屏幕分辨率	[14, 23,27, 28]	高+	输入模式	[5,13,23]	高
Cookies	[14, 27]	高+	多因素启用状态	[12, 13]	高
语言	[14, 21, 27]	高+	请求间隔	[14, 20, 26]	高
操作系统	[14,23,28]	中	DNS 查询信息	[27]	中
浏览器	[14, 23,28]	中	访问频率	[18, 19, 28]	中
设备名称	[23, 28]	中	社交媒体活动	[29]	低-
登录时间	[14, 23, 26, 27, 28]	低	滚动页面行为	[20]	中

其中，IP 地址、操作系统、浏览器和分辨率特征广泛应用于测试触发风险身份认证[14]，IP 地址、操作系统、浏览器和分辨率等特征在测试触发风险身份认证时扮演着重要角色。这些特征提供了关于用户身份和设备的有用信息，有助于识别异常活动和潜在风险。通过对这些特征的分析和比对，系统可以更准确地判断是否需要触发额外的身份验证步骤，从而增强安全性并防范潜在的欺诈行为。因此，综合考虑 IP 地址、操作系统、浏览器和分辨率等特征可以有效地提高身份认证系统的准确性和可靠性。本文使用了 6 个特征，包括 IP 地址、浏览器、屏幕分辨率、操作系统和登录时间、登录地址，因为这 6 个特征与用户的行为习惯息息相关也是触发风险身份认证的重要特征因素。

本文使用 RBA 的用户在线流行行为混合特征计算模型。根据[14]指出在线服务网站：谷歌和亚马逊和领英可能使用了这种特征计算方法。用户 u 的风险计算得分 S，若干特征数据集 F（f_1，f_2，…，f_n），其中第 k 个特征数据的计算如下式（1）：

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

其中，$p(F_k)$是目标用户历史登录的特征值 F_k 在全局登录数据中的出现概率，$p(F^k|u,legit)$是该特征值在用户 u 合法历史登录中的条件概率。用户在合法登录条件下的可能性概率为 $p(u|legit)$= 用户合法登录次数/总登录次数。攻击者潜在的攻击概率 $p(u|attack)=1/U$，其中，U 是用户 u 的数据集。

在 Freeman 等人的统计 RBA 模型[21]中，攻击数据 $p(u|attack)$基于用户失败登录的次数。计算这需要用户有一个非常高的错误登录次数来满足目标数据的计算。其中，$p(u|attack)$=1/总登录次数。用户的特征计算数据如下：

$$S_u(F)=\left(\prod_{k=1}^{d}p(\text{ attack}|F^k)\frac{p(F^k)}{p(F^k|u,\text{ legit})}\right)\frac{p(u|\text{ attack})}{p(u|\text{ legit})} \quad (2)$$

然而，$p(u|attack)$的参数在不同的用户之间差距非常大且没有明显的规律，难以进行计算，因而用户的真实样本的攻击模型概率分布 $S_u(F)$是难以估计的，从而通过推断获取数据集的后验分布 $p(u|legit)$以获取用户行为习惯的分布往往也很困难。

在式（2）中，训练数据的有效性高度依赖于用户错误登录次数及其历史行为结果的分布。在部分固定阈值 RBA 实现中，风险规则、特征重要度或准入阈值通常需要预先配置，并在运行期间保持相对固定[29]。近年来，越来越多的数据保护法规旨在保护用户免受在线服务对大规模数据的收集。这些法规包括 CCPA[30]和 GDPR[31]，均强调个人数据处理中的目的限制、必要性及数据最小化原则。对于基于风险的认证（RBA），这意味着一旦特征数据不再符合 RBA 目的，就应立即处理。为获得更具代表性的特征分布，本文基于历史用户的周期化登录数据进行学习，并在不同场景下使用实验账户访问在线服务，以观察行为变化。

Rivera 等[32]利用网络时延特征进行登录风险识别，而 Dasu 等[33]则基于多维登录行为构建动态风险评分机制，以实现风险驱动的自适应认证。考虑到登录历史、自动化测试干扰、功能复杂度以及反馈机制等因素可能对结果造成偏差，本文进一步构建了多个虚拟身份进行对照训练，并采用经过广泛修补的 Chromium 版本进行数据采集，从而最大限度地降低环境噪声与检测干扰的影响。

3.2 攻击者模型

通过经验账户和现实世界的场景相关性，本文参考 Wiefling 等人[20]的做法，考虑以下四种攻击：（1）朴素的攻击，基于用户登录信息的攻击；（2）基于 VPN 的攻击，对所在国家的广泛攻击；（3）定向攻击，包括一些基于登录地址、设备、浏览器的定向攻击者。（4）超级定向攻击，已经被验证的违规行为但是没有触发网站的风险验证攻击。这些攻击者的攻击能力和范围如图 3 所示。

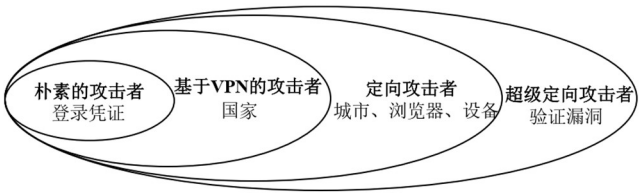


图 3 本文使用的四类攻击者模型及其逐步增强的合法用户模仿能力（参考并扩展自 Wiefling 等人[20]）。

朴素的攻击者[20]通过使用失败登录尝试的 IP 地址[34]来模拟攻击者，系统观察到大量的失败登录尝试可能表示用户成为攻击目标。为了更全面地模拟攻击者，系统只考虑每个城市中出现最多的 IP 地址，并提取了两种类型的攻击者，分别代表随机尝试登录和具有额外资源的攻击者。通过重新计算结果并评估数据集的权重，系统强调了对现实世界在线服务的风险分析的重要性，包括来自 180 个国家的攻击。

基于 VPN 的攻击者获取用户所在国家，借助 VPN 连接来伪造 IP 地理位置，并采用流行的用户代理字符串。这些 IP 地址未必来自 VPN 服务提供商，本文选用了用户的已知攻击 IP 地址来模拟这种攻击[34]。由于这类攻击并非特定于某一用户，本文从失败的登录尝试中随机抽样这些攻击 IP 地址。而在合法登录尝试中，攻击者则使用最为常见的用户代理。

定向攻击者了解受害者的位置、设备和浏览器相关的特征参数。因此，攻击者会使用用户可能选择的 IP 地址和用户代理字符串。本文选择了位于用户主要国家的已知攻击 IP 地址来模拟这类攻击者[34]。由于攻击者的目标是特定的用户群体，本文对出现频率最高的失败登录尝试的 IP 地址和用户代理进行了采样。

超级定向攻击者是目前最强大的攻击者模型，它只攻击受害者一次，但会收集所有相关信息。这可能包括访问受害者的设备或网络，例如通过恶意软件。本文通过重放 100 次成功的账户接管来模拟这个攻击者。本文这样做是为了检查在攻击数据 p 中 RBA 系统会检测到多少这些非常有针对性的攻击。

4 基于风险的动态身份认证 RBDA

本节介绍 RBDA 的组成和功能，RBDA 由三个协同的子模块组成：（i）群体-个体时区对齐的特征重标定模块；（ii）周期化短期行为一致性建模模块；（iii）历史感知的动态多级安全阈值决策模块。该框架的核心思想是：不再依赖单一的全局静态阈值，也不要求长期保存完整行为轨迹，而是将当前会话风险映射到一个个性化、分级化的认证决策中。RBDA 的整体流程如图 2 所示：

Step 1：抽取当前登录会话的特征向量（IP、浏览器、屏幕分辨率、操作系统、登录时间、登录位置等）。

Step 2：将该会话特征同时与（a）同一时区/地区用户群体的参考分布 G_z 和（b）该用户近期（短周期窗口）行为分布 H_u 进行对齐，得到群体偏差 Δ_{group} 与个体偏差 Δ_{user}。

Step 3：将这些偏差与特征重要性权重一起整合为基础风险分数 S_{raw}（参见式（5）），并结合用户近期一周内的周期性行为模式对 S_{raw} 进行时间一致性修正，得到 S_{adj}（t）（参见式（6））。

Step 4：根据用户累积的登录历史稳定度，动态回归一个个性化的访问阈值区间，并据此将 $S_{adj}(t)$ 映射到不同等级的附加验证策略，而不是总是触发同一种二次认证。这一决策即为 RBDA 的最终输出。

表 2 是主要的公式参数及相关的解释。

表 2 主要参数对照表

符号	含义	类型
f_k	第k个行为特征（例如 IP、浏览器类型、屏幕分辨率等）	特征变量
z	用户所属时区或区域编号	分类索引
$G_z(f_k)$	群体分布：时区z下所有用户在特征f^k上的分布	概率分布
$H_u(f_k)$	个体分布：用户u在过去N次登录中对特征f^k的分布	概率分布
Δ_k	分布差异度，用于衡量群体和个体在特f^k上的偏差	标量
w_k	特征权重（feature importance）	权重参数
S_{raw}	初始风险得分（未经过时间修正）	风险分值
$S_{adj}(t)$	时间一致性修正后的风险得分	风险分值（函数）
$Tu(t)$	用户u在时间t的个性化动态阈值	动态阈值函数
$L(S_{adj}(t),Tu(t))$	最终分级认证决策函数	输出等级（低、中、高、危险）

4.1 群体-个体时区对齐的特征重标定模块

不同用户在相同风险特征上的容忍度存在显著个体差异，统一阈值会导致误报率上升。群体–个体特征重标定模块关注的是跨用户的空间分布差异（例如不同时区用户的典型登录时间、IP 变化频率），用于校准特征权重；而短期行为一致性建模模

块关注的是同一用户的时间稳定性，用于修正风险得分而非调整权重。两者分别作用于“权重层”和“评分层”，功能互补。该算法包含下面四个阶段。在公式（3）中，本文定义群体分布 G_z（f_k）（同一时区/区域 z 下的平均特征），定义用户近期分布 H_u（f_k）（用户过去 N 次登录的特征），Δ_k 用于更新 w_k，本文使用最大均值差异（Maximum Mean Discrepancy, MMD）衡量分布差异，通过比较群体与个体的统计分布表示，模型在决策阶段无需直接使用完整的长期原始行为轨迹，从而降低对原始日志持续参与风险计算的依赖。分布偏差计算（群体-个体特征对齐）：

$$\Delta_k=\mathrm{MMD}(G_z(f_k),H_u(f_k)) \quad (3)$$

使用最大均值差异度量群体与个体在特征空间的距离。其中，Δ_k 越大表示个体近期行为与群体典型模式差异越显著。

本文提出的动态身份认证方案。该方法由 3 部分组成。首先，通过测试以往触发风险身份认证的特征来进行中国网站的风险身份认证。尽可能通过控制变量来将触发风险身份认证的特征进行权重分类。

算法 1：群体–个体时区对齐的特征重标定（RBDA-Module I）

INPUT: 当前会话特征集合 $F = \{f_1, f_2, ..., f_n\}$, 用户 u 的短期行为分布 $H_u(f_k)$ (k=1..n), 时区索引 z 及群体分布 $G_z(f_k)$ (k=1..n), 超参数 α
OUTPUT: 分布偏差向量 $\Delta = \{\Delta_1,...,\Delta_n\}$ 与特征权重映射 $W = \{w_k\}$

$\Delta \leftarrow \varnothing$
$W \leftarrow \varnothing$
FOR k = 1 to n **DO**
  $\Delta_k \leftarrow \mathrm{MMD}(G_z(f_k), H_u(f_k))$ //公式(3)
  $\mathrm{wk} \leftarrow \exp(-\alpha \cdot \Delta_k)$ // 特征可信度/权重更新
  $\Delta \leftarrow \Delta \cup \{\Delta_k\}$
  $W[\mathrm{fk}] \leftarrow w_k$
END FOR
RETURN (Δ, W)

在公式（4）中，本文进行权重重标定（Adaptive Feature Weighting）：

$$w_k' = \frac{w_k}{1+\alpha\cdot\Delta_k} \quad (4)$$

其中α为调整系数，差异越大，特征可信度越低。

其次，引入最大均值差异距离来度量整体用户身份认证环境特征与目标用户身份认证环境特征空间之间的距离。通过最小化该距离，提取出二者的共性身份认证特征表示，并区分各自的特性分布表示。最终目标是基于相似的条件概率分布来表征整体身份认证环境特征与目标用户的行为特征空间。

算法 2：MMD 距离计算与特征分布表示（RBDA-Module I）

INPUT: 时区索引 z, 群体分布 $G_z(f_k)$ (k=1..n),个体短期分布 $H_u(f_k)$ (k=1..n), 核函数 $K(\cdot,\cdot)$
OUTPUT: 分布差异向量 $\Delta = \{\Delta1,...,\Delta n\}$, 群体特征分布表示 $\Phi G = \{\Phi G_1,...,\Phi G_n\}$, 个体特征分布表示 $\Phi H = \{\Phi H_1,...,\Phi H_n\}$

$\Delta \leftarrow \varnothing$
$\Phi G \leftarrow \varnothing$; $\Phi H \leftarrow \varnothing$
FOR k = 1 to n **DO**
  $\Phi G_k \leftarrow$ Represent($G_z(f_k)$) //群体分布表示（统计均值嵌入）
  $\Phi H_k \leftarrow$ Represent($H_u(f_k)$) //个体分布表示
  $\Delta_k \leftarrow \mathrm{MMD}(\Phi G_k, \Phi H_k; K)$ //计算最大均值差异
  $\Delta \leftarrow \Delta \cup \{\Delta_k\}$
  $\Phi G \leftarrow \Phi G \cup \{\Phi G_k\}$
  $\Phi H \leftarrow \Phi H \cup \{\Phi H_k\}$
END FOR
RETURN (Δ, ΦG, ΦH)

最后，增加一个短期循环时间段内用户行为特征与时间的相关关系。本文对比了不同的时间范围内用户的行为特征与时间的相关关系，以筛选出既能保证用户行为特征与时间的相关性又能保证用更少的用户行为数据来使得用户的隐私安全得到更多的保护。按照以往的触发风险身份认证的特征对用户在不同时间节点触发的特征进行分类。通过时间节点相关性分析，学习一个强大的周期循环非线性分类器，提取与触发风险认证相关的信息，从而提高模型的周期分类效果。

4.2 用户登录数据

网站选取及账户创建。本文致力于报告互联网用户所面临的用户身份验证安全状态。为了选择测试网站，本文在 2023 年 6 月的中国 Webmaster 排名中进行了筛选。该排名描述了中国网站的排名情况，综合考虑了网站流量、质量、技术能力和社交影响力等方面。本文首先选择了八个主题网络服务，这些主题包括旅行、日常生活、学习、娱乐、社交、买卖交易、政务和新闻。本文重点关注了排名前 200 的 100 个网站，然后在排名 200 到 500 之间的数据集中随机选择了 12 个网站，以增加研究广度。

尽管本文选择的网站范围广泛，但仍有许多其他主题需要进一步探索，比如电子银行和电子健康。本文未包含它们的主要原因是它们通常依赖多因素身份验证技术。尽管如此，本文的样本涵盖了当前在线网络服务中最受认可和领先的部分，这些服务吸引了大部分访问流量，它们所采用的 RBA 策略将影响大多数最终用户。在未来的工作中，本文

计划将每个主题的网站平均数量增加到 20 个，以提供更多切实可行的 RBA 部署方案的支持。

大多数网站提供基本类型的用户账户，但有些网站提供多种账户类型。例如，淘宝允许用户以买家或卖家身份注册，本文根据网站的主要服务对象类型创建账户。根据 Wiefling 等人的长期 RBA 实证研究[19]，在其具体数据集和 EXTEND 模型设置下，约 10 条成功登录历史可支持对 99.92%定向攻击尝试的稳定检测。因此，本文在正式触发测试前首先进行 10 次正常登录，以建立初始账户行为历史。在账户创建和身份绑定过程中，本文遵循以下步骤：首先访问可创建账户的网站，记录认证方式并优先选择基于口令的账户创建认证。判断是否需要绑定身份信息以及具体的绑定方法，确保信息的有效性和准确性。然后，本文在相同环境下进行十次登录和正常浏览行为，以训练账户。三天后，本文尝试在不同场景下登录以触发 RBA。网站评估用户登录风险主要通过比较当前登录特征和历史登录特征，根据特征值的变化判断登录行为的合法性。尽管 RBA 的具体身份验证方案和程序尚未公开，本文缺乏 RBA 属性和配置的开源资料，但本文选择可能帮助网站识别用户登录的指标，以分析可能触发网站 RBA 策略的用户特征。

登录环境特征选取。在基于风险的认证部署中，本文综合利用不同特征如登录行为、时间、地点来评估各网站采用的 RBA 形式。通过黑盒测试诸如输入口令错误、屏幕分辨率变化、非常规登录时间、IP 地址变化等特征来触发 RBA，并观察额外身份验证的反应。这有助于本文探索 RBA 触发结果，并有选择地使用特征以引发额外认证需求。同时，本文的 RBA 模型以 IP 地址和用户代理字符串作为特征，包括 IP 地址的国家和城市，以及用户代理字符串的浏览器、操作系统和设备类型。本文选择这种方法是为了与相关研究进行公平比较，因为这一特征集可视为 RBA 实践的状态。

特征优化及数据采集。庞大的数据集规模显著影响了计算时间，限制了计算能力。为了应对这一挑战，本文进行了 RBA 算法的优化以提高计算速度。这种优化使本文能够快速计算所有合法登录的风险评分。尽管如此，本文仍需将攻击者的计算范围限制在一个数据子集内，以在合理时间内得出结果。

本文采用了分层方法，基于包含 20%用户的层次化数据集，计算攻击者对朴素攻击者、基于 VPN 的攻击者和定向攻击者的风险评分。通过根据登录次数对用户集进行分层，本文尽可能保留有关登录历史的数据属性。分层数据集在机器学习解决方案中是常见的做法。为了重点关注现实世界的攻击，本文选择包括那些登录周期行为在相应时间段稳定的用户，即可能成为目标的用户。

为了确保分层不会影响本文研究结果的普适性，本文将所有合法登录用户的结果与使用分层数据集的结果进行了比较。结果显示，对于少于 100 次登录历史记录的大多数情况，两种方法的结果一致。因此，本文认为分层样本适用于本文的分析。

4.3 周期化短期行为一致性建模模块

用户登录行为在短时间尺度内呈现显著的时间相关性与周期性，而固定阈值难以适应这种动态变化。一种数据最小化方法是在一定时间后删除全局登录历史记录，这也有助于维持认证速度。为了评估最小化方法的潜力，在公式（5）中，本文进行了初始风险得分（Raw Risk Score）评估：

$$S_{\text{raw}} = \sum_{k=1}^{K} w_k' \cdot R(f_k) \qquad (5)$$

其中 $R(f_k)$表示该特征在当前会话中的风险指示值（如异常标志、黑名单命中率等）。

登录历史最小化与攻击注入设置。为评估在数据最小化约束下 RBDA 的安全性与可用性，本文基于第一项研究中构建的模拟用户登录序列，分别采用滑动窗口保留用户最近 1 周、1 个月与 1 个季度的登录历史，并将保留 1 年历史的设置作为对照组。为保证风险模型在每次认证时均可计算特征统计量，本文对每位用户的历史记录设置最小保留约束，即任意时刻至少保留 2 条历史登录记录。

为保证不同历史保留窗口之间的比较仅受到可用历史信息量的影响，本文采用固定相对位置的攻击注入策略。首先按照时间顺序排列每名用户的合法登录序列，并在满足最小历史长度要求的用户中选择预设分位位置 q 作为攻击注入锚点 t^*。

本文在不同保留窗口下统计两类指标：（1）安全性指标：攻击拦截率（TPR）；（2）可用性与成本指标：合法用户二次验证触发率（以中位数呈现）与历史存储量（以保留记录条数/容量近似衡量）。

公式（6）首先进行归一化稳定度，然后计算了一周七天的周期函数时间一致性修正：

$$Stability_u(t) = e^{-V_u(t)}$$

$$S_{adj}(t) = S_{raw}(t)[1+\beta(1\text{-}Stability_u(t))] \qquad (6)$$

根据过去一周风险得分波动率（行为稳定性）进行修正，β 控制时间一致性权重。其中，$S_{\text{raw}}(\tau)$表示在最近一个时间窗口$[t-\Delta t, t]$内，用户风险原始

得分序列$S_{\text{raw}}(\tau)$的（样本）方差，用于度量该用户在短期内的风险评分波动程度。

算法 3：周期化短期行为一致性建模与时间一致性修正（RBDA-Module II）

INPUT: 当前会话特征 $\{f_k\}k=1..K$ 及其风险指示 $R(f_k)$，重标定后的特征权重 $\{w'k\}_{k=1..K}$，用户 u 的历史风险序列 $\{S_{raw}(\tau)\}\tau\in H_u$ 及最小化窗口参数 Δ_t，参数 β 与最小保留约束 min_hist = 2

OUTPUT: 初始风险得分 $S_{raw}(t)$ 与修正风险得分 $S_{adj}(t)$

$S_{raw}(t) \leftarrow 0$
FOR $k = 1$ to K **DO**
  $S_{raw}(t) \leftarrow S_{raw}(t) + w'_k \cdot R(f_k)$ //式(5)
END FOR
$H \leftarrow \{S_{raw}(\tau) \mid \tau \in [t-\Delta t, t]\}$ //滑动窗口保留
IF $|H| <$ min_hist **THEN**
  $H \leftarrow H \cup \{S_{raw}(t)\}$ //保证至少 2 条历史
END IF
$V \leftarrow V_{ar}\tau\in[t-\Delta t,t]\ (S_{raw}(\tau))$ //短期波动率
$S_{adj}(t)\leftarrow S_{raw}(t)\times(1+\beta\cdot(1-Stability_u(t)))$ //式(6)
RETURN $(S_{raw}(t), S_{adj}(t))$

本文建模用户在同一星期几、相近时间段的典型特征分布，衡量当前登录与其“周期自画像”一致性，作为 S_{adj}（t）的一部分。与依赖全年登录历史的（Linear）不同，RBDA 仅使用一周量级窗口即可稳定地区分合法行为与攻击者行为（第 5 节），显著降低了行为留存需求，同时保留了可用性和检测能力。这一点对满足 GDPR/CCPA 等数据最小化要求尤其关键。

在获得用户近期行为稳定度后，RBDA 进一步依据式（7）生成用户(u)在时间(t)的个性化动态访问阈值(T_u(t))。当用户近期行为较稳定时，阈值相应提高，以减少不必要的附加验证；当用户行为稳定性下降时，阈值相应降低，使异常会话更容易进入增强认证流程。

$$T_u(t) = T_0 + \gamma \cdot \text{Stability}_u(t) \quad (7)$$

当用户近期行为稳定时，阈值上升（减少验证概率）；不稳定时，阈值下降（增加验证概率）。

算法 4：周期自画像一致性与迭代更新（基于周周期与 MMD）

INPUT: 当前会话特征向量 $F(t)$
用户 u 的一周历史窗口 H_uW
周期分桶函数 $b(t) = (\text{weekday}(t), \text{timeslot}(t))$
阈值 δ 及核函数/参数 K

OUTPUT: 周期一致性偏差 Dper 与更新后的周期画像 Π_u

$b \leftarrow b(t)$
$H_b \leftarrow \{F(\tau) \mid \tau \in H_uW \text{ and } b(\tau) = b\}$ // 同一星期几同时段
IF $|H_b| = 0$ **THEN**
  $\Pi_u[b] \leftarrow \text{InitProfile}(F(t))$ //当前未验证会话不能直接进入可信画像；验证失败不得更新画像
  RETURN $(0, \Pi_u)$
END IF
$\mu_b, \sigma b^2 \leftarrow \text{MeanVar}(H_b)$ //均值为共性 方差为个性
Dper ← MMD(H_b,{F(t)}) //用 MMD 对齐均值表征，周期画像核均值距离
IF Dper > δ **THEN**
  Verification //提高验证强度；只有附加验证确认会话属于合法用户后，才允许更新周期画像。
END IF
$\Pi_u[b] \leftarrow \text{ReplaceOrUpdate}(\Pi_u[b], F(t))$ //循环迭代习惯替换
RETURN (Dper, Π_u):

本文比较每周同一天的数据的最大均值和异常值，以便有效地计算历史数据（例如星期一）与当前登录时间的用户行为数据之间的差异。通过引入最大均值差异损失来处理历史数据和当前行为数据的均值参数，本文可以确保两者的概率分布均值参数具有相似的分布，同时通过方差参数来体现差异。均值参数描述了用户历史数据行为与当前行为的共性特征，而方差特征则突显了用户历史行为数据和当前行为数据的个性特征。如果用户的未来行为与已确定的历史行为数据不符，在公式（8）中，本文进行循环迭代，以新的行为习惯取代历史行为数据。其中，多级认证决策计算如下：

$$L\left(S_{adj}(t), T_u(t)\right) =$$

$$\begin{cases} \text{Pass}, & S_{adj}(t) < T_1 \\ \text{Soft Verify (Captcha)}, & T_1 \le S_{adj}(t) < T_2 \\ \text{Hard Verify (2FA)}, & T_2 \le S_{adj}(t) < T_3 \\ Block, & S_{adj}(t) \ge T_3 \end{cases} \quad (8)$$

4.4 历史感知的动态多级安全阈值决策模块

风险并非二元状态，而是具有连续性，不同风险等级应匹配不同强度的认证响应。当前的 RBA（基于风险的认证）文献[20,21,29,32]指出，准入阈值通常是一个静态的、单一数值，用于区分两类风险状态（有风险/无风险）。然而，这种二值化决策机制在大规模在线服务中存在明显局限：一方面，固定阈值难以适应不同用户行为稳定度的差异；另一方面，统一的二次认证方式无法区分风险程度的高低，容易导致低风险合法用户遭遇不必要的高强度验证，从而降低系统可用性。

为解决上述问题，本文提出一种历史感知的动态多级安全阈值决策机制。该机制的核心思想并非仅生成一个动态阈值，而是将连续风险评分空间划分为多个风险等级区间，并为不同区间配置差异化的认证响应策略，从而实现风险感知的分级决策。

算法 5：历史感知的动态阈值与多级认证决策（RBDA-Module III）

INPUT: 修正风险得分 $S_{adj}(t)$,用户稳定度 $Stability_u(t)$ //可由短期波动率等归一化得到
基准阈值 T_0 与参数 γ,两级分界偏移 η_1, η_2, η_3 //用于生成 T_1,T_2,T_3

OUTPUT: 分级认证决策 $L(S_{adj}(t), T_u(t))$
$T_u(t) \leftarrow T_0 + \gamma \cdot \text{Stability_u}(t)$　// 式(7)
$T_1 \leftarrow T_u(t) + \eta_1$
$T_2 \leftarrow T_u(t) + \eta_2$
$T_3 \leftarrow T_u(t) + \eta_3$
IF $S_{adj}(t) < T_1$ **THEN**
　　L ← Pass
ELSE IF $S_{adj}(t) < T_2$ **THEN**
　　L ← Soft Verify (Captcha)
ELSE IF $S_{adj}(t) < T_3$ **THEN**
　　L ← Hard Verify (2FA)
ELSE
　　L ← Block
END IF
RETURN L

（1）多级风险区间的定义

在 RBDA 中，系统以时间一致性修正后的风险评分$S_{\text{adj}}(t)$作为决策输入，并结合用户登录历史长度L_u与行为稳定度σ_u，为每个用户动态生成一组个性化阈值边界$\{T_u^{(1)}, T_u^{(2)}, T_u^{(3)}\}$。据此，将风险评分空间划分为四个等级区间：

Low：$S_{adj}(t) < T_u(1)$，表示当前登录行为与用户历史模式高度一致；

Medium：$T_u^{(1)} \le S_{adj}(t) < T_u^{(2)}$，表示存在轻微异常；

High：$T_u^{(2)} \le S_{adj}(t) < T_u^{(3)}$，表示较高风险；

Critical：$S_{adj}(t) \ge T_u^{(3)}$，表示高度异常或潜在攻击行为。

上述区间划分使 RBDA 能够区分不同程度的风险状态，而非简单地进行“通过或者拒绝”的二值判断。

（2）风险区间到身份认证行为的映射关系

在完成风险区间划分后，RBDA 将不同风险等级映射到分级认证策略：

输入：S_{adj}（t）、用户登录历史长度 L_u、历史稳定度 σ_u
输出：一个风险区间分配 {Low，Medium，High，Critical}
再映射到动作：
Low → 无额外验证
Medium → 人机验证/图形验证码
High → 短信验证码
Critical → 阻断

这种分级响应机制使系统能够在风险逐步上升时平滑增加认证强度，避免在低风险场景下引入高成本验证，从而在安全性与可用性之间取得更优折中。

（3）动态阈值生成与多级决策的关系

动态访问阈值在 RBDA 中并非最终决策结果，而是多级风险区间的边界生成机制。为生成个性化阈值，系统基于合法用户历史登录行为，利用回归模型学习风险评分随登录历史长度变化的趋势。实验观察表明，合法用户的风险评分在初期登录阶段快速下降，随后趋于稳定。为降低异常值对阈值学习的影响，本文对风险评分最高的 5%数据进行截断处理，并确保生成的阈值始终为正值；当阈值不满足约束条件时，系统保留上一阶段的有效阈值。通过上述设计，RBDA 实现了“个体化阈值回归+分级认证决策”的联合机制，而非固定阈值 RBA 中的单阈值触发策略。图 2 给出了该多级决策机制与参数更新逻辑之间的关系示意。

5 实验分析

本文针对节 4.2 中筛选出的 112 个中国网站展开 RBA 研究，旨在设计面向国内网站的 RBA 策略的安全阈值形式和触发结果，并通过用户调研评估不同 RBA 策略的可行性。实验结果显示，国内网站对 RBA 的触发率较低，仅为 59%（66/112）。此外，这些网站在确定 RBA 安全阈值方面存在明显困难，难以精确地区分合法登录和非法登录。具体而言，在采用 RBA 策略的网站中，有 37%（41/112）的网站将普通用户的合法登录行为视为高风险，并要求额外的身份验证（例如：通过手机验证码验证登录者身份），增加了合法用户登录的复杂性和时间成本。参与者招募方式、知情同意、匿名化方式、获得南开大学伦理审批号和伦理豁免说明。

在我们的 100 名中国用户调研中，大多数用户的登录频率为每日登录（50.0%）或每周登录（35.0%）。用户使用的设备主要是手机（75.0%）和台式机（25.0%），其他设备类型包括机器人和无法识别的设备。台式机设备的操作系统为 Windows（85.0%）、macOS（10.0%）和 Linux（5.0%）。大多数移动设备使用的操作系统为 Android（80.0%）和 iOS（20.0%）。在线浏览器主要为 Chrome（40.0%）、QQ 浏览器、UC 浏览器、360 浏览器等其他浏览器（40.0%）、Safari（10.0%）、Edge（5.0%）和 Firefox（5.0%）。这些浏览器的使用比例反映了中国市场的浏览器份额。为评估 RBDA 在长期使用场景下的行为稳定性，本文构建了一个模拟用户账户登录行为数据集。该数据集在时间跨度、登录频率分布以及周期性特征上对齐主流在线服务中用户行为的统计特性。

表 3 展示了不同类型网站在各种特征变化下触发 RBA（风险基础认证）的情况。网站的分类包括社交、政务、买卖交易和旅游四类，百分比数据用于展示每类网站在特定条件下触发 RBA 的频率及其触发方式的占比。图表共包含四组数据：人机交

互占比、短信验证占比、有 RBA 触发的网站占比以及触发 RBA 的总占比。从数据中可以看出，买卖交易类网站在触发 RBA 的情况下，人机交互（50%）和短信验证（40%）的占比较高。这意味着用户在这些网站上进行交易或登录时，较为常见的是通过人机验证或短信验证进行额外的身份认证和[23,35,36]的调研结果一致。这类网站对安全要求较高，更频繁地启用二次身份验证来保护用户账户安全。相比之下，旅游类网站在人机交互（20%）和短信验证（10%）方面的占比明显较低，反映出这类网站对安全认证的要求相对较少，可能是因为这类网站的交易或敏感数据较少。此外，社交和政务类网站的数据位于中间，分别在人机交互和短信验证上有适度的应用。触发 RBA 的总占比也因网站类型而异。买卖交易类网站的总触发率最高，达到 12%，这与其更高的安全需求相符合。旅游类网站的触发 RBA 总占比仅为 5%，显示出其对身份认证风险的敏感度较低。

为了深入理解不同特征对基于风险的身份验证（RBA）触发的影响，本文进行了网站统计分析，并将结果汇总在表 3。该表展示了各个特征触发 RBA 的总占比，使本文能够评估每个特征在 RBA 机制中的重要性。登录特征包括登录行为、Cookie、登录地点和登录时间，触发 RBA 的百分比，范围从 0%至 100%。总体来看，触发 RBA 的特征排序为：登录行为>Cookie>登录地点>登录时间。

表 3　不同网站类别及部分典型实验条件下的 RBA 触发结果(n=112，保留整数)

网站分类	人机交互占比	短信身份验证占比	有 RBA 机制网站占比	触发 RBA 的占比	触发 RBA 特征	触发 RBA 的占比
社交	**40%**	**30%**	**50%**	**10%**	**用户登录行为**	**80%**
政务	**35%**	**25%**	**45%**	**9%**	**Cookie**	**70%**
买卖交易	**50%**	**40%**	**60%**	**12%**	**登录地点**	**50%**
旅游	**20%**	**10%**	**25%**	**5%**	**登录时间**	**30%**

登录行为的异常变化，即用户在登录账户时的行为与其正常登录习惯出现显著偏差，通常被视为一种常见的安全风险。这可能意味着账户被未经授权的人员访问，从而增加了账户被攻击或滥用的风险。例如，当用户忘记用户名和口令时，可能会尝试多次不成功的登录操作；或者用户在不同设备或浏览器上进行登录，这些都属于非典型的登录行为。调研结果显示，有 27%的网站在用户输入口令错误次数达到一定阈值后，会暂时锁定账户，禁止用户继续登录，锁定时间从 5 分钟到 24 小时不等。少数（3.6%）网站会向用户发送警报通知，提醒用户有异常登录尝试，并建议用户检查账户安全情况。例如：66law.cn，该网站还同时显示当前登录的 IP 地址和咨询电话。

少数（3.6%）网站为了确保账户安全，要求用户在一定时间内无法重新尝试登录，同时提供口令重置功能。用户可以通过验证身份的方式（如邮箱验证、手机验证码等）重新设置口令。90%以上的网站在多次验证口令失败后增加了验证码的检测环节，验证码包括数字、图形和智能验证等形式。其中，3.6%的网站会在锁定账号的同时更改登录方式，如小程序或手机端 APP 扫码登录。3.6%的网站在智能验证的同时增加了短信验证。

口令输入错误的处理方式主要包括账户锁定、发送警报通知、重置口令和安全验证等，其中安全验证策略最为常见。另外，9%的网站会在用户从注册时使用的屏幕分辨率更换至其他参数时触发 RBA 策略，额外增加图形或短信验证。8%的网站在用户从注册时使用的浏览器更换至其他浏览器登录时会触发 RBA 策略，额外增加图形验证或数字验证。这表明一些网站会根据用户的登录环境变化对其进行额外身份验证的要求。7.6%的网站会在用户登录时出现手机熄屏现象时触发 RBA 策略，其中 50%会增加手机验证码，50%增加图形验证码。

针对异常登录时间的实验结果显示，10%的网站在非高峰时段（例如，当地时间凌晨 02:00）尝试登录时会触发 RBA，其中 80%的网站要求用户输入手机验证码，20%的网站要求用户输入图形验证码。这意味着一些网站在罕见登录时间进行登录时会采取额外的验证措施。实验结果还显示，4%的网站会在用户隔天登录时触发 RBA，要求输入验证码，这意味着一些网站会在登录频率异常时对用户进行额外的验证来确保账户安全。然而，实验结果显示，通过虚拟机登录的 RBA 触发率仅为 14%（16/112），低于多次口令错误的 59%（66/112）、关闭 Cookie 的 39%（44/112）和改变登录地址的 17%（19/112）。这一发现表明，在中国网站的 RBA 策略中，低权重特征的组合不能达到与高权重特征同水平的触发率，意味着中国网站 RBA 策略中各特征的权重比例差异极大。

85%的网站触发 RBA 的形式为增加人机验证或者通过短信验证码的形式来验证用户身份。其中，80%

的网站实施了人机验证，作为一种有效的手段来减少自动化工具的攻击。短信验证码作为一种身份验证手段，其使用频率相对较低，只有 25%的网站采用了短信验证码。值得注意的是，为了提供更全面的账户安全保护，20%的网站会根据具体情况，结合使用人机验证和短信验证码。例如，在用户首次注册或登录时，网站会首先使用人机验证来确认用户的真实性；而在检测到潜在的安全风险时，如异常登录行为，进一步要求用户输入短信验证码，以确保账户的安全。这种多层次的安全验证策略有助于提高网站的安全性，满足不同网站的安全需求。

5.1 评价指标及基准模型

为了评估 RBA 在大型在线服务中的性能，本文采用了与原 RBA 分析相同的方法，这使本文能够有效地比较结果。

- 安全性：攻击拦截能力（TPR）
- 可用性：合法用户被要求二次验证的频率
- 误报控制：合法登录被判为高风险的比例
- 数据留存与隐私合规成本
- 动态多级阈值 vs 固定挑战

首先，本文根据每个攻击者模型的拦截攻击者百分比（真阳性率，TPR）[20,21]校准了风险评分。接下来，本文重演了在线服务上发生的所有合法登录尝试。对于每次登录尝试，本文（i）恢复登录时的数据集状态，（ii）计算风险评分，（iii）应用给定 TPR 和攻击者模型的访问阈值，并（iv）记录访问决策。

本文将所提方法 RBDA 在四种攻击模型：朴素的攻击，基于 VPN 的攻击，定向攻击，超级定向攻击中进行了测试，同时与 RBA（Linear）[21]、基于机器学习的 RBA[20]进行了比较。在对比实验中，本文对（Linear）方法进行了规范化实现。该基线采用与 RBDA 相同的特征集合，包括 IP 地址、浏览器、操作系统、屏幕分辨率、登录时间与登录地点，并基于特征异常程度计算会话风险评分。风险评分通过对各特征相对于用户历史行为分布的偏离情况进行加权汇总。在决策阶段，固定阈值 RBA 使用单一全局固定阈值判断是否触发额外身份验证，该阈值在验证集上进行经验性选取，以平衡攻击拦截率与合法用户二次验证频率。需要说明的是，该基线方法不包含个性化阈值调整、时间一致性修正或多级认证策略，从而确保其与 RBDA 的对比仅反映阈值策略与决策机制层面的差异。

为了提高 RBA 的可用性，用户需要注意与 MFA 的差异[18]。一个基准是要求不频繁地进行二次身份验证[20]。因此，本文将所需的登录历史数据大小定义为在其中位数重新身份验证率保持在 0.5 以下的点。

根据这一定义，大多数 TPR 低于 0.99 的情况下，一个稳定的设置需要一个或没有登录历史记录。其他情况下，TPR 需要的条目数分别为四个（对于简单攻击者，TPR 为 0.995；对于 VPN 攻击者，TPR 为 0.99），六个（对于有针对性的攻击者，TPR 为 0.99），八个（对于 VPN 和有针对性的攻击者，TPR 为 0.995），以及 18 个（对于非常有针对性的攻击者，TPR 为 0.8276）。这将建立一个基准设置，其中中位数情况下不到一半的时间需要进行重新身份验证。

研究表明，用户登录频率会影响维持稳定 RBA 配置所需的历史记录数量。Wiefling 等人[20]以合法用户的中位重新认证率低于 0.5 作为稳定配置标准。在其数据集和实验设置下，当目标 TPR 为 0.995 时，每日活跃用户在定向攻击场景中使用约 4 条历史记录即可达到该稳定标准，而其他较低频用户通常需要约 8 条历史记录。对于能够高度模拟合法用户特征的超级定向攻击者，若要求进一步提高攻击拦截率，则需要显著增加历史数据量，在部分实验条件下约 40 条历史记录后才能获得较稳定的安全－可用性折中。

这些数值反映的是既有 RBA 方法在特定数据集和攻击模型下的表现，不应直接推广为 RBDA 的最优历史长度。本文在此基础上进一步研究 RBDA 中登录频率与二次验证开销之间的关系，并在 TPR=0.995 的定向攻击场景下采用统计检验比较不同登录频率用户，结果见表 5。

RBA 模型提供了两种包含攻击数据的可能性（成功触发 RBA/未触发 RBA），这些可能性尚未被研究。

攻击拦截能力（TPR）。本节评估了攻击数据的变化。本文模拟了与之前研究相同的登录行为。然而，在风险评分计算中加入了两种不同的攻击数据变化。在第一种变化中，本文使用了从攻击数据中计算的用户攻击概率 P（a|用户），该概率是根据用户的循环周期登录尝试次数计算的。当加入用户攻击概率时，朴素攻击者、VPN 攻击者和定向攻击者的重新身份验证率显著增加（$p<0.0001$）。然而，对于超级定向攻击者，当 TPR 低于 0.94 时，重新身份验证率显著下降（$p<0.0001$）。与之前相比，超级定向攻击者可以更好地与合法用户区分开，而其他攻击者则相反。在第二种变化中，本文使用了包含用户攻击概率 P（a|用户）和特征攻击概率 P（特征值|a），该概率是根据失败的登录尝试数据计算的。当也包括特征攻击概率时，所有攻击者模型的重新身份验证率显著增加（$p<0.0001$），攻击者无法有效识别合法用户。

5.2 实验方法

本文采集自中国网站和模拟用户行为的身份认证模拟数据集作为实验数据集，并执行基于用户行为

和短期时间循环的用户行为预测。

本文从 6 个触发账户风险通知的特征中选择一个作为目标，然后依次使用其他 5 个特征中的数据作为干扰因素。例如，当用户登录的 IP 地址作为目标时，存在其他五个特征的二分类预测组合。由于本文主要关注这 6 个特征在用户行为数据集中的预测，因此本文没有对数据集中具有相同度量集的同一组项目进行缺陷预测。总共有来自 6 个小组的 64 种可能的预测组合。

由于数据集针对的是 RBA 系统，尤其是 Freeman 等人[21]的模型，所有用户在全局和局部的分类数据的统计特征概率是相同的。为了保持特征之间的逻辑关系和时间顺序，其他数据都是随机生成的。时间戳包含随机性，并且不完全相同。与 IP 地址和用户代理字符串相关的特征值是通过公开可用的数据随机生成的，因此这些值很可能不在真实数据集中。RTT（往返时间）类似于真实值，但在每个地理位置中随机分配给用户。因此，RTT 条目在原始数据集中可能位于其他位置。

设备类型与真实数据集相同。基于设备类型，本文随机分配操作系统，并基于操作系统分配浏览器信息。从这些信息中，本文随机生成用户代理字符串。因此，用户代理的所有逻辑关系与真实数据集中的保持一致。RTT 是从登录成功状态和合成的地理位置数据中随机抽取的，以确保 RTT 值的现实性。

基于本文的评估，该数据集能够再现本文关于 RBA 模型行为的研究结果，使用的特征包括 IP 地址（IP 地址、国家和城市）和用户代理字符串（完整字符串、操作系统名称和版本、浏览器名称和版本、设备类型）。

5.3 实验结果与分析

为了评估基于风险的身份认证研究方法的有效性，本文研究了以下三个问题。

• RQ1：不同攻击者能力与目标 TPR 如何影响合法用户的二次验证开销？

• RQ2：用户登录频率与历史行为窗口长度如何影响 RBA 的安全性和可用性？

• RQ3： RBDA 与固定阈值 RBA、基于机器学习的 RBA 相比表现如何？

RQ1：不同攻击者能力与目标 TPR 如何影响合法用户的二次验证开销？

为了验证这个问题的结果，本文将 RBDA 模型与四种攻击者模型进行了对比实验。这些攻击者模型分别模拟了不同类型的攻击者行为，包括基于用户基础攻击数据的简单攻击模型、基于特征攻击数据的高级攻击模型、结合多种特征和行为数据的复杂攻击模型以及模仿真实用户行为的混合攻击模型。通过这种多维度的实验设计，本文评估 RBDA 模型在不同攻击场景下的性能，实验的结果见表 4。

表 4 四类攻击者模型下 RBDA 的攻击检测与分级认证结果

攻击者	真阳性率	未触发风险处置	合法用户二次验证率中位数	人机验证率 Soft Verify	二次验证率 Hard Verify	驳回率
朴素攻击者	99.21%	0.79%	18.90%	9.92%	29.76%	59.53%
VPN 攻击者	98.55%	1.45%	同一基线	19.71%	44.35%	34.49%
定向攻击者	98.12%	1.88%	同一基线	44.15%	44.15%	9.82%
超级定向攻击者	82.85%	17.15%	同一基线	58.00%	20.71%	4.14%

表 4 展示了 RBDA 在四类不同攻击能力模型下的攻击拦截率及分级认证结果。随着攻击者对合法用户环境和行为特征的掌握程度不断增加，RBDA 的总体攻击拦截率由朴素攻击者场景下的 99.21%逐步下降至超级定向攻击者场景下的 82.85%。同时，认证处置结果呈现明显的风险等级迁移。对于朴素攻击者，由于 IP 地址、登录位置、设备及浏览器等多个高权重特征同时偏离合法用户历史分布，被检测攻击主要集中于高风险和危险等级，其中 59.53%的攻击被直接阻断，29.76%进入高强度身份验证。

VPN 攻击者能够部分模拟用户所在区域及常见用户代理，其 Critical 风险比例下降，Hard Verify 成为主要处置方式，占 44.35%。进一步地，定向攻击者能够掌握用户的位置、设备和浏览器信息，其环境特征与合法用户高度重合，因此系统更多依赖短期行为一致性与个体历史稳定度识别异常，Soft Verify 和 Hard Verify 分别占 44.15%。在超级定向攻击场景下，攻击者能够高度模拟合法用户的环境信息，RBDA 主要依据周期行为偏差进行识别，因此被检测攻击主要落入 Medium 风险区间，58.00%的攻击触发轻量验证，而直接阻断比例仅为 4.14%。上述结果表明，RBDA 能够依据攻击行为与合法用户画像之间的偏离程度自适应调整认证强度，而非对所有风险会话统一触发高成本二次认证。

RQ2：用户登录频率与历史行为窗口长度如何影响 RBA 的安全性和可用性？

为分析用户登录频率对二次验证开销的影响，本文按照相邻登录间隔将用户划分为五类：每日登录用户（<1 天）、几天登录用户（1－7 天）、每周登录用户（7－14 天）、几周登录用户（14－30 天）以及低

于每月一次登录的用户（>30 天）。在固定 TPR=0.995 和定向攻击者条件下，本文比较五类用户的二次验证开销，并采用非参数多组比较及 Dunn-Bonferroni 事后检验分析组间差异。为提高表格可读性，校正后(p>0.2)的比较结果以“–”表示，(p<0.05)的结果以粗体标记，具体结果见表 5。

数据留存与隐私合规成本。本文对比了在不同历史保留长度下的（1 周、1 个月、1 个季度、1 年）系统表现。当 RBDA 只保留最近一周的行为窗口时，TPR 相比保留完整历史（1 年）下降幅度小于 0.3 个百分点；同时，二次验证触发率基本不变（18.9%vs19.1%，差异不显著，p>0.05）。表明 RBDA 无需长期保存完整行为轨迹即可维持安全性和可用性。

表 5　不同登录频率用户二次验证开销的 Dunn-Bonferroni 校正后(p)值（TPR=0.995，定向攻击者）

登录频率	每天	几天	每周	几周	>30 天
每天	-	**0.0231**	-	**0.0098**	<0.0001
几天	**0.0231**	-	-	-	**0.0932**
每周	-	-	-	<0.0001	<0.0001
几周	**0.0098**	-	<0.0001	-	**0.0191**
>30 天	<0.0001	**0.0932**	<0.0001	**0.0191**	-

RQ3：RBDA 与其他算法相比表现如何？

本文模拟四类攻击者：朴素攻击、VPN 攻击、定向攻击、超级定向攻击。比较对象包括传统静态 RBA（固定特征权重+固定阈值）、机器学习式 RBA 阈值优化方案，以及本文的 RBDA。

为保证比较公平，本文在相同的用户登录数据、特征集合、训练/测试划分及攻击注入位置下重新实现 Wiefling 等人[20]提出的三种动态访问阈值方法 Linear、Poly 和 Hybrid。Linear 使用线性回归拟合合法用户风险评分随登录历史长度的变化；Poly 使用多项式回归拟合初期快速下降的风险评分趋势；Hybrid 在每个历史长度上选择 Linear 与 Poly 生成的较低有效访问阈值。对于三种基线和 RBDA，均使用相同的四类攻击样本计算 TPR，从而避免由数据集和攻击条件不同造成的直接比较偏差。

表 6 给出了在统一数据集、攻击样本及训练/测试划分下，Linear、Poly、Hybrid 与 RBDA 在四类攻击场景中的 TPR。对于朴素攻击者，Linear、Poly、Hybrid 和 RBDA 的 TPR 分别为 99.05%、99.16%、99.16%和 99.21%；对于 VPN 攻击者，分别为 97.97%、98.06%、98.30%和 98.55%；对于定向攻击者，分别为 96.92%、98.24%、98.28%和 98.12%，其中 RBDA 与 Poly 和 Hybrid 基本接近；对于超级定向攻击者，分别为 81.62%、81.75%、81.75%和 82.85%。总体来看，RBDA 在朴素、VPN 和超级定向攻击场景下取得最高 TPR，而在定向攻击场景下与 Poly 和 Hybrid 保持相近水平。

总体来看，RBDA 在多个攻击者模型下的表现优于或接近于机器学习模型，特别是在对抗复杂攻击者时显示了更强的抵御能力。因此，基于风险的动态身份认证 RBDA 相对于大多数基于风险的身份认证方法有较大优势，并更具有指导意义和实践价值。

表 6　四种不同基于风险的身份认证模型生成的动态访问阈值 TPR

	TPR			
攻击者模型	Linear	Poly	Hybrid	RBDA[本文]
朴素攻击者	**0.9905**	**0.9916**	0.9916	0.9921
VPN 攻击者	**0.9797**	**0.9806**	0.9830	0.9855
定向攻击者	**0.9692**	**0.9824**	0.9828	0.9812
超级定向攻击者	**0.8162**	**0.8175**	0.8175	0.8285

5.4 消融实验

为验证 RBDA 各模块设计的必要性与合理性，本文开展了系统性的消融实验。在相同数据与攻击注入设置下，逐一移除关键模块并对比性能变化。包含群体-个体特征重标定、历史感知的一致性修正、动态多级阈值决策及异常值清理。

图 4(a)结果表明，移除任一模块均会削弱系统的安全–可用性。其中，去除**多级阈值决策**使合法用户二次验证触发率由 18.9%上升至约 20%+（p<0.01），而攻击拦截率未获得对应提升，验证了分级响应在降低用户干扰方面的有效性。去除**时间一致性修正**显著增大风险评分波动，误触发率上升约 4 个百分点，说明稳定性建模对抑制偶发异常至关重要。去除**群体–个体特征重标定**在跨时区与异质用户场景下表现下降更为明显，表明群体先验对个体风险评估具有重要约束作用。未进行异常值清理会抬高动态阈值并劣化整体性能，体现了鲁棒性处理的必要性。

此外，不同历史窗口的消融结果显示，仅保留 1 周登录历史即可在攻击拦截率损失<0.3%的情况下维持稳定的二次验证触发率（18.9%–19.1%，p>0.05），同时显著降低数据留存成本，验证了 RBDA 对隐私最小化要求的适配性。我们通过朴素攻击者和 VPN 攻击者的攻击效力下，周期相关性证据（L=7 峰值）与边际收益递减对比（3→7 提升显著、7→30 提升趋

缓）共同确定 Δt=7 天，为满足隐私最小化的同时保持接近长窗口的攻击触发能力。在朴素攻击者和 VPN 攻击者场景下，一周窗口已能获得接近长历史窗口的检测性能；对于定向和超级定向攻击者，更长的历史窗口带来进一步收益有限。

5.5 特征权重分配影响分析

为了充分探讨在不同情况下，本文的算法受模型参数的影响状况，分别讨论了（1）如何选取影响用户风险身份认证的相关特征；（2）如何分配与用户身份认证相关的特征权重参数。

首先，在现代网络安全环境中，影响账户触发风险身份认证的特征种类繁多，通常包括但不限于以下几方面：IP 地址（IP address），用于识别用户的网络位置，当 IP 地址异常时，如来自未曾使用的地理区域，可能触发额外的身份认证；浏览器（Browser），用户使用的浏览器类型和版本，当浏览器信息与账户的历史使用模式不符时，可能被认为是潜在的安全风险；屏幕分辨率（Screen resolution），用户设备的屏幕分辨率，当屏幕分辨率变化时，可能表示用户在不同设备上登录，从而触发安全警告；操作系统（Operating system），用户设备的操作系统类型和版本，当操作系统变化时，可能指示用户更换了设备或发生了其他潜在风险；登录时间（Login time），用户的登录时间，在非正常时间段登录可能引发安全警告；登录地址（Login location），用户的地理位置，当登录位置与常用位置显著不同，可能触发额外的验证步骤。此外，还有用户行为模式（Behavioral patterns），包括用户的鼠标点击、键盘输入节奏等，当行为模式与历史行为显著不同，可能被识别为风险；设备指纹（Device fingerprinting），综合设备的各种硬件和软件特征，用于唯一识别设备，当设备指纹发生变化，可能被视为风险信号；多重因素认证状态，当用户未使用 MFA 或 MFA 状态异常时，可能触发风险认证；登录设备的安全状态，包括操作系统的安全补丁、杀毒软件的安装状态等，未安装安全补丁或未运行安全软件的设备可能被视为风险；用户账户历史活动（Account history），包括交易历史、账户设置变更等，当用户账户的行为模式突然发生显著变化时，可能触发认证。

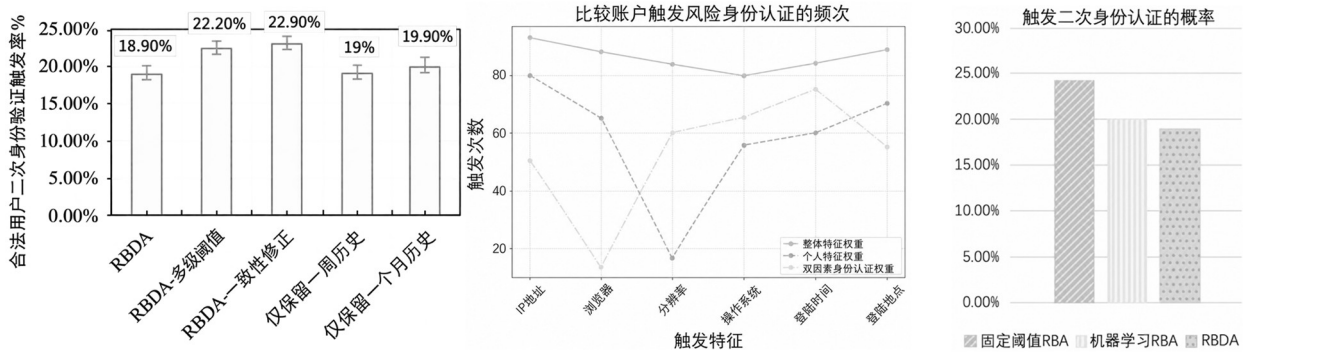
图 5 展示了在众多可能的风险特征中，不同特征在触发风险身份认证中的相对重要程度。本文选择“IP 地址”“浏览器”“屏幕分辨率”“操作系统”“登录时间”和“登录地址”作为重点研究对象，并非假设所有特征均与时区强相关，而是基于其在统计可区分性、工程可获得性以及区域行为差异等方面的综合考量。从特征属性来看，上述特征可大致分为三类。首先，登录时间与登录地址属于强时区相关特征，不同时区或区域用户在日常活跃时间段以及地理位置分布上存在显著差异；其次，IP 地址、浏览器与操作系统等特征虽然不直接由时区决定，但在不同国家或区域的用户群体中呈现出稳定的统计分布差异，体现出明显的区域相关性；最后，屏幕分辨率等设备特征更多反映终端生态差异，其与时区的关联性相对较弱，但在群体层面仍具有一定的判别价值。

因此，群体–个体特征重标定模块并不要求所有特征均与时区存在直接因果关系，而是将时区或区域作为群体划分的统计维度，用于刻画不同用户群体在特征分布上的系统性差异，从而为特征权重的自适应调整提供参考基准。该过程作用于特征权重层，回答“在当前区域条件下，哪些特征更具判别力”的问题。

在此基础上，短期行为一致性建模模块进一步针对单一用户展开，仅利用一周尺度的行为窗口评估当前登录是否偏离其近期行为节律，用于修正风险评分而非重新设定特征权重。两者在建模对象与作用层级上相互独立、功能互补，并不存在重复建模。

在中国主流站点场景（112 个网站，覆盖社交、政务、交易和旅游）下，本文对模拟“日常活跃用户”（每天或几乎每天登录）进行了评估。实验结果表明，在传统 RBA（全局固定阈值）下，合法用户的二次验证中位触发率为 24.3%；在机器学习式 RBA 阈值优化方案下，该值降至 20.1%；而在 RBDA 框架下进一步降低至 18.9%。相较于固定阈值 RBA，RBDA 使活跃用户的额外验证频率中位数下降约 22%（$p<0.01$），且未对攻击拦截率产生显著影响，表明其在安全性与可用性之间实现了更优平衡。

合法用户被要求二次验证的频率。在我们对 112 个站点的触发实验中，观察到 2023 年的数据，超过 70%的互联网用户在至少一个网站或应用程序上启用了多因素认证，其中以短信验证码为主要形式。然而，由于多因素认证的普及率相对较低，很多网站依然依赖用户名口令登录方式，因此上述六种特征（“IP 地址”、“浏览器”、“屏幕分辨率”、“操作系统”、“登录时间”和“登录地址”）在现阶段的身份认证过程中扮演着至关重要的角色。



(a) 不同模块对二次验证触发率的影响 (b) 整体-个人-双因素身份认证 (c) 合法用户不同 RBA 条件二次验证概率

图 4 RBDA 在特征敏感度、模块消融和二次验证开销方面的实验结果

在中国，针对用户登录的安全措施主要集中在以下几个方面：IP 地址，超过 80%的安全系统会重点关注 IP 地址的变化，特别是在跨地域登录时，该特征尤为重要；浏览器和操作系统，这些特征的匹配和历史数据比对是中国大部分电商和金融服务网站的标准安全措施；登录时间，非正常时间的登录尝试会被标记为高风险行为，特别是在金融交易类网站上。综合来看，这些特征的权重决定了账户是否会触发额外的身份验证，从而在平衡安全性和用户体验之间发挥着重要作用。本文对中国主流站点场景（112 个网站，覆盖社交、政务、交易、旅游）中“日常活跃用户”（每天或几乎每天登录）做了评估：传统 RBA（全局固定阈值）下，合法用户的二次验证中位触发率为 24.3%。机器学习式 RBA 阈值优化下，该值降到 20.1%。RBDA 下，该值进一步降到 18.9%。即相比固定阈值 RBA，RBDA 可让活跃用户的额外验证频率中位数下降约 22%（从 24.3%降到 18.9%）。这一差异在显著性检验中呈现 $p<0.01$。RBDA 在不牺牲拦截率的前提下，显著减少了对正常用户的打扰。

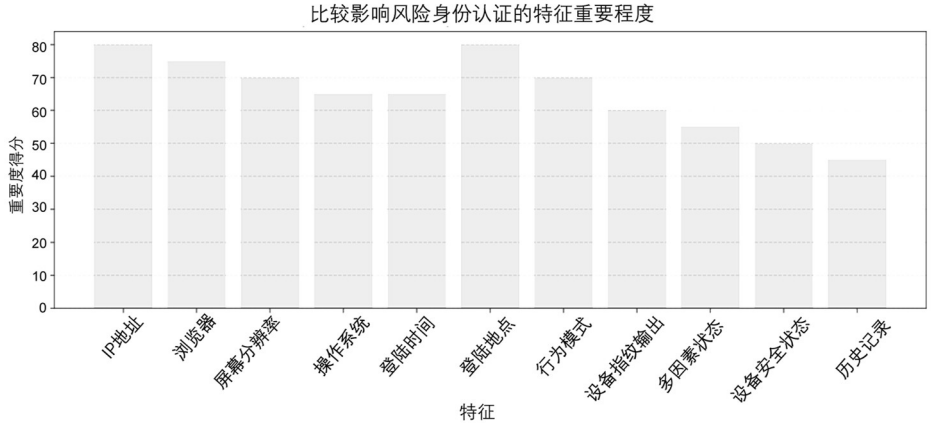


图 5 调研中 112 个受测网站中不同特征触发额外身份验证的比例

图 4(b)展示了在不同触发特征下，整体特征权重、用户个人特征权重以及双因素/多因素认证策略的认证触发次数。具体而言，整体特征权重参数（浅粉色曲线）在各个特征上的触发次数相对均衡，尤其是在“登录地址”（70 次）和“IP 地址”（80 次）上，触发次数较高，表明这些特征在整体用户群体中的安全性检测中起到了重要作用。相比之下，用户个人特征权重参数（浅绿色曲线）在“登录时间”（75 次）和“操作系统”（65 次）上触发次数更高，表明这些特征在个性化安全策略中更加敏感，能够有效识别出与个人行为模式不符的异常活动。此外，多因素身份认证（MFA）（浅紫色曲线）在所有特征上均表现出较高的触发次数，尤其在“IP 地址”（95 次）和“浏览器”（90 次）特征上体现尤为明显。这说明 MFA 能够在这些关键特征发生变化时更为有效地检测潜在威胁，显著增强了账户的安全性。整体来看，MFA 在各个特征下均显示出优于其他认证策略的安全防护能力，而个人特征权重则显示了在特定领域（如登录时间和操作系统）上更为突出的敏感性，这为基于特征的动态身份认证策略提供了数据支持。

合法登录被判为高风险的比例。这里本文用“高风险误报率”表示：合法登录被直接判定为高风险并

触发强校验（短信验证码/多因素验证）的概率。固定阈值 RBA 由于使用统一特征权重和统一阈值，高风险误报率在本文的站点集合中约为 7.4%。当引入仅“机器学习式阈值回归”时，可降到 5.9%。RBDA 引入“群体-个体时区对齐重标定+短期行为一致性”后，误报率进一步降到 4.8%。即 RBDA 最主要的收益在于：它不会简单地把“跟平常不太一样”直接当成“高风险”，因为它会判断这种差异到底是“个体自己的正常节律”还是“真正异常”。这是由于 RBDA 在计算基础风险分数时，先对齐同一时区群体分布 G_n 与目标用户近期分布 H_u，得到特征偏差 $\Delta group$ 与 $\Delta user$，然后按周级窗口评估一致性并进行动态阈值映射。

动态多级阈值 vs 固定挑战。固定阈值 RBA 往往一旦“判高风险”，就统一要求强挑战（例如短信验证码）。RBDA 不是这么做：它用动态多级安全阈值，将风险区间映射到不同强度的附加验证。在 RBDA 中，被判为“中等风险”的会话中，约 62%只需要轻量提醒或人机验证（而不是直接触发短信/多因素认证）。RBA（Linear）方案里，同等风险分数的会话有 91%都会直接进入高强度校验（短信验证码或 MFA），因此用户体验更差。RBDA 不只是“触发/不触发”，它能以更细粒度的方式把大部分边界情况“降级处理”。

6 结束语

基于风险的身份认证方法旨在解决单一口令认证带来的安全问题，同时缓解多因素身份认证中对外界设备的依赖，提升系统的可用性。本文提出了一种基于风险的动态身份认证方法，针对现有方法在学习和适配用户登录环境特征变化、拟合用户行为特征分布等方面的不足，进行了创新性改进。通过分析用户短期行为的循环特征，并结合账户风险触发的共性与差异性特征，本文进行了时间相关的行为特征分析。进一步地，通过引入用户行为的时间特征及其条件概率分布，验证了所提模型在特征抽象和身份认证中的有效性。实验结果表明，与固定阈值 RBA 和基于机器学习的 RBA 方法相比，本文提出的基于风险的动态身份认证方法不仅优化了身份认证的特征权重分配和安全阈值的动态学习能力，还显著提升了多因素身份认证的效率和性能。

致 谢 衷心感谢《计算机学报》编辑部老师和各位评审专家花费宝贵的时间和精力对本论文进行审阅！

参考文献

[1]邹云开,汪定. 口令猜测研究进展. 密码学报, 2024, 11(1): 67-100.

[2] Wang D, Zhang Z, Wang P, Yan J, Huang X. Targeted online password guessing: An underestimated threat//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. Vienna, Austria, 2016: 1242-1254.

[3] Duan M, Xu M, Zhang S, Han W. MoPE: A Mixture of Password Experts for Improving Password Guessing//Proceedings of the 2026 IEEE Symposium on Security and Privacy, 2026: 3017-3036.

[4] Li Z, Wang D. Targeted Password Guessing Using k-Nearest Neighbors. //Proceedings of the 33rd Network and Distributed System Security Symposium, 2026.

[5] K. Viezelyte, Juggling Security: How Many Passwords Does the Average Person Have in 2026?, May 2026, https://nordpass.com/blog/how-many-passwords-does-average-person-have/.

[6] Yunkai Zou, Maoxiang An, Ding Wang. Password Guessing Using Large Language Models. //Proceedings of the 34th USENIX Security Symposium , 2025: 7799-7818.

[7] Inglesant P G, Sasse M A. The true cost of unusable password policies: Password use in the wild//Proceedings of the 28th International Conference on Human Factors in Computing Systems. Atlanta, USA, 2010: 383-392.

[8] Shay R, Komanduri S, Durity A L, Huh P, Mazurek M L, Segreti S M, Ur B, Bauer L, Christin N, Cranor L F. Designing password policies for strength and usability. ACM Transactions on Information and System Security, 2016, 18(4): 1-34.

[9] Wang D, Shan X, Dong Q, Shen Y, Jia C. No single silver bullet: Measuring the accuracy of password strength meters//Proceedings of the 32nd USENIX Security Symposium. Anaheim, USA, 2023: 947-964.

[10] Bonneau J, Schechter S. Towards reliable storage of 56-bit secrets in human memory//Proceedings of the 23rd USENIX Security Symposium, 2014: 607-623.

[11] Woods N, Siponen M. Too many passwords? How understanding our memory can increase password memorability. International Journal of Human-Computer Studies, 2018, 111: 36-48.

[12] Wang D, Wang P. Two birds with one stone: Two-factor authentication with security beyond conventional bound. IEEE Transactions on Dependable and Secure Computing, 2016, 15(4): 708-722.

[13] Ometov A, Bezzateev S, Mäkitalo N, Andreev S, Mikkonen T, Koucheryavy Y. Multi-factor authentication: A survey. Cryptography, 2018, 2(1): 1.

[14] Wiefling S, Lo Iacono L, Dürmuth M. Is this really you? An empirical study on risk-based authentication applied in the wild//Proceedings of the 34th IFIP International Conference on ICT Systems Security and Privacy Protection. 2019: 134-148.

[15] Wiefling S, Patil T, Dürmuth M, Lo Iacono L. Evaluation of risk-based re-authentication methods//Proceedings of the 35th IFIP International Conference on ICT Systems Security and Privacy Protection. Maribor, Slovenia, 2020: 280-294.

[16] Xu Bing-Quan, Zhang Yuan, Yang Min. GrantDroid: A method supporting

in-context permission granting for Android. Computer Applications and Software, 2014, 31(8): 232-236 (in Chinese).
(徐冰泉, 张源, 杨珉. GrantDroid: 一种支持 Android 权限即时授予的方法. 计算机应用与软件, 2014, 31(8): 232-236.)

[17] Hu Ya-Lan, Zhang Yan. Common static password risks and protection suggestions in graded protection evaluation. Cyberspace Security, 2018, 9(4): 59-61 (in Chinese).
(胡亚兰, 张艳. 等级保护测评过程中常见的静态密码风险及防护建议. 网络空间安全, 2018, 9(4): 59-61.)

[18] Wiefling S, Dürmuth M, Lo Iacono L. More than just good passwords? A study on usability and security perceptions of risk-based authentication//Proceedings of the 36th Annual Computer Security Applications Conference. Austin, 2020: 203-218.

[19] Wiefling S, Dürmuth M, Lo Iacono L. What's in score for website users: A data-driven long-term study on risk-based authentication characteristics//Proceedings of the 25th International Conference on Financial Cryptography and Data Security. Berlin, Heidelberg, 2021: 361-381.

[20] Wiefling S, Jørgensen P R, Thunem S, Lo Iacono L. Pump up password security! Evaluating and enhancing risk-based authentication on a real-world large-scale online service. ACM Transactions on Privacy and Security, 2022, 26(1): 1-36.

[21] Freeman D, Jain S, Dürmuth M, Biggio B, Giacinto G. Who are you? A statistical approach to measuring user authenticity//Proceedings of the 23rd Annual Network and Distributed System Security Symposium. San Diego, USA, 2016: 1-15.

[22] Das A, Bonneau J, Caesar M, Borisov N, Wang X. The tangled web of password reuse//Proceedings of the 21st Annual Network and Distributed System Security Symposium. San Diego, USA, 2014: 23-26.

[23] Gavazzi A, Williams R, Kirda E, Lu L, King A, Davis A, Leek T. A study of multi-factor and risk-based authentication availability//Proceedings of the 32nd USENIX Security Symposium. Anaheim, USA, 2023: 2043-2060.

[24] Akhtar N, ul Haq F. Real time online banking fraud detection using location information//Proceedings of the International Conference on Computational Intelligence and Information Technology. Berlin, Heidelberg, 2011: 770-772.

[25] Cser A, Maler E. The Forrester Wave: Risk-Based Authentication [R], Q1 2012.

[26] Golan L, Orad A, Bennett N, inventors; EMC Corp, assignee. System and method for risk based authentication [P]. United States Patent US8572391, 2013.

[27] Hurkała A, Hurkała J. Architecture of context-risk-aware authentication system for web environments//Proceedings of the Third International Conference on Informatics Engineering and Information Science. Lodz, Poland, 2014: 219-228.

[28] Spooren J, Preuveneers D, Joosen W. Mobile device fingerprinting considered harmful for risk-based authentication//Proceedings of the Eighth European Workshop on System Security. Bordeaux, France, 2015: 6:1-6:6.

[29] Steinegger R H, Deckers D, Giessler P, Abeck S. Risk-based authenticator for web applications//Proceedings of the 21st European Conference on Pattern Languages of Programs. Kaufbeuren, Germany, 2016: 1-11.

[30] State of California. California Consumer Privacy Act. Assembly Bill No. 375, 2018. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

[31] European Union. General Data Protection Regulation (GDPR). Regulation (EU) 2016/679, May 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[32] Rivera E, Tengana L, Solano J, Castelblanco A, López C, Ochoa M. Risk-based authentication based on network latency profiling//Proceedings of the 13th ACM Workshop on Artificial Intelligence and Security. Virtual Event, USA, 2020: 105-115.

[33] Dasu L S, Dhamija M, Dishitha G, Vivekanandan A, Sarasvathi V. Defending against identity threats using risk-based authentication. Cybernetics and Information Technologies, 2023, 23(2): 105-123.

[34] FireHOL. FireHOL IP Lists: All Cybercrime IP Feeds [EB/OL]. 2021, https://iplists.firehol.org/.

[35] Wei T, Wang D, Li Y, Wang Y. "Who is trying to access my account?" Exploring user perceptions and reactions to risk-based authentication notifications//Proceedings of the 32nd Network and Distributed System Security Symposium. San Diego, USA, 2025: 1-27.

[36] Wiefling S. Usability, security, and privacy of risk-based authentication [Ph.D. dissertation]. Ruhr University Bochum, Bochum, Germany, 2023.



WEI Tong-Xin holds a PhD in Computer Technology. Her research focuses on usable security and privacy in authentication systems.



WANG Ding, Ph.D., is a professor and a member of the China Computer Federation (CCF). His research focuses on secret key security and identity authentication.

Background

Risk-Based Authentication (RBA) evaluates contextual and behavioral attributes to dynamically assess login risks and trigger additional verification only when anomalies are detected, thereby enhancing security while reducing user disruption. However, most existing RBA frameworks are derived from foreign designs and rely on proprietary risk evaluation algorithms, which conflict with the goal of achieving autonomous and controllable cyberspace security. Once these foreign algorithms contain backdoors or undisclosed vulnerabilities, attackers can exploit them to bypass risk detection or extract sensitive behavioral information, posing severe security threats to individuals, enterprises, and even national digital infrastructure.

In particular, RBA systems are widely adopted in large-scale online services because of their flexible deployment and minimal interference with user experience. Yet, researchers have shown that conventional RBA often relies on static feature weighting and globally fixed risk thresholds, which fails to capture individual behavioral dynamics and regional variations. To mitigate false alarms, some systems attempt to tune these parameters manually, but such static adjustments cannot adapt to evolving user behaviors or contextual shifts. Moreover, most existing RBA frameworks assume that behavioral profiles remain consistent and complete over time, which is rarely true in practice. Once a user changes device, location, or network environment, these systems tend to misclassify legitimate logins as high-risk events, triggering secondary verification. Although several adaptive or machine-learning-based risk models have been proposed, they require long-term behavioral storage and centralized data aggregation. This increases privacy exposure and introduces compliance concerns under regulations (e.g., GDPR and CCPA), thereby limiting their deployability in diverse service and regional contexts.

This paper proposes a Risk-Based Dynamic Authentication (RBDA) framework to enhance the adaptability and deployability in wild. On the one hand, it addresses the limitations of static feature weighting and global thresholds by introducing a group–individual behavioral recalibration mechanism, which aligns regional user behavior distributions with individual login histories to dynamically adjust feature importance during risk evaluation. On the other hand, it mitigates privacy and compliance challenges by employing a periodic short-term behavioral consistency model, which evaluates login legitimacy within week-scale temporal windows, thereby eliminating the need for long-term behavioral retention while preserving detection accuracy. Furthermore, RBDA incorporates a history-aware multi-level threshold decision mechanism that adaptively generates authentication thresholds according to user behavioral stability and assigns corresponding secondary verification strategies. This design enables fine-grained balancing between security, usability, and privacy protection, offering a practical pathway toward deployable, privacy-preserving RBA for modern online service ecosystems.

This work is supported by the National Natural Science Foundation of China under Grant 62222208 and by the Youth Project (Category A) of the Tianjin Natural Science Foundation, Tianjin, China, under Grant 25JCJQJC00230.