

大数据环境下用户口令认证风险分析及对策

付永贵^{1,2} 朱建明¹

(中央财经大学信息学院 北京 100081)¹ (山西财经大学信息管理学院 太原 030031)²

摘要 认证是信息安全的基本服务之一,口令认证是最常用的认证方法,但是目前用户口令设置存在许多隐患和风险。在分析目前用户口令设置存在的问题的基础上,提出了大数据环境下用户口令防护的攻防博弈模型,分析指出攻击者借助大数据分析技术能提高对用户口令的破译能力,而用户要想确保其安全性或更进一步降低其风险性则需要设置更有效的口令,使用身份交叉认证技术或动态跟踪用户访问信息系统行为的技术,降低大数据分析成本。提出相应的对策并使用用户数据画像思想建立大数据环境下信息系统用户身份交叉认证模型,通过模拟实验对模型的有效性进行验证。

关键词 大数据,口令认证,风险分析,数据画像,身份交叉认证模型

中图分类号 TP309 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.6.032

Risk Analysis and Countermeasure for User Password Authentication in Big Data Environment

FU Yong-gui^{1,2} ZHU Jian-ming¹

(School of Information, Central University of Finance and Economics, Beijing 100081, China)¹

(School of Information Management, Shanxi University of Finance and Economics, Taiyuan 030031, China)²

Abstract Authentication is one of the basic service for information security, and password authentication is the most common authentication method, but currently there is much risk in setting user password. On the basis of analyzing the current setting user password problem, we presented user password protection's offensive-defensive game model in big data environment, and pointed attacker could improve ability for deciphering user password with big data analysis technology. However, for user to ensure security or reduce risk, more effective password, identity cross-certification technology or dynamic track user access information system behavior technology, as well as lower big data analysis cost are needed. Countermeasure was presented and user data portrait thinking was used to establish information system user identity cross-certification model in big data environment. Validity of model was verified through simulation experiment.

Keywords Big data, Password authentication, Risk analysis, Data portrait, Identity cross-certification model

1 引言

目前,口令认证仍然是用户确保信息系统安全的一种主要手段,是对用户身份进行认证的主要方式之一,比如用户电子邮箱、微信、QQ、电子商务交易系统、日常事务管理信息系统都是通过口令认证来实现其系统安全的。因此,用户口令设置的方式也直接决定着信息系统使用的安全性。2011年12月CSDN网站遭到黑客攻击,600万用户明码电子邮箱的用户名、口令、邮件地址资料泄露,对其资料进行统计分析,发现用户邮箱口令具有以下特征:

(1)连续或者相同的数字串、字母串以及数字串与字母串的组合,比如:“12345678”、“123456789”、“11111111”、“qw-erasdf”、“abc1122334455”等;

(2)英文词汇、词汇的组合、习惯用语以及词汇与特征数字串的组合,比如:“pageuppageup”、“safecopy123”、“happy2000”等;

(3)手机号,比如:“136****407”、“138****084”、

“139****541”等;

(4)口令与用户名、邮箱名相同或相近,比如:用户名为“liuyu6645”,口令为“liuyu6645”;用户名为“a156694629”,口令为“156694629”;邮箱名为“zhang8699078@163.com”,口令为“zhang8699078”等;

(5)出生日期,比如:“19880916”、“19860706”、“19970701”等;

(6)姓名及其与出生日期、连续字符串等的组合,比如:“yeyanjuan1029”、“renjiandao”、“liuqing111”等。

经统计,这些常用的数字串、字母串、词汇、习惯用语、具有身份特征的字符及其组合等构成的邮箱口令占总口令数量的半数以上。调查发现,很多用户在自己注册使用的不同信息系统中经常使用相同或相近的口令,这样一旦不法分子或者黑客破译了用户其中一个信息系统的口令以后,就有可能通过“撞库”访问用户其他信息系统。中国移动广东公司中山分公司网络部副总经理李海健提出:在用户端,使用频率最高的前100个密码,覆盖了22%的用户^[1];CSDN董事长蒋涛坦

到稿日期:2015-01-03 返修日期:2015-04-17 本文受国家自然科学基金项目(61272398),国家社会科学基金重点项目(13AXW010)资助。

付永贵(1976—),男,博士生,讲师,主要研究方向为信息安全、信息管理、信息经济;朱建明(1965—),男,教授,博士生导师,主要研究方向为信息安全。

言包括 CSDN 在内的诸多大型网站都存在安全意识薄弱的问题:有 80% 的网站存在漏洞,60% 的安全类网站也存在漏洞,70% 的密码库可以被破解^[2]。另外,对山西财经大学 3000 名学生进行随机抽样填写调查问卷,证实了用户在设置邮箱口令时习惯使用常用的数字串、字母串、词汇、习惯用语、具有自己或者亲近社交人群身份特征的字符及其组合等构成的邮箱口令,而且这些口令占总调查记录的半数以上。

CSDN 泄露的这 600 万用户明文电子邮箱口令信息具有一般用户电子邮箱口令设置的特征,同时也说明用户在其他需要口令认证的信息系统设置口令也具有类似的特征。因此,目前用户在需要口令认证的信息系统中设置的口令具有很大的安全漏洞,这样的口令设置方式虽然有利于用户对信息系统口令的记忆,但也提高了不法分子或者黑客破译用户口令进入用户信息系统的风险。

学术界针对信息系统口令认证风险问题的研究已经很多,有些是针对用户借助身份特征、简单数字或常用字符串等设置系统口令造成风险进行分析研究的,比如 Jason Hong 等^[3]提出构建口令内容的具体办法,以确保口令的安全性;魏为民等^[4]以 CSDN 和 RenRen 网泄露的用户口令为例,分析了国内用户设置信息系统口令的特点,对其漏洞进行分析并提出了设置安全口令的一些思路,提出了系统安全防范的一些措施;WILLIAM CHESWICK^[5]提出用户在设置口令时不要让别人易于猜到,而且为了确保口令安全要经常更换设置的口令;Alan S. Brown 等^[6]分析了用户借助个性化特征设置口令以及不同系统重复使用同一口令的风险,并提出了自己的建议。Moshe Zviran 等^[7]探讨了用户建立信息系统口令的特征;这一类研究虽然指出了目前用户信息系统口令设置的漏洞并提出了改进思路,但其改进思路是以牺牲用户口令设置的便宜性及以增加用户记忆的难度为代价的。有些提出使用多重认证的方式来提高信息系统对用户身份认证的效率,比如 Mudhafar M. Al-Jarrah^[8]提出包含口令、系统生成密码、打字节奏在内的身份多重认证方法,但其认证思路比较僵化,缺乏灵活性,一定程度上增加了用户使用信息系统的复杂性。有些提出使用图形认证的方式来确保信息系统的安全,比如 John Charles Gyorffy 等^[9]提出使用嵌入式设备与图形口令来对用户身份进行认证;Thanh An Nguyen 等^[10]提出使用图形作为用户使用信息系统的口令。这类认证方式由于实现难度大、操作困难不适宜普通的信息系统使用。有些提出使用系统外装置来认证用户口令,比如 Mohammad Mannan 等^[11]提出使用独立的装置来输入用户的长期口令,但这种方式由于实现困难也不适于普通的信息系统使用。有些研究是针对用户口令结构进行分析的,比如邹静等^[12]提出通过对高概率的口令结构进行分析,作为破译用户口令的具体方法,但该文并未具体分析用户设置口令的非结构化问题和多重性问题。

大数据是移动互联网、云计算以及人们社会活动发展的产物,大数据研究机构 Gartner 将大数据定义为“大数据是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产”^[13]。大数据时代的来临,给人们获取信息提供了更广阔的途径,也为不法分子或黑客获取信息系统中的用户信息提供了更加广阔的数据源,为信息系统用户身份认证带来了新的挑战。大数据时代信息系统对用户身份认证的问题已经得到了学术界的重

视,一些学者也进行了相应的研究,比如董杨慧等^[14]提出大数据时代数据泄露的规模与范围在不断扩大,通过实例分析了泄露数据的特征并提出了相应的安全管理策略。但该文并未详细分析大数据造成数据泄露的原因,所提出的安全管理策略加大了信息系统管理的力度,造成了用户对信息系统管理的难度。

对国内外口令认证风险研究成果进行分析,目前的研究一方面以牺牲用户对口令记忆的便利性为代价,构建复杂、难记忆口令来提高信息系统口令认证的安全性,另一方面以牺牲用户对信息系统操作的便利性为代价,构建复杂的认证体系以提高信息系统口令认证的安全性。没有考虑常用信息系统对用户口令认证便利性的基本要求以及大数据环境下信息系统攻击者获取用户信息的广博性和攻击者与信息系统之间新的攻防博弈关系,虽然有个别研究者提出了大数据环境下口令认证风险提高的问题,但也未进行具体分析并提出有针对性的解决方案。本文将信息系统用户口令认证放到大数据环境下进行研究,分析了攻击者获取用户口令信息的全新手段,指出了攻击者与信息系统之间新的攻防博弈关系,针对大数据环境下信息获取的优势,提出了大数据环境下信息系统安全防范的新举措及用户口令认证模型,并通过模拟实验分析了模型的有效性。

2 大数据环境下用户口令防护的攻防博弈

目前有关攻防博弈的研究成果已经很多,比如朱建明等^[15]提出了在攻防双方信息不对称情况下具有学习机制的攻防演化博弈模型,黄启发等^[16]通过攻防博弈、共同防御博弈、联合攻击博弈研究了社交网络用户隐私的攻防博弈过程,但国内外针对大数据环境下用户口令防护的攻防博弈进行的研究尚为空白。

大数据环境下,信息系统攻击者可以借助大数据分析技术获取用户的姓名、出生日期、手机号码、ID 卡号、学历、毕业院校等结构化信息,也可以获取用户爱好、生活习惯、工作业绩等非结构化信息,还可以获取用户社交人群相应的结构化和非结构化信息。这些信息会在一定程度上反映用户对信息系统口令的设置,这样信息系统攻击者借助大数据分析技术破译用户口令的概率会大幅度提高,即大数据环境下单纯使用口令对信息系统进行防护的安全性会大大降低。当用户在自己使用的多个信息系统设置相同或者相近口令进行身份认证时,一旦其中一个信息系统口令被破译,其他信息系统口令被破译的风险也会迅速提高。因此,大数据对通过口令认证信息系统用户身份的安全性提出了新的挑战。

大数据对信息系统用户口令认证也带来了新的契机,信息系统可以借助用户大数据获取用户姓名、出生日期、学历、职业、电话号码、收入状况、爱好、社交人群、设备型号、打字频率、登录时间范围、登录时长等,在口令认证的基础上实现用户身份交叉认证。

为了方便研究,做以下定义或假设:

(1) 设攻击者与防御者都为理性人;攻击者的攻击能力指攻击者攻击意愿及攻击期望收益的水平,防御者的防御能力指防御者通过防御避免自身期望收益降低的水平;

(2) 设攻击者的攻击能力 A 受攻击成本 C_1 、获取防御者信息能力 f_1 、攻击成功收益 P_1 、攻击失败被捕获受到惩罚 Q_1 的影响;信息系统作为防御者的防御能力 B 受防御成本 C_2 、

防御者保护信息能力 f_2 、防御成功收益 P_2 、防御失败损失 Q_2 的影响；

(3) 设攻击成本对攻击能力造成影响系数值为 K_1 ，防御成本对防御能力造成影响系数值为 K_2 ；

(4) 设不使用大数据分析技术时攻击者攻击能力为 A_0 ，攻击成本为 C_{10} ，获取防御者信息能力为 f_{10} ；防御者防御能力为 B_0 ，防御成本为 C_{20} ，防御者保护信息能力为 f_{20} ；

设使用大数据分析技术时，攻击者攻击能力为 A' ，攻击成本为 C_{11} (C_{10} 及获取防御者大数据成本)，获取防御者信息能力为 f_{11} ；防御者防御能力为 B' ，防御成本为 C_{21} (C_{20} 及借助大数据分析技术防御攻击者成本)，防御者保护信息能力为 f_{21} ；

对攻击者与防御者在使用大数据分析技术前后进行比较， P_1, Q_1, P_2, Q_2 取值不发生变化， $C_{11} \geq C_{10}, C_{21} \geq C_{20}, f_{11} \geq f_{10}, f_{21} \geq f_{20}$ ；

(5) 设攻击者攻击成功概率函数值为 $F(A, B, \frac{A}{B})$ 。

因为攻击者攻击意愿及 $F(A, B, \frac{A}{B})$ 仅取决于 $\frac{A}{B}$ ，单纯受 A, B 取值的影响可以忽略不计，所以 $F(A, B, \frac{A}{B})$ 可以简化为 $F(\frac{A}{B})$ ；攻击者攻击意愿及 $F(\frac{A}{B})$ 是 $\frac{A}{B}$ 值的增函数。

则 $A' = A_0 + (f_{11} - f_{10}) - K_1(C_{11} - C_{10})$ ， $B' = B_0 + (f_{21} - f_{20}) - K_2(C_{21} - C_{20})$ ， $\frac{A'}{B'} = [A_0 + (f_{11} - f_{10}) - K_1(C_{11} - C_{10})] / [B_0 + (f_{21} - f_{20}) - K_2(C_{21} - C_{20})] = [(A_0 - f_{10} + K_1 C_{10}) + (f_{11} - K_1 C_{11})] / [(B_0 - f_{20} + K_2 C_{20}) + (f_{21} - K_2 C_{21})]$ 。 $K_1 C_{11}$ 随 f_{11} 增大而增大，但增大的绝对值较 f_{11} 小，否则攻击者会在 f_{11} 的某一取值以后放弃对 f_{11} 的增大。

当 A 增加时， B 也要相应增加，以确保 $\frac{A'}{B'} \leq \frac{A_0}{B_0}$ ，否则信息系统的的天性风险会增大，攻击者的攻击意愿及信息系统被攻击成功的概率函数值 $F(\frac{A}{B})$ 会增大。其解决的办法是 B_0 、 f_{21} 增大或者 $K_2 C_{21}$ 减小，即要提高防御者的基础防御技术，提高防御者借助大数据的防御攻击能力，或者通过降低信息系统大数据分析成本来实现信息系统的安全性。

因此，在大数据环境下，攻击者借助大数据分析技术可以提高对信息系统攻击的能力，而用户要想确保其使用信息系统的安全性或更进一步降低其风险性，则需要通过设置更有效的口令，使用身份交叉认证技术或动态跟踪用户访问信息系统行为的技术，降低大数据分析成本来实现。

3 大数据环境下用户口令认证对策

大数据环境下，用户的数据信息会更容易收集、提炼和分析，为完整地抽象出一个人的信息全貌提供了可能。当用户在信息系统注册时，通常需要填写自己的身份信息，比如姓名、出生日期、学历、职业、电话号码等，通过其他数据源可以进一步收集用户的其他身份信息，这些信息逐步修正、叠加就会形成用户的完整身份信息；同样从其他数据源还可以收集到用户的其他固定信息，比如收入状况、爱好、社交人群等信息，用户的身份信息及其相关的固定信息称做这一用户的静态信息。另外，用户的信息还包含与其相关的动态信息，比如

用户登录信息系统设备型号、打字频率、登录地域、登录时间范围、登录时长、操作行为信息、浏览内容信息、一定时期社会活动信息，甚至某一时期的情绪信息等，用户的静态信息及其动态信息构成了用户的信息全貌，即用户画像数据。

在信息系统对用户进行身份认证时，如果检测到不符合合法用户画像数据特征的用户数据（静态或动态信息），就需要对用户身份进行进一步认证。

攻击者也可以通过获取用户大数据对用户进行画像、破译用户口令、模仿用户行为，以实现对信息系统的攻击。因此信息系统要通过设置不利于攻击者破译的口令、制定大数据的保护措施、使用严密的用户行为监管技术及身份交叉认证技术来实现对信息系统的保护。

对于用户口令设置提出以下对策：

(1) 使用字母、数字和符号相结合的字符串作为用户口令；

(2) 避免使用常用的、易于记忆的字符串作为用户口令；

(3) 字符串长度要尽量较长；

(4) 为了便于记忆可以使用一些具有特性的字符串，但这些字符串必须是经过用户“加密”处理过的字符串；

(5) 避免使用用户的静态数据信息作为用户口令，如果使用也需要先进行“加密”处理；

(6) 尽量使用一些用户容易记忆而通过大数据分析技术难于收集的私密信息作为用户口令。

大数据环境下信息系统一方面要安全有效地保管用户大数据，防止用户大数据泄露造成对用户的损害，另一方面需要具备对用户数据画像的能力，对于通过口令进入系统但不符合用户数据画像特征的使用者，要对其实行身份交叉认证。

信息系统用户画像的特征数据分为定性数据和定量数据，对于有量级或者取值范围的数据要设定其量级或者取值范围；对于直接判断对错的数据，要确定其固定值，作为合法用户与非法用户进行比较是否一致的依据。用户的特征变量分为静态特征变量和动态特征变量，使用模糊综合评价法确定各特征变量对用户数据画像贡献的权重，为了便于表述，做以下定义和假定：

(1) 将表征用户特征的各静态特征变量定义为 $x_1, x_2, \dots, x_n$ ；表征用户特征的各动态特征变量定义为 $y_1, y_2, \dots, y_m$ ；

(2) 设各静态特征变量 $x_i (1 \leq i \leq n)$ 对用户数据画像贡献的权重为 w_i ，各动态特征变量 $y_k (1 \leq k \leq m)$ 对用户数据画像贡献的权重为 t_k ；且 $0 \leq w_i \leq 1, 0 \leq t_k \leq 1, \sum_{i=1}^n w_i + \sum_{k=1}^m t_k = 1$ ；

(3) 设 $L(x_1, x_2, \dots, x_n; y_1, y_2, \dots, y_m)$ 表示合法用户的数据画像函数，简记为 L ； $L'(x_1', x_2', \dots, x_n'; y_1', y_2', \dots, y_m')$ 表示合法用户以外的非法用户的行为数据函数，简记为 L' ；

(4) 对于有量级或者取值范围的数据，将 x_i 的量级或者取值范围设定为 G_i 个， y_k 的量级或者取值范围设定为 H_k 个；如果 x_i 与 x_i' 属于同一量级或者取值范围，则 $|x_i - x_i'| = 0$ ，否则， $|x_i - x_i'| = 1$ ，如果 y_k 与 y_k' 属于同一量级或者取值范围，则 $|y_k - y_k'| = 0$ ，否则， $|y_k - y_k'| = 1$ 。对于直接判断对错的数据，如果 x_i 与 x_i' 相同，则 $|x_i - x_i'| = 0$ ，否则， $|x_i - x_i'| = 1$ ，如果 y_k 与 y_k' 相同，则 $|y_k - y_k'| = 0$ ，否则， $|y_k - y_k'| = 1$ ；

(5) 设共有 S 个专家参与各特征变量对用户数据画像贡

献的权重初始评估,其中第 j 个专家($1 \leq j \leq S$)对各静态特征变量对用户数据画像贡献的初始评估值为 w_{ij} ,对各动态特征变量对用户数据画像贡献的初始评估值为 t_{kj} ,则 $w_i = \sum_{j=1}^S w_{ij} / S, t_k = \sum_{j=1}^S t_{kj} / S$ 。

用多元判别分析模型(也可以使用其他相关模型)表示非法用户与合法用户数据画像函数值的差异,其表达式设为:

$$\Delta L = |L' - L| = \sum_{i=1}^n (w_i |x_i - x_i'|) + \sum_{k=1}^m (t_k |y_k - y_k'|)$$

则这一数据画像函数值的差异可以作为判断用户是否是合法用户的依据。当一非法用户通过破译口令进入用户信息系统时,信息系统将根据非法用户的行为路径不断产生的画像数据与系统保存的合法用户的基础数据进行对比计算,并根据其差异值大小提示非法用户输入进一步的认证口令(对于同一变量对应的数据整个操作过程如果有差异只记录一次)。设 $\delta_1, \delta_2, \dots, \delta_q$ 为信息系统设置的 L' 与 L 差异临界值,当判别式 $\Delta L = |L' - L| \geq \delta_1$ 时,信息系统要求用户输入除进入信息系统输入口令外的信息系统内部第 1 层口令;当进一步根据非法用户行为路径判别式 $\Delta L = |L' - L| \geq \delta_2$ 时,信息系统将要求用户输入第 2 层口令;...;当判别式 $\Delta L = |L' - L| \geq \delta_q$ 时,信息系统将要求用户输入第 q 层口令,判别过程直至用户访问信息系统行为结束。这样通过用户行为数据画像和交叉认证就实现了对用户的身份认证。

大数据环境下信息系统用户口令认证流程如图 1 所示。

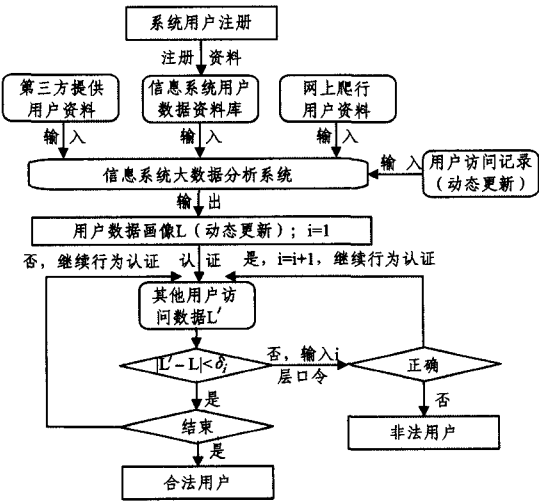


图 1 大数据环境下信息系统用户口令认证流程

4 模拟实验分析

以邮箱系统为例做模拟实验并进行分析,提出现行邮箱系统在未来功能扩充的基本思路。邮箱系统涉及用户的信息:用户在邮箱注册的基本身份信息、用户使用邮箱行为信息以及邮箱系统大数据分析系统借助大数据分析技术获得用户身份信息 and 行为信息,通过大数据分析技术获得的用户身份信息是对用户基本身份信息的修正、补充,用户使用邮箱的行为信息是对用户使用邮箱的特点、习惯、邮箱内容等的描述,通过大数据分析技术从其他数据源得到的用户行为信息是对用户个性特点、生活内容等的描述,这些信息作为用户的标签,形成对用户的数据画像,其将被邮箱系统大数据分析系统进行记录和安全保管。当非法用户盗取合法用户口令进入信

息系统进行操作时,信息系统将沿着非法用户的行为路线逐渐判断用户的操作行为与合法用户画像数据是否相符合(对于同一变量对应的数据,整个操作过程如果有差异只记录一次),将不相符合数据差异按权重计算相加后与邮箱系统预先设置的临界值进行比较并逐步完成相应的口令认证。

邮箱系统将用户的数据分为静态数据和动态数据,静态数据内容包括:用户注册基本身份数据(姓名、出生日期、手机号码)(SAD_1)及行为数据(收入状况(SBD_1)、爱好(SBD_2)、社交人群(SBD_3));动态数据内容包括:基本数据:设备型号(AAD_1)、登录邮箱地域(AAD_2)、打字频率(AAD_3)及行为数据(登录时间范围(ABD_1)、登录时长(ABD_2)、邮箱使用行为特点(ABD_3)、操作邮件学科分类(ABD_4))。

对邮箱系统用户的变量数据进行分析, SAD_1 、 AAD_1 、 AAD_2 变量数据只有对错两种结果,在用户进入系统后,只需判断其对错就可以了; SBD_1 、 SBD_2 、 SBD_3 、 AAD_3 、 ABD_1 、 ABD_2 、 ABD_3 、 ABD_4 变量数据属于有量级或者取值范围的数据,需要根据变量数据的量级或者取值范围来判断访问系统的用户是否是合法用户。经过 10 名 IT 专业人员讨论、分析、计算后,对变量数据量级或者取值范围的定义如表 1 所列,各变量对应权重值如表 2 所列。

表 1 变量数据量级或者取值范围定义表

名称	量级或者取值范围	编号	名称	量级或者取值范围	编号
SBD ₁	3000 元/月以下	1	AAD ₃	50 字/分钟以下	1
	3000~5000 元/月	2		50-100 字/分钟	2
	5000~10000 元/月	3		100 字/分钟以上	3
	10000~30000 元	4		6:00-12:00(上午)	1
	30000 元/月以上	5		12:00-14:00(中午)	2
SBD ₂	旅游	1	ABD ₁	14:00-20:00(下午)	3
	棋牌类	2		20:00-24:00(晚上)	4
	球类	3		00:00-6:00(夜间)	5
	舞蹈、音乐类	4	ABD ₂	3 分钟以下	1
	读书、科研	5		3-5 分钟	2
SBD ₃	其他类	6		5-10 分钟	3
	企业家(非 IT)	1		10 分钟以上	4
	企业职员(非 IT)	2	ABD ₃	操作某一分类邮件	1
	IT 业界高级人员	3		删除某一分类邮件	2
	军人	4		更改用户信息	3
	生产、运输人员	5		其他	4
SBD ₃	演艺人员	6	ABD ₄	经济管理类	1
	中小學生	7		高等教育、心理、体育类	2
	大中专学生	8		新闻、传播、图情类	3
	研究生	9		数学、计算机类	4
	教师、科研人员	10		政治、哲学类	5
其他	行政、党群人员	11	其他类	中小学教育类	6
	其他	12		其他类	7

表 2 邮箱系统大数据分析系统用户特征数据及其权重

SAD ₁	SBD ₁	SBD ₂	SBD ₃	AAD ₁	AAD ₂	AAD ₃	ABD ₁	ABD ₂	ABD ₃	ABD ₄
0.25	0.05	0.05	0.05	0.15	0.13	0.12	0.05	0.05	0.05	0.05

模拟实验由合法用户 A 给出其邮箱系统相关变量静态数据及动态数据的基础数据值,基础数据值有量级或取值范围部分如表 3 所列。

表 3 合法用户 A 各变量基础数据值(有量级或取值范围部分)

变量	SBD ₁	SBD ₂	SBD ₃	AAD ₃	ABD ₁	ABD ₂	ABD ₃	ABD ₄
基础数据编号	3	2	3,10	2	3,4	1	1	4

另外, SAD_1 = “王兵,1982 年 1 月 3 日生,138 * * * * * 55”, AAD_1 = “DELL V5470-4528S”, AAD_2 = “北京”。

从若干大学随机抽取 300 名学生进行问卷调查(学生处于假期离校期间,调查结果数据通过邮件发送),实验过程中,A 将自己的邮箱名“A@qq.com”及口令“*****”告知给 300 名其他学生。由这 300 名学生分别登录邮箱系统进行使用,将使用过程按变量记录相关的静态数据和动态数据行为数据值。

给出差异临界值 $\delta_1=0.3, \delta_2=0.5, \delta_3=0.8$ (在具体应用过程中也可以根据系统要求设置不同的 $\delta_1, \delta_2, \delta_3$ 值),对实验结果进行分析,90.7%的用户需要填写第 1 层用户口令(确认为非法用户,即 $\Delta L = |L' - L| \geq \delta_1 = 0.3$ 者),70.3%的用户需要填写第 2 层用户口令(即 $\Delta L = |L' - L| \geq \delta_2 = 0.5$ 者),11%的用户需要填写第 3 层用户口令(即 $\Delta L = |L' - L| \geq \delta_3 = 0.8$ 者)。这说明 90.7%的非法用户将被认证出来,即使有个别用户侥幸通过第 1 层口令,原来 70.3%的非法用户还需要进行第 2 层口令的认证,有些还需要进行第 3 层口令的认证,这种多层口令认证系统大大加强了信息系统对非法用户的防御力度。为了避免合法用户在使用信息系统时的非常规行为造成多级口令认证带来使用上的不便,信息系统在设置临界值 $\delta_1, \delta_2, \delta_3$ 时要进行前期调研,尽量使其合理,本文对临界值的设置仅服务于模拟实验,不对现实应用形成结论性指导。

图 2 表示这 300 名学生所记录的变量数据与合法用户 A 的基础变量数据对应不相符的数量比例(即各变量不相符人数/300)。

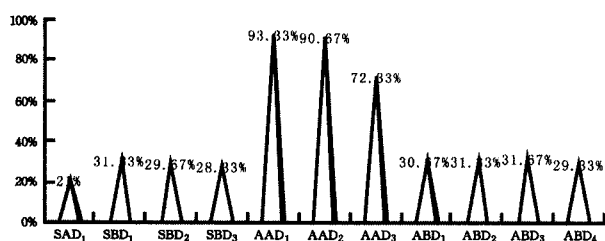


图 2 差异变量数据占比(差异变量人数/300)

从图 2 可以看出,与 A 的基础数据进行比较后,静态变量数据(SAD₁、SBD₁、SBD₂、SBD₃)中出现比较差异的数据占比较小,说明用户在访问邮箱时将更少对静态变量数据的操作记录;ABD₁、ABD₂、ABD₃、ABD₄ 中出现比较差异的数据占比也较小,而且这些变量的权重值也较小,说明这些变量数据对于检验非法用户行为具有不敏感性。AAD₁、AAD₂、AAD₃ 中出现比较差异的数据占比较大,权重值也较大,说明这些变量更能有效地认证用户是否是真实用户。因此,通过模拟实验的研究,以后在设置认证变量时要尽量使用认证效率更高的敏感变量,对于不敏感变量要重新分析定位,确实不必要的变量在认证模型中则应剔除。

整个实验属于模拟实验,对邮箱系统用户特征数据变量及其权重设置具有很大的主观性,所抽取的 300 名学生样本数量较小,身份特性也呈单一化,其定量化的数值结果不具备对现实应用的指导性,但通过该模拟实验的研究也可以很好地说明大数据分析技术在用户邮箱系统对用户进行身份交叉认证方面的价值。因此,在大数据分析技术迅速发展的时代,邮箱服务提供商需要顺应信息时代的发展,借助大数据分析技术不断完善邮箱系统对非法入侵用户的监测和合法用户信

息的管理。

对于其他口令认证信息系统也可以参照以上方法进行相应的研究,这里将不再赘述。

结束语 本文针对目前信息系统口令设置常规化、特征化造成安全性差的问题,对国内外口令认证研究进行了分析,指出了大数据环境下信息系统攻击者与信息系统之间新的博弈关系,提出了大数据环境下口令认证的相应对策并使用用户数据画像思想建立了信息系统用户身份交叉认证模型,最后通过模拟实验对模型进行了验证。目前由于大数据分析技术相应的理论研究和应用尚处于刚起步阶段,因此未能将系统实现,只进行了模拟实验分析,以后将继续进行这一领域应用方面的研究,建立模拟邮箱系统大数据分析系统并将其研究成果投入使用。

参考文献

- [1] 李海健. CSDN:互联网服务端近 80%密码库可破解[J]. 移动通信, 2012(Z1):112
Li H J. CSDN: nearly 80 percent of password databases can be cracked in internet server[J]. mobile communication, 2012(Z1): 112
- [2] 杨汛, 王琰. 泄密用户七成未改密码[N]. 北京日报, 2012-01-12(14)
Yang X, Wang L K. seven tenths leaked users unchanged password[N]. Beijing Daily, 2012-01-12(14)
- [3] Hong J, Reed D. Passwords Getting Painful, Computing Still Blissful[J]. Communications of the ACM, 2013, 56(3):10-11
- [4] 魏为民, 陈为召, 李红娇. 国内网络用户密码分析[J]. 上海电力学院学报, 2013, 29(6):584-587
Wei W M, Chen W Z, Li H J. domestic network users password analysis[J]. electronic college journal of shanghai, 2013, 29(6): 584-587
- [5] Cheswick W. Rethinking Passwords[J]. Communications of the ACM, 2013, 56(2):40-44
- [6] Brown A S, Bracken E, Zoccoli S, et al. generating and remembering passwords[J]. Applied Cognitive Psychology, 2004, 18(6): 641-651
- [7] Zviran M, Haga W J. password security: an empirical study[J]. Journal of Management Information Systems, 1999, 15(4): 161-185
- [8] Al-Jarrah M M. a multi-factor authentication scheme using key-stroke dynamics and two-part passwords[J]. International Journal of Academic Research, 2013, 5(3):98-102
- [9] Györfy J, Tappenden A, Miller J. Token-based graphical password authentication[J]. International Journal of Information Security, 2011, 10(6):321-336
- [10] Nguyen T A, Zeng Y. A vision based graphical password[J]. Journal of Integrated Design and Process Science, 2010, 14(2): 43-52
- [11] Mannan M, van Oorschot P C. Leveraging personal devices for stronger password authentication from untrusted computers[J]. Journal of Computer Security, 2011, 19(4):703-750
- [12] 邹静, 林东岱, 郝春辉. 一种基于结构划分概率的口令攻击方法[J]. 计算机学报, 2014, 37(5):1206-1214
Zou J, Lin D D, Hao C H. a password attack method based on

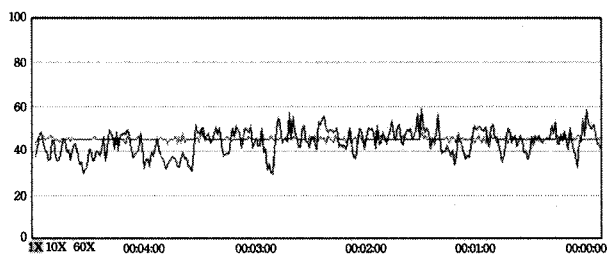
- structural division probability[J]. Chinese journal of computers, 2014,37(5):1206-1214
- [13] 徐迪威. 大数据与科技管理[J]. 科技管理研究, 2013(24):216
Xu D W. Big Data and Technology Management[J]. Technology Management Research, 2013(24):216
- [14] 董杨慧, 谢友宁. 大数据视野下的数据泄露与安全管理[J]. 情报杂志, 2014,33(11):154-158
Dong Y H, Xie Y N. Data Disclose and Secure Management in Gig Data Vision[J]. Information Magazine, 2014, 33(11): 154-158
- [15] 朱建明, 宋彪, 黄启发. 基于系统动力学的网络安全攻防演化博弈模型[J]. 通信学报, 2014,35(1):54-60
Zhu J M, Song B, Huang Q F. Evolution Game Model of Offense-defense for Network Security Based on System Dynamics[J]. Journal on Communications, 2014,35(1):54-60
- [16] 黄启发, 朱建明, 宋彪, 等. 社交网络用户隐私保护的博弈模型[J]. 计算机科学, 2014,41(10):184-189
Huang Q F, Zhu J M, Song B, et al. game model of user's privacy-preserving in social networks[J]. Computer Science, 2014,41(10):184-189

(上接第 114 页)

表 4 实时通信视频性能指标比较

方法	码率	帧率	CPU 占用	图像质量	画面连贯性
ABR	画面动态时候会超过较多, 码率统计曲线波动大	不稳定, 在 15~25fps 范围频繁跳动, 动态时帧率急剧下降	~50%	块效应少一些, 但细节欠清晰, 动态画面恢复到清晰的过程较 TMN8 慢	运动时流畅性下降, 且偶有停顿
TMN8	基本没有超过设定带宽, 码率统计曲线平滑	稳定在 24~25fps	相当 (~50%)	块效应略多, 细节清晰, 背景细腻, 恢复快	保持流畅, 无停顿现象

此外, 本文特别地用专业的流量测试软件 Netmon 统计了带宽的使用情况, 结果如图 3 所示。可以看出, TMN8 的码率波动曲线明显较 ABR 平滑, 基本上没有出现 ABR 那种瞬时码率上升很高的情形。众所周知, 在带宽受限且得不到 QoS 保障的公网上, 瞬时的突发性码流是引起网络丢包的一个最主要的因素, 因此, TMN8 特别适合于公网视频通信。



上下波动较大的线表示 ABR, 上下波动较小的线表示 TMN8

图 3 码率波动曲线

结束语 本文对 TMN8 模型在 H. 264 码率控制中的改进和应用展开研究。文章首先从实际应用环境和 TMN8 文档中描述的仿真环境的差异出发, 对 TMN8 模型提出 3 点完善措施, 包括帧级目标比特数计算、图像金字塔感知加权、分层次量化步长控制。每个改进措施都有不同的针对性: 引入“负”缓冲概念是为了在实际采样帧率达不到目标帧率的情况下, 更好地利用网络带宽资源; 采用金字塔加权是考虑到在交互式视频应用中, 人眼对位于图像中心的头肩部画面更敏感的缘故; 分层次控制量化步长则是为了更好地保障视频画面的流畅性, 因为从主观感受来看, 帧率高但不流畅的视频反而不如帧率稍低但流畅的视频的视觉感受好。针对 TMN8 在 H. 264 中应用面临的 RDO 运动估计和码率控制互为因果的矛盾, 本文提出了一种两遍运动估计的方案, 充分利用预运动估计阶段的附带信息来加速和优化编码过程。本文最终实现了一种基于 TMN8 模型的 H. 264 码率控制方法。离线和在线测试结果都表明本文方法显著改善了 H. 264 码率控制的

精度和实时视频通信的品质。由于 H. 265 采用了与 H. 264 一致的率失真编码框架, 本文技术有望被推广到 H. 265 的码率控制中, 下一步我们将在 H. 265 环境下开展测试验证。

参考文献

- [1] ISO/IEC JTC1/SC29/WG11/93-225b, MPEG-2 Test Model 5 [S]. Test Model Editing Committee, 1993
- [2] ITU-T Video Coding Experts Group. Video Codec Test Model, Near-Term, Version 8 [S]. Portland, 1997
- [3] Jordi R C, Lei S. Rate control in DCT video coding for low-delay communications [J]. IEEE Trans. on Circuits and Syst. for Video Technol., 1999,9(12):172-185
- [4] Lee H J, Chiang T H, Zhang Y Q. Scalable rate control for MPEG-4 video [J]. IEEE Trans. on Circuits and Syst. for Video Technol., 2000,10(6):878-894
- [5] JVT-G012-r1, Adaptive basic unit layer rate control for JVT [S]. Pattaya II, Thailand. 2003
- [6] ITU-T Recommendation H. 265 & ISO/IEC HEVC, High Efficiency Video Coding [S]. 2013
- [7] Choi H, Nam J, Yoo J, et al. JCTVC-H0213, Rate control based on unified RQ model for HEVC [S]. San José, 2012
- [8] Grois D, Hadar O. Complexity-aware adaptive bit-rate control with dynamic ROI pre-processing for scalable video coding [C]// IEEE Int. Conf. on Multimedia and Expo(ICME). 2011:1-4
- [9] Qu T S, Huang Y H, Chen H H. SSIM-based perceptual rate control for video coding [J]. IEEE Trans. on Circuits and Syst. for Video Technol., 2011,21(5):682-691
- [10] Wu Guan Lin, Fu Yu Jie, Huang Sheng, et al. Perceptual quality-regulable video coding system with region-based rate control scheme [J]. IEEE Trans. on Image Process, 2013,22(6):2247-2258
- [11] X264 Source Code[OL]. <http://developers.videolan.org/x264.html>