DOI:10.1145/2594445

神经科学与密码学交汇：通过密码原语抵御软磨硬泡式攻击

作者：Hristo Bojinov、Daniel Sanchez、Paul Reber、Dan Boneh 和 Patrick Lincoln

摘要

密码系统常常依赖于提供给用户的密钥的机密性。但在攻击者强制用户提供密钥时，许多方案都无法抵御胁迫式攻击。此类攻击称为*软磨硬泡式破译*，通常是攻破密码的最简单方式。本文介绍胁迫攻击的一种抵御方式，它利用来自认知心理学的*内隐学习概念*。内隐学习指的是一种在不知不觉中学习的方式。我们使用一款精心制作的计算机游戏，让用户通过内隐方式学习密码，而用户却对训练的密码没有任何明显或有意识的认知。虽然训练的密码可用于身份验证，但参与者无法被强迫提供此密码，因为他们根本不知道密码。我们利用 Amazon 的 Mechanical Turk 进行了一系列用户研究，确认了参与者在一段时间后仍可再次通过身份验证，但无法重新构建或明确识别训练的密码。

1. 引言

想象这样一个场景：某一安保措施严密的设施中运用了复杂的身份验证系统，只有知道密钥、持有硬件令牌并具备授权生物识别信息的人才能进入其中。门卫确保只有成功通过身份验证的人才可进入该设施。假设一个聪明的攻击者俘获到一名通过身份验证的用户。攻击者能够盗取该用户的硬件令牌，伪造其生物识别信息，借助橡胶管等武器迫使受害者泄露其密钥。此时，攻击者可以假冒受害者，攻破该设施中部署的造价昂贵的身份验证系统。

所谓的软磨硬泡式攻击一直是安全系统的天敌，也常常是攻破密码的最简单方式。[12] 原因在于，用户必须持有身份验证凭据才能通过身份验证，而这些凭据可以通过武力 [10] 或其他方式获取。

本文介绍一种可预防某类软磨硬泡式攻击的新方法，它利用来自认知心理学的*内隐学习* [2, 7] 概念。一般认为，内隐学习涉及大脑中称为*基底神经节*的部分，它支持通过重复执行任务来学习骑自行车或打高尔夫球等技能。以内隐学习为主要课题的实验表明，通过这种方式学习的知识无法被受训练的人有意识地访问。[7] 这种现象在日常生活中的一个例子是骑自行车：我们知道如何骑车，但无法解释我们是怎么做到的。第 2.1 节中提供了相关神经科学的更多背景。

内隐学习为设计出抗胁迫式的安全系统提供了一个令人向往的工具。在本文中，我们重点关注用户身份验证：利用内隐学习，训练人类大脑掌握可在身份验证期间检测、但无法被用户明确描述的密码。此系统避免了人们因受他人劝说而泄露密码的问题。要使用此系统，首先需要训练参与者执行一项称为串行拦截序列学习（Serial Interception Sequence Learning，SISL）的特定任务，如下一节中所述。训练时将使用一款主要依赖内隐学习的计算机游戏，训练结果是掌握一组充当身份验证密码的特定按键序列。在我们的实验中，训练活动为时大约 30 到 45 分钟，参与者将掌握一个拥有约 38 比特熵的随机密码。我们开展的实验表明，参与者在训练之后无法重新构建其受训的序列。

今后进行身份验证时，参与者将再次执行包含多个嵌入序列的 SISL 任务，其中包括之前已训练序列的元素。若在训练元素上的表现稳定优于未训练元素，参与者可在 5 到 6 分钟内验证其身份。不知道训练序列的攻击者无法呈现用户的表现特征（在训练结束时衡量）。请注意，身份验证过程是一个互动式游戏，服务器知道参与者的密码训练序列，并使用它来验证参与者身份。读者如果想要试玩该系统，可以在 brainauth.com/testdrive 上查看训练任务。

本文的最初版本刊登在 2012 年《第 21 届 USENIX 安全性专题研讨会论文集》* 中。

虽然本文关注的是抗胁迫用户身份验证系统，但身份验证仅仅是冰山一角。我们希望可以利用内隐学习设计许多其他的抗胁迫安全原语。

威胁模型。我们所提出的系统，被设计成一个本地密码机制，该机制需要用户本人在场。也就是说，我们考虑的是安全位置入口处的身份验证，门卫可以确保真人在不借助电子仪器的前提下参与测试。

为了欺骗身份验证测试，对手可以拦截一名或多名受过训练的用户，使他们（或许通过胁迫方式）尽可能泄露信息。然后，对手亲自参与到实时身份验证测试中，其目标是通过该测试。

我们要强调的是，该系统的设计和标准密码身份验证一样，不能抵御窃取式攻击，如在身份验证过程中偷窥密码输入。尽管问答式协议是标准的窃取预防手段，但要基于内隐学习来设计问答式协议，目前依然是尚待解决的问题。我们将在本文结尾部分重新讨论这一问题。

胜于生物识别身份验证的优点。训练的机密序列可以视为一种能够验证受训练参与者身份的生物识别密钥。不过，与生物识别密钥不同的是，这种身份验证信息不能被秘密复制，参与者即使愿意也无法泄露所训练的机密。此外，如果训练序列被攻破，可以训练新的身份识别序列作为替代，从而导致更换密码。

在相关的文章中，Denning 等人[1]提出使用图像来训练用户以内隐方式记忆密码。这种方式可能无法抵御软磨硬泡式攻击，因为用户将记住哪些图像他们看到过，哪些则没有。此外，基于图像的方式还需要准备大量的图像，并且每个用户仅使用一次，这使得系统的部署难度更高。我们这种基于组合学的方法，可以让我们对被学习的密码的熵有一个下界，设置简单，也可经过设计，不留下任何训练序列的有意识踪迹。

用户研究。为了验证我们的方案，我们使用 Amazon 的 Mechanical Turk 进行了一系列用户研究。我们询问了以下核心问题，探索通过内隐学习进行身份验证的可行性：

- 个人身份识别是否可靠？即，被训练的用户能否重新进行身份验证，可否在过段时间后依然能够验证？
- 攻击者能否通过从受训练参与者那里轻松获取的表现数据来对该序列进行反向工程？

通过三个实验，我们展示了颇有前景的初步结果，可以为设计的实际实施提供支持。首先，我们展示了通过相对较短的训练和简单测试来进行身份识别是可以实现的。其次，用户习得的信息可以保持一到两周的时间：虽然有些人一周就已忘记，两周后又有一些人忘记，但这表明了比较长（指数形）的遗忘曲线。最后，在第三个实验中，我们审视了基于下述内容的攻击：请参与者完成包含所有最短长度片段的序列，以此尝试重建身份识别序列。我们的结果显示，参与者无法在此情况下表达对序列的可靠认知，这表明底层序列信息能够抵御攻击，直到攻击者正确猜测出更长的子序列为止。

2. 人类记忆系统概述

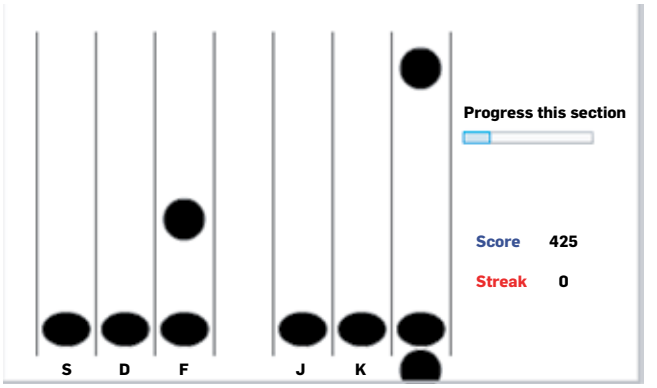
任何掌握了某种技能专长的人应该都熟悉，了解如何使用某项学好的技能和解释如何使用，这两者是有区别的。这种分离性反映了人类大脑中存在多个记忆系统。[5] 对于可口头表达的事实和事件的记忆依赖于内侧颞叶记忆系统（包括海马体）。然而，即便是内侧颞叶受损的患者（如因老年痴呆症而导致）也具备以内隐方式掌握新信息的完好能力，包括正常学习某些运动任务。[3] 通过数十载的实验性认知心理学研究，科学家们已经开发出可选择性地依赖这种内隐式、无意识学习系统的任务。

2.1.SISL 任务

在串行拦截系列学习 (SISL) 任务[7]中，人类参与者在不知不觉中学习一个字母序列。该任务要求参与者拦截以预先确定的序列送出的移动物体（圆圈），具体的操作与热门游戏“吉他英雄”非常相似（图 1）。

在我们的修改版 SISL 中，每个圆圈出现在一列的顶部（共有六列），以恒定的速度垂直落下，直到其到达底部的“槽口”为止，此时它就会消失。玩家的目标是在物体接近槽口时将它拦截。拦截的方式是在物体处于正确的垂直位置时，按下与物体所在列对应的按键。按键错误或未按任何键将导致该物体的结果

图 1.进行中 SISL 任务的屏幕截图。



为不正确。在典型的 30–60 分钟训练活动中，参与者完成数千次尝试，在其中 80% 的尝试中，提示的顺序将遵循一个秘密嵌入的重复序列。该任务经过设计，可通过逐渐改变圆圈掉落的速度，使击中率达到大约 70%，从而使每个用户达到（但不超过）其能力的极限。通过将提示遵循所训练序列期间的表现水平（准确率）与提示遵循非训练序列期间的水平进行对比，就可以评估用户对嵌入的重复序列的掌握情况。

向用户呈现的所有序列都经过设计，防止出现显而易见、容易记忆的模式。具体而言，训练序列和随机序列都经过设计，其包含的每个有序字符对在一行中仅出现一次，无一字符出现两次（因此序列长度必须为 $6 \times 5 = 30$）。其结果为，虽然训练序列的表现要好于非训练序列，但参与者通常不能有意识地识别出训练序列。为了在实验中确认这一点，在 SISL 之后，参与者通常会被要求完成外显识别的测试，指出他们对各个不同序列的熟悉程度。

与使用 4 个键的原版 SISL 任务相比，我们的版本将可能序列的范围从仅 256 提高到超过 2400 亿。此外，我们在视觉显示的中部加入了一个间隙，使它更容易地将每一列与负责的左、右手产生正确的关联。

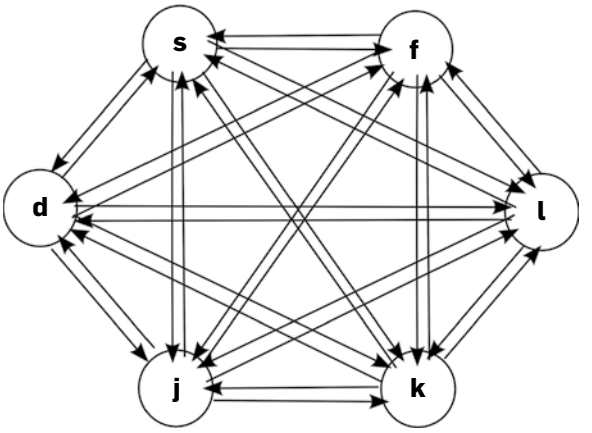
SISL 任务作为一个 Flash 应用程序通过网页浏览器提供给用户。参与者前往我们的网站 (www.brain-auth.com)，然后会看到一份同意书。他们同意参加后，网页小程序即会下载一个随机训练序列，参与者就可以开始执行任务。在完成训练和测试尝试后，我们进行外显识别测试，然后将结果上传到服务器。讨论完身份验证系统后，我们会回过头来介绍 SISL 小程序在我们有多名用户参与的大型实验计划中发挥怎样的作用。

3. 利用内隐学习的基本身份验证系统

借助 SISL 任务，可以在人类大脑中存储可在身份验证期间检测、但无法被用户明确描述的机密密钥。这样的系统避免了人们因受他人劝说而泄露密码的问题，可以形成抗胁迫身份验证协议的基础。如果信息被泄露，可以训练新的身份识别序列作为替代，从而导致更换密码。

身份识别系统的运作分两个步骤：训练，然后身份验证。在训练阶段，用户学习的机密密钥如扩展 SISL 任务中所示，即一个由 30 个字符组成的序列，集合 $S = \{s, d, f, j, k, l\}$。我们仅使用 30 字符的序列，其与图 2 中的图上的一个 Euler 圈 (即每一条边恰出现一次的圈) 对应。这些序列有一个特性，S 集合中的每个非重复双字组（如 sd、dj 和 fk）都仅出现一次。为了预测下一项（例如，用于展示表现成绩提高），需要学习由三个或更多个项组成的小组之间的关联性。这将消除学习字母频率或常见字母对的可能，也就减少了对嵌入重复序列的有意识识别。[2]

图 2.我们生成的机密密钥是来自这一有向图中所有Euler 圈的随机 30 字符序列。生成的序列所包含所有双字组仅出现一次，并且不含重复字符。



我们用 Σ 表示所有可能机密密钥的集合，即与图 2 中 Euler 圈对应的 30 字符序列的集合。此图中的 Euler 圈的数量可通过 BEST 定理计算 [11]（其中 K_6 图的生成树数量为 6^4（根据 Cayley 公式），每个顶点的入度为 5，因此 $\Pi_v (deg(v)-1)!$ 是 24^6），得出下列等式

$$\#\ 密钥 = |\Sigma| = 6^4 \cdot 24^6 \approx 2^{37.8}。$$

因此，学习的随机机密为大约 38 比特的熵，比标准可记忆密码的熵大得多。

训练。用户在可信环境中玩 SISL 任务游戏，学习一个随机的 30 项的密钥 $k \in \Sigma$。在训练用户时，我们实验了下列程序：

- 在执行 SISL 任务期间，向受训人员提供包含 30 项的密钥序列（重复三次）以及从其他随机序列中选择的 18 个项（但有一个限制，即不能出现同一提示的连续重复），总计为 108 个项。
- 这一序列重复五次，所以一共向受训人员显示 540 个项。
- 在这一序列末尾，SISL 任务出现一个简短的暂停，然后包含 540个项的完整序列（包括其末尾的暂停）再重复六次。

在整个训练活动中，一共向受训人员显示 7 × 540 = 3780 个项，需要大概 30–45 分钟时间完成。在训练阶段完成后，受训人员进行身份验证测试（如下所述），确保训练成功。系统记录用户达到的最终游戏速度。

SISL 身份验证。 过段时间后再次身份验证时，向训练的用户呈现 SISL 任务，其提示结构中包含了来自训练的身份验证序列的元素，以及用于比较的非训练元素。若在训练元素上的表现成绩稳定优于未训练元素，参与者即可验证其身份。具体而言，我们对下列身份验证过程进行了实验：

- 我们假设 k_0 是训练的包含 30 个项的序列，k_1 和 k_2 是另外两个从 Σ 中随机选取的包含 30 个项的序列。所有身份验证过程中都使用相同的序列 (k_0, k_1, k_2)，这样就不会显露出有关 k_0 的附加信息。
- 系统选择 (0, 1, 2, 0, 1, 2) 的随机排列 π（如 π = (2, 1, 0, 0, 2, 1)），向用户呈现包含下列序列（540 = 18 × 30 项）的 SISL 任务：

  $$k_{\pi 1}, k_{\pi 1}, k_{\pi 1}, \ldots, k_{\pi 6}, k_{\pi 6}, k_{\pi 6}。$$

  即，k_0、k_1、k_2 中的每个都向用户刚好显示六次（两组三个重复），但次序是随机的。任务开始的速度与用户在训练结束时的速度相同。

- 对于 $i = 0, 1, 2$，使 p_i 等于用户在输入序列 k_i 的所有轮数中密钥输入正确的比率。满足以下条件时，系统宣告身份验证成功：

  $$p_0 > \text{average}\,(p_1, p_2) + \sigma \qquad (3.1)$$

  其中 $\sigma > 0$ 足够大，可以将偶然出现此差距的可能性降到最低，又不会导致身份验证失败。

在上述初步构想中，身份验证流程存在易受以下攻击的可能：非训练用户在两个区块之间降低其表现水平，以呈现出有利于训练序列的人为表现差异（获得通过身份验证的 1/3 概率）。我们将在第 5 章节中讨论抵御这种情形的可靠方式。但现在，对于这个简单的评估过程，我们可以通过两个简单的预防措施提供一些保护。首先，验证身份验证者是真人，确保难以持续改变陪衬区块 k_1 和 k_2 之间的表现。其次，在掌握序列过程中得到的最终训练速度对身份验证服务器是已知的，攻击者无法匹配训练区块和陪衬区块之间的表现成绩差异。表现成绩差别如果与训练后获得的存在显著不同，这就表明存在攻击。

分析。 以下两个章节讨论此系统的两个重要方面：

- 有用性：受训练的用户能否在以后可靠地完成身份验证任务？
- 安全性：如果拦截到受训练用户并胁迫其提供足够多的信息，攻击者可否通过身份验证？

4. 有用性实验

我们在初步实验中的报告表明了 SISL 身份验证系统的可行性和前景。我们分为三个阶段开展这些实验。首先，我们确认，使用 Mechanical Turk 可以通过 SISL 任务的全新扩展版观察到可靠的学习成果。其次，我们确认用户在 1 周和 2 周后仍保持对训练序列的认知。最后，我们对可用于重新构建原始系列的最小片段进行采样，并据此调查了对参与者的序列知识进行攻击的有效性。

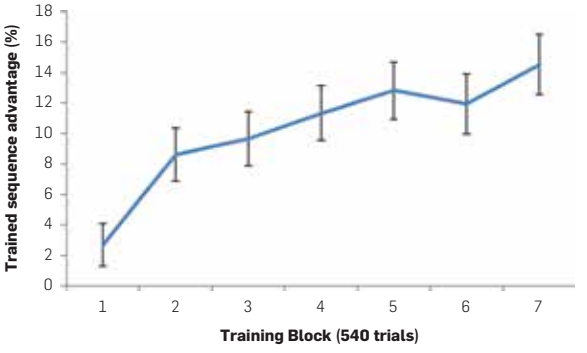
这些实验在 Amazon 的 Mechanical Turk 平台中在线开展。Mechanical Turk 的优势在于可以招募到基本不受限制的参与者，成本也将对较低。在线实验有一个缺点，即对于那些之后回来重复进行评估的用户，我们相对缺乏控制力。

4.1. 实验 1：内隐学习和外显学习

我们第一个实验确认了内隐学习可以被明确地检测到，而且对序列的外显有意识认知达到最小。分析中包含了来自 35 名参与者的实验数据。

该实验使用了前一章节中所述的训练过程，其中训练阶段包含总共 3780 次尝试，大约耗时 30-45 分钟。注意，该训练由七个各包含 540 次尝试的训练块组成。完成训练活动后，参与者进行了 SISL 身份验证测试，

图 3.在训练期间，参与者逐渐开始表现出对重复序列的认知，即在训练序列上的表现成绩优于随机散布的干扰段上的表现。请注意，任务的总体表现成绩始终保持在大约 70%，原因在于该任务的自适应性，即随着参与者在 SISL 上的总体表现的改善，其速度也会加快。



该测试将训练序列上的表现与两个随机测试序列上的表现作对比。

如图 3 所示，训练序列的学习成果取决于训练序列相对于随机出现的干扰段的表现优势（正确回答百分比的提升）。在训练之后的测试块中，参与者完成 SISL 的平均正确率为：训练序列 79.2%，非训练序列 70.6%。正确率相差 8.6% (SE[a] 2.4%) 表明，训练序列的表现存在可信的优势（成对样本 *t*-test 与零比较，$t(34) = 3.55, p < 0.01$）。

群体层面上的表现差别常常体现在内隐学习的测试中，但对于身份验证方法而言，必须能够进行可靠的个体评估。就个体参与者而言，在 540 测试尝试中，35 个案例中有 25 个可以辨别出训练序列与非训练序列的表现差别（卡方分析，$p < 0.05$）。出于身份验证的目的，需要通过更长的训练来确立内隐习得的序列，以此进一步加强评估的个体可靠性。然而，SISL 任务的特点就是能够通过相对较短的训练在大量个体中识别学习成果，而大多数内隐学习测试都不具备此特点。[7] 传统而言，内隐学习的测量依赖于评估个人群体的表现，不能在个体层面上识别学习成果。[9]

外显识别测试。完成训练和测试模块后，向参与者显示五个不同的动画序列，并询问对每一个序列的熟悉程度（从 0 到 10 打分）。在这五个序列中，其中一个是训练序列，另外四个是随机选择的陪衬序列。此测试评估了训练序列的外显识别记忆。

在识别测试中，参与者将对序列的熟悉程度按照 0 到 10 分进行打分，其平均分为：训练序列 6.5 分 (SE 0.4)，新的非训练序列 5.15 分 (SE 0.3)。从群体层面看，训练序列的识别率稍高是确凿的（$t(34) = 3.69$，$p < 0.01$），但并不与 SISL 表现关联 ($r = 0.13$)，表明这对内隐表现的好坏没有什么作用。内隐学习实验中经常会看到训练序列识别率稍高的现象，因为健康的参与者会在练习之后发现训练序列的某些部分比较熟悉。值得一提的是，内隐记忆并不转变为外显知识，即使经过重复使用也是如此；而且，训练的结构和长度以及测试序列都经过了专门设计，以降低随时间推移而积累外显知识的可能性。

识别率的总体差别较小（5.15 对 6.5）表明，参与者无法回忆出包含 30 个项的序列。这意味着，他们无法有意识地生成训练信息（例如，用于破坏身份验证方式的安全性）。我们将在第三个实验中进一步讨论重建问题。

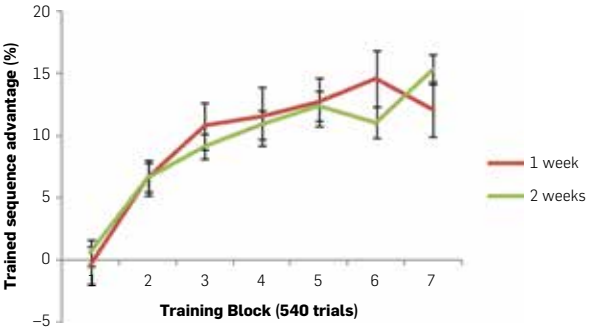
4.2. 实验 2：长期保持

只有在密码被记忆了一段时间以后身份验证依然可以准确执行，这样的身份验证机制才有用。我们在实验 2 中确认，用户获得的与序列相关的知识可以保持比较长的时间。尽管学得的技能通常会维持一段时间，但此前从未进行过具有较长延期和大量参与者的基于 SISL 的测试。

在实验 2 中，参与者同意分两个阶段完成 SISL 任务。参与者在第一阶段中完成与实验 1 中结构相同的训练序列。训练之后立即进行相同的 SISL 测试，以评估延期前的序列知识。一个由 32 名参与者组成的小组在 1 周后，返回在线小程序以进行训练序列的保持力测试和识别评估。另一个由 80 名参与者组成的小组在 2 周后进行保持力测试和识别测试。对于 1 周后返回的小组，测试由一个包含 540 次尝试的内隐序列学习评估组成。对于 2 周后返回的小组，测试的时长加倍，以进一步评估更长的测试能否提高对个体序列知识的敏感度。对于两个小组，延期测试的初始测试速度都设为与参与者在训练活动末尾执行该任务时的速度相匹配。通过一个由 180 次尝试组成的热身块来调整这一初始速度，以便参与者能够在保持力测试开始时，以大约 70% 的目标正确率执行任务。

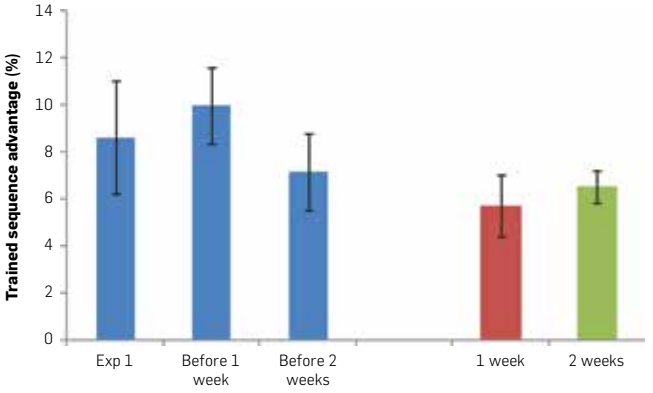
图 4 显示，和实验 1 一样，两组人都在第一阶段中逐渐地学习训练序列。图 5 显示了立即测试和延期测试的内隐序列知识。在所有五个评估中，参与者在

图 4.在训练期间，参与者逐渐开始表现出对重复序列的认知，即在训练序列上的表现成绩优于随机散布的干扰段上的表现。和预计的一样，两个组的学习表现相似，并与实验 1 相似。



[a] SE 是 *Standard Error*（标准误差）的缩写。换言之，如果训练序列和非训练序列的正确率测量遵循相同的正态分布，由 $N = 35$ 样本（所以有 $N - 1 = 34$ 个自由度）计算出来的 *t* 值应当接近于零，小于此处获得的 *t* 值（3.55），这表示有 99% 的概率这一差距是显著的。*t*-test 是一种标准的统计方法，用于确认被控变量（此处为序列类型）对测得的变量（正确率）的影响。

图 5.参与者在两个即时评估（实验 1 所示，以及实验 2 的两种状况）上展现了可靠的序列知识，即在测试时，训练序列上的表现成绩优于未训练的新序列上的表现。1 周和 2 周延期测试都体现了序列知识的可保持性。尽管在两个延期后都体现出知识的表达有一些缩减，但 1 周到 2 周之间没有显著的进一步衰减，这表明信息有可能会在 2 周后还能保持比较长的时间（许多类型的记忆都观察到指数或幂律衰减曲线）。



整体上展现了可靠的序列学习成果，$ts > 4.3$，$ps < 0.01$。在 1 周延期测试中，32 名参与者中有 15 名在个体上展现了可靠的序列知识。不过，对于 2 周延期小组，80 名参与者中有 49 名在个体上展现了可靠的序列知识，反映出使用较长的评估测试可以提高敏感度。未来的研究不仅将考察延长的训练时间，也将考察对个体知识敏感度更高的评估测试，以根据 SISL 表现提供可靠而准确的识别方式。

即使是延期 1 周和 2 周后，参与者仍表现出同样的趋势（虽然这个趋势并不明显），即对训练序列的识别度更高，$ts > 2.8$，$ps < 0.05$。需要重申的是，识别表现和序列知识表达没有关联（$rs < 0.16$），也没有人能够回忆整个包含 30 个项的训练序列。

5. 安全性分析

在这一章节中，我们将分析第 3 章节中的基本身份验证协议的安全性，并提出可提高安全性的一系列拓展。我们还对一次特定的攻击进行了实验，该攻击试图一次一个片段地从用户处获取机密序列。我们的 Mechanical Turk 实验显示，这种攻击对人类效果不佳。

5.1. 内隐学习作为密码原语

我们先看看通过内隐学习实现的新功能的抽象模型。在传统的建模中，密码协议中的参与者被建模为拥有对手不知的机密的实体。这些假设在面对胁迫时遭到瓦解，因为此时可以从参与者身上提取到所有的机密。

内隐学习提供了以下新的抽象功能：训练阶段将断言

$$p : \Sigma \to \{0, 1\}$$

嵌入到用户大脑中，对某个很大的集合 Σ。任何人都可以要求用户评估其断言 p 在某个点上 ($k \in \Sigma$) 的值。用户习得了 k 时，断言评估为 1，否则为 0。p 评估为 1 的输入数量相对较小。在绝大多数情形下，p 仅在一个点上评估为 1，也就是说，用户仅针对一个机密序列接受训练。

内隐学习的主要特征是，即便在受到胁迫的情况下，也无法从用户提取 $p(k) = 1$ 的点 $k \in \Sigma$。这一抽象属性抓住了这一事实，即机密序列 k 是用户通过内隐方式习得的，无法被有意识地访问。在这篇论文中，我们使用内隐学习原语来构建身份验证系统，但我们可以设想它在安全系统中得到更广泛的使用。

第 3 章节中所述的身份验证过程提供了 Σ 中某序列 k_0 的断言 $p(\cdot)$ 的实现。如果该过程宣告成功，我们可以认为 $p(k_0) = 1$，否则 $p(k_0) = 0$。断言 p 在训练活动中嵌入到用户的大脑中。

基础胁迫威胁模型。第 3 章节中的 SISL 身份验证系统经过设计，可抵御企图欺骗身份验证测试的对手。我们假设测试要求真人在场并从活性检查开始，以确保真人在没有任何仪器的协助下参加测试。为了欺骗身份验证系统，允许对手进行以下步骤：

- 提取步骤：拦截一名或多名受训练的用户，使他们（或许通过胁迫方式）尽可能泄露信息。
- 测试步骤：对手亲自提交信息到身份验证测试，其目标是通过该测试。在现实中，这可能意味着对手出现在安全设施的入口处，试图通过那里的身份验证测试。如果失败，他可能被扣留下来质询。

这种基础威胁模型给予攻击者一次机会挑战身份验证测试。本章节稍后部分中，我们将考虑另一种模型，即攻击者可能会重复提取和测试步骤，并在提取和测试之间交替进行。

我们也应注意，基础威胁模型假设，在训练阶段（即用户被授予凭据时），用户会遵循相关指示，而不会有意尝试误导训练过程。实际上，对手仅被允许在训练过程完成之后胁迫用户。

显而易见，第 3 章节中的系统在这一基础威胁模型下是安全的（假定训练过程将内隐习得的断言 p 嵌入在用户的大脑中）。事实上，如果攻击者拦截 u 名

受训练用户，使每一人遭受 q 次查询，其能够找到有效序列的概率最多为 $qu/|\Sigma|$。由于每一测试用时约五分钟，我们可以假设每个被俘用户的最多尝试次数为 $q = 10^5$ 次（此数量大概是每名用户不间断测试大约一年，而这可能会干扰用户习得的密码，导致用户对攻击者无用；或者向安全管理员发出用户不在场警报，导致凭据被撤销）。因此，即使在俘获了 $u = 100$ 名用户后，攻击者成功的可能性也仅为

$$100 \times 10^5/|\Sigma| \approx 2^{-16}.$$

让攻击者更麻烦的是，使一个人通过 SISL 查询许多个随机序列可能会导致其忘掉已习得的序列，或者导致其学会不正确的序列，从而使提取变得不可能。

我们应注意，设计为抵御胁迫攻击的身份验证系统需要真人在场。如果系统支持远程身份验证，那么攻击者可以胁迫受训练用户通过远程服务器进行身份验证，并劫持认证会话。

安全性增强。上述安全模型给予攻击者一次身份验证的机会，攻击者必须有较高的成功概率。如果攻击者被允许进行多次身份验证尝试，即重复执行提取和测试步骤并在两者之间交替进行，那么该协议可能会变得不安全。其原因是，攻击者在身份验证尝试期间能够看到三个序列 k_0、k_1 和 k_2，有可能会记住其中之一（30 个符号）。然后他可以对该序列进行离线训练，这样在下一次身份验证尝试时他可以拥有 1/3 的成功概率。如果攻击者能够记住所有三个序列（90 个符号），然后再重建 SISL 任务，他就能够迫使受训练的用户离线接触所有三个记住的序列，通过用户的表现可靠地判断哪个是正确的身份验证序列。之后攻击者可以针对该特定序列训练自己。这样，他在下一次身份验证尝试时就能*确保*获得成功。我们要补充的是，这一攻击者很难实现其目标，因为对于人类攻击者而言，要以执行任务的速度记住整个序列非常困难。

第 3 章中提到了另一种可能的攻击，攻击者碰巧是一名高手玩家，但有意在所提供的其中序列上降低其表现成绩。他有 1/3 的概率可以在正确的序列上展示出表现差别，从而通过身份验证测试。我们在第 3 章节中介绍了几种抵御方式。这里我们介绍一种更鲁棒的方式。

以上两种攻击都可通过组合学击破。我们训练用户时不针对单个序列，而是针对若干序列，例如四个。实验 [8] 表明，人类大脑可以学习多个序列，而且这些习得的序列不会互相干扰。另外，我们进行的新实验也表明，用户可以在 24 到 48 小时的间隔时间内训练多个包含 30 字符的序列，序列之间并没有出现可测量到的干扰。同样，我们也能够针对较长的序列训练用户，并使用其片段来进行身份验证。尽管上述数据表明，最短的片段（3 个项）无法用于评估对较长序列的认知，但最近我们发现，针对较长的片段（如 5 - 7 个项），该序列知识却能够可靠地表达出来。[6] 因此，通过在初始训练中投入更多时间对更多信息进行编码，我们可以使用基于片段的测试来提高对上述窃取式攻击的抵御能力。

在身份验证期间，我们不使用一个正确序列和两个陪衬序列，而是使用四个正确的序列（或片段）并随机搭配 8 个陪衬序列。如果攻击者在 12 个预设序列中的 4 个正确序列上显示出可测量到的表现差别，那么就可通过身份验证。而在随机序列上速度减慢的攻击者现在最多有 $1/\binom{12}{4} \approx 1/500$ 的几率通过测试。可以通过调节训练序列数量 (4) 和陪衬序列数量 (8)，在安全性和有用性之间达成可接受的平衡。

与之类似，少量的身份验证尝试将不能帮助直接攻击者通过测试。不过，记住身份验证测试（360 个符号），稍后再呈现给受胁迫的用户，可以给对手带来优势。为进一步防御这种记忆式攻击，我们在身份验证过程中加入了一个额外步骤：一旦身份验证服务器发现用户无法在部分训练序列上展现可测量的表现差别，所有剩余的训练序列就被替换为随机陪衬序列。这样一来，若是攻击者在没有先前知识的情况下尝试身份验证，就不能看到全部的训练序列，因而就无法从受胁迫的用户那里提取到所有训练序列。因此，就无法对受胁迫用户发起*“一击致命”的*攻击。然而，通过重复这一过程，即进行身份验证测试、记忆观察到的序列，再在受胁迫的训练用户身上测试它们，攻击者或许最终能学会所有训练序列，并成功欺骗身份验证测试。但在这一过程中，攻击者必须参与身份验证测试，在该测试中，他能够证明自己对严格的一小组训练序列的认知，但无法证明对所有序列的认知。这给予系统一个受到攻击的明确信号，此时参与身份验证的人员可能会被扣留进行质询，而合法的用户将被阻止通过该系统进行身份验证，直到其重新训练了一组新的序列。

窃取安全性。传统的密码身份验证容易遭受窃取（通过客户端恶意软件或肩窥方式），此处所示的身份验证系统亦是如此。窃取者如果获取了受训练用户的多个有效身份验证脚本，就可重新构建学习的序列。设计一套抗胁迫系统并在服务器的问答式协议中使用

内隐习得的机密，是一个很有前景的未来研究方向。我们将在本文结尾部分重新讨论这一问题。

5.2. 实验：提取序列片段

我们的系统可能会受到这样的攻击：怀有恶意的一方分析出合法用户的知识特征并使用该信息对训练序列进行反向工程，进而通过身份验证测试。虽然可能的训练序列数量过多，无法对任何一个序列进行穷举测试，但每一序列的构建都有已知的局限，掌握序列片段或许可让攻击者能够重建原始序列或者其足够的部分，以至通过身份验证测试。

训练序列被限制为以均等的频率使用所有 6 个回答按键，因此对个别回答概率的分析无法提供有关训练序列的信息。与此类似，所有 30 个回答按键对（6 × 5 = 30，因为按键都不重复）在训练期间以均等的频率出现，这意味着双字组频率也不能提供有关训练序列的信息。不过，每个包含 30 个项的序列拥有 30 个唯一三字组（共 150 个可能）。如果特定的训练三字组片段可以被识别，那么其中含有的训练序列就有可能被重建。

基于这一信息的攻击将必须让受训练用户执行 SISL 测试，该测试包含频率均等的所有 150 个三字组。如果用户在 30 个训练三字组上的表现优于在 120 个非训练三字组上的表现，那就可以重建该序列。此攻击能够削弱该方法对外部压力的相对抵抗力，而致泄露身份验证信息。

然而，虽然可以在三字组层面上确定序列信息，但目前尚不清楚参与者在如此短的片段上能否可靠地展现序列知识。在实验 3 中，我们评估了在这类三字组测试上的表现，以此衡量重建序列信息的可能性。

我们再一次通过 Mechanical Turk 招募参与者，并完成了与实验 1 和 2 中相同的训练活动。在测试时，参与者所执行的序列被构建为提供 150 个三字组，每个都刚好提供 10 次，即构建 10 组不同的包含 150 次尝试的单元，每一组以不同的顺序包含所有可能的三字组。将每个三字组上的表现成绩作为当前回答与前两个回答的函数来衡量（表示为正确率）。

为评估此数据是否能用于重新构建序列，我们逐一计算各个三字组的正确率，并逐一创建所有三字组的排名。如果训练三字组的表现优于其他三字组，那么训练三字组的排名往往较低（例如，成绩表现会导致序列三字组成为前 30 个最佳回答）。然而，在平均排名和平均正确率上，训练三字组和非训练三字组之间没有明显的差别。参与者并未在这一类型的测试上展现出他们的训练序列知识，这表明无法借助基于三字组的方式攻击其序列知识。更具体而言，我们针对每一用户比较了 30 个训练序列三字组和 120 个剩余三字组的平均正确率测量结果。34 名参与者的三字组平均正确率为：训练序列 73.9% (SE 1.2%)，其余 73.2% (SE 1.1%)。这一差别不具足够的说服力。

虽然三字组测试并不能表现出序列知识，但也存在通过一些更长的片段评估出参与者序列知识的可能。实际上，我们的进一步实验证实了这一理论：当片段长度为 4 时，我们发现片段中的第三个字符确实会显现较好的表现成绩，与训练一致；当片段长度为 5 时，第三和第四个字符显现表现成绩的提高。有趣的是，向用户呈现的训练片段中的最后一个字符并未显现出优于非训练序列平均表现的成绩。其原因可能是，下一片段中的第一个（意料之外的）字符“重置”了用户的表现。

使用片段分析的攻击难度很高，因为即便是中等长度的片段，其数量也过多，而且分析每一片段的用时也过长。例如，对于长度为 4 的片段（四字组），存在 750 种可能性并且需要运行多次。在训练序列中添加可变时序可以进一步减少对协议的这一类攻击，这种方式可以快速扩展组合空间。通过实验我们发现，时序变化可以“抹掉”在以不同时序模式训练的序列上的表现。[4]

6. 总结和未来研究

我们介绍了胁迫攻击的一种新型抵御方式，它利用来自认知心理学的*内隐学习*概念。我们介绍了一个概念验证协议，以及通过 Mechanical Turk 开展的初步实验，这些实验让我们相信，构建抗软磨硬泡式攻击的身份验证系统是可行的。

尚有许多工作要去完成。我们希望进一步分析内隐习得的密码的遗忘速度，以及巩固训练的所需频率。此外，我们也希望找到方法来检测或预测个体用户可靠学习的时间（收集更多用户的人口统计数据，以及开展多阶段长期实验，或许是这一方向的良好开端）。我们还希望探索这一方法的其中一些限制。例如，通过确定习得序列中对攻击者与合法身份验证者存在差别的部分的最短长度，以及加强测试过程和分析，来提高在更大比例用户之中的可靠性，或缩短必要的测试时间，减少误报和漏报等。在提示之间使用可变时序，并将用户表现成绩作为任务速度的函数来测量，可进一步提高测试协议的可靠性。多个凭据的内隐学习同样可从更多实验中获益。这些实验所依据的以往研究工作已发现，用户在学习不同的包含 12 个项的序列时

没有出现互相干扰的迹象，同时用户还能以内隐方式学习最长为 80 个项的序列。

这一研究工作的另一未来方向是测试能否以内隐方式学习更为复杂的结构，如 Markov 模型。我们希望利用此类学习来构建可抵御窃取和胁迫的问答式身份验证体系。最后，除了身份验证外，我们也希望探究各种以内隐学习为基础的密码原语的构建。

致谢

在此感谢所有有偿志愿者的参与，他们为我们的用户研究做出了重要的贡献。本研究获得了 NSF 和 MURI 项目的资助。

参考资料

1. Denning, T., Bowers, K.D., van Dijk, M., Juels, A. Exploring implicit memory for painless password recovery.In *CHI*. D. S. Tan, S. Amershi, B. Begole, W. A. Kellogg, and M. Tungare, eds. ACM, 2011, 2615–2618.
2. Destrebecqz, A., Cleeremans, A. Can sequence learning be implicit?New evidence with the process dissociation procedure.*Psychonomic Bull.Rev. 8* (2001), 343–350.
3. Gobel, E., Blomeke, K., Zadikoff, C., Simuni, T., Weintraub, S., Reber, P. Implicit perceptual-motor skill learning in mild cognitive impairment and Parkinson' s disease.*Neuropsychology*, *27*, 3 (2013), 314–321.
4. Gobel, E., Sanchez, D., Reber, P. Integration of temporal and ordinal information during serial interception sequence learning.*J. Exp.Psychol.Learn. Mem.Cognit.37*, 4 (2011), 994–1000.
5. Reber, P. Cognitive neuroscience of declarative and non-declarative memory.*Parallels in Learning and Memory*.M. Guadagnoli, M.S. deBelle, B. Etnyre, T. Polk, and A. Benjamin, eds.North-Holland, 2008, 113–123.
6. Sanchez, D., Bojinov, H., Lincoln, P., Boneh, D., Reber, P. Statistical learning in perceptual-motor sequences and planning effects in performance.In *Poster at the Meeting of the Society for Neuroscience* (2012).
7. Sanchez, D., Gobel, E., Reber, P. Performing the unexplainable: implicit task performance reveals individually reliable sequence learning without explicit knowledge.*Psychonomic Bull. Rev. 17* (2010), 790–796.
8. Sanchez, D., Reber, P. Operating characteristics of the implicit learning system during serial interception sequence learning.*J. Exp.Psychol.Hum. Percept.Perform.38*, 2 (2012), 439–452.
9. Schwarb,H., Schumacher,E.H.Generalized lessons about sequence learning from the study of the serial reaction time task.*Adv. Cognit.Psychol.8*, 2 (2012), 165–178.
10. Soghoian, C. Turkish police may have beaten encryption key out of TJ Maxx suspect, 2008. news.cnet.com/8301-13739_3-10069776-46.html.
11. van Aardenne-Ehrenfest, T., de Bruijn, N.G.Circuits and trees in oriented linear graphs.*Simon Stevin 28* (1951), 203–217.
12. Wikipedia.Rubber-hose cryptanalysis, 2011.

Hristo Bojinov 和 Dan Boneh 来自美国加利福尼亚州斯坦福的斯坦福大学

Daniel Sanchez 和 Paul Reber 来自美国伊利诺伊州埃文斯顿的西北大学

Patrick Lincoln 来自美国加利福尼亚州门洛帕克的斯坦福国际研究所

译文责任编辑：孙晓明