一种进行密码猜测的深度学习方法

布里兰德，希塔杰，保罗·加斯蒂，朱塞佩·阿特尼塞*和费尔南多·佩雷斯克鲁兹*
*史蒂文斯理工学院计算机科学系
电子邮件：史蒂文斯。edu
美国纽约理工学院计算机科学系
电子邮件： pgasti@nyit.edu

摘要-最先进的密码猜测工具，如哈猫和开膛手约翰，允许用户每秒数十亿密码检查密码散列。除了直接的字典攻击之外，这些工具还可以使用密码生成规则来扩展字典。这些规则定义了转换，如单词的连接(e。g.，"密码123456”）。g.，"密码“变成了“p4s5w0rd”）。尽管这些规则在当前的密码数据集上表现良好，但创建为新数据集进行优化的新规则是一项费力的任务，需要专门的专业知识。

在本文中，我们设计了一种基于机器学习的基于理论的密码生成方法来取代人类生成的密码规则。这一努力的结果是PassGAN，一种利用生成式对抗网络（GANs）来增强密码猜测的新技术。PassGAN通过训练GAN处理泄露的密码列表来生成密码猜测。因为GAN的输出与它的训练集紧密分布，所以使用PassGAN生成的密码很可能与尚未泄露的密码相匹配。PassGAN是对基于规则的密码生成工具的一个重大改进，因为它从密码数据自动推断密码分发信息，而不是通过手动分析。因此，它可以毫不费力地利用新的密码泄露来生成更丰富的密码分发。

我们的实验表明，这种方法是很有前途的。当我们在两个大型密码数据集上评估PassGAN时，我们能够比开膛手约翰的间谍实验室规则平均高出2倍，并且我们与HashCat中最好的64代和第2代规则竞争——我们的结果在HashCat规则的2倍范围内。更重要的是，当我们将PassGAN的输出和HashCat的输出相结合时，我们能够比单独使用HashCat就多匹配18%-24%的密码。这是值得注意的，因为它表明PassGAN可以生成相当数量的当前工具无法获取的密码。

Ⅰ.介绍

密码是最流行的身份验证方法，主要是因为它们易于实现，不需要特殊的硬件或软件，而且用户和开发人员都很熟悉。不幸的是，多个密码数据库泄露表明，用户倾向于选择易于猜测的密码[14]，[18]，[42]，主要由公共字符串(e。g.，密码，123456，我爱你），及其变体。

密码猜测工具为识别弱密码提供了一个有价值的工具，特别是当它们被存储在哈希形式[54]，[58]。密码猜测软件的有效性依赖于对每个密码散列快速测试大量非常有可能的密码的能力。密码猜测工具没有彻底尝试所有可能的字符组合，而是使用字典中的单词和以前的密码泄露的单词作为候选密码。最先进的密码猜测工具，如约翰开膛手[71]和HashCat [28]，进一步采取这种方法，定义密码转换的启发式，其中包括多个单词的组合(e。g.，，混合字母的情况下(e。g.，然后开始说话（例如，你）。这些启发式方法，结合马尔可夫模型，允许开膛手约翰和哈什猫生成大量新的极有可能的密码。

虽然这些启发式在实践中相当成功，但它们是特别的，基于用户如何选择密码的直觉，而不是基于对大型密码数据集的连贯和有原则的分析构建的。此外，开发和测试新的启发式方法是一项耗时的任务，需要专业知识。为了解决这些缺点，本文提出了一种基于深度学习和生成对抗网络（GANs）[23]的生成密码猜测的新方法PassGAN生成方法。GANs是最近引入的机器学习工具，设计用于在高维空间[23]中执行密度估计。一个GAN由两个深度神经网络组成：一个生成式深度神经网络(G)和一个有区别的深度神经网络(D).D被设计用来区分“真实样本”和G生成的“假样本”。这两个深度神经网络通过多次迭代而相互作用。在每次迭代中，G的假样本给D。然后将D的输出提供给G，G将其作为反馈，生成假样本，并分布得越来越接近真实样本。经过足够次数的迭代后，G的输出就变成了GAN的输出。PassGAN利用这种技术来生成新的密码猜测。我们的核心想法是使用一个泄露的密码列表（真实的样本）来训练D。因此，在每次迭代中，PassGAN（假样本）的输出更接近原始泄漏中的密码分布，因此更有可能匹配真实用户的密码。据我们所知，这项工作是第一次使用GANs的目的。 PassGAN代表了一个原则和理论基础的产生密码猜测。我们探索不同的

神经网络配置、参数和训练程序，以确定学习和过拟合之间的适当平衡，并报告我们的结果。具体来说，我们的贡献如下： (1)我们证明了GANs可以生成高质量的密码猜测。在我们的实验中，我们能够匹配RockYou数据集[61]的真实用户密码中的2,774,26969个（46.86%），以及领英数据集[40]的43,354,871个密码中的4,996,980个（11.53%）。此外，由PassGAN生成的绝大多数与我们的测试集不符的密码仍然“看起来像”人类生成的密码；(2)我们证明了我们的技术与最先进的密码生成规则具有竞争力。尽管这些规则是专门针对我们评估中使用的数据集进行调整的，但PassGAN的输出质量相当（在HashCat的情况下），或更好（在开膛手约翰的情况下）密码规则；(3)我们的结果还表明，PassGAN可以用来补充密码生成规则。在我们的实验中，我们成功地使用PassGAN生成了没有由任何密码规则生成的密码匹配项。当我们将PassGAN的输出与HashCat的输出结合起来时，与单独的HashCat相比，我们能够匹配18%到24%的额外唯一密码；(4)与密码生成规则相比，PassGAN可以生成几乎无限数量的密码猜测。我们的实验表明，新的（唯一的）密码猜测的数量随着GAN生成的密码总数而稳步增加。这一点很重要，因为目前使用规则生成的唯一密码的数量最终会受到用于实例化这些规则的密码数据集的大小的限制。

我们认为这项工作是迈向全自动生成高质量密码猜测的第一步。我们的研究结果表明，当使用足够大的密码数据集进行训练，并且使用足够复杂的神经网络架构进行实例化时，GANs可以优于基于规则的密码猜测。此外，PassGAN实现了这个结果，而不需要通常与密码猜测规则设计相关的用户努力。

我们认为这项工作是相关的、重要的和及时的。与此相关，因为尽管有许多替代方案[55]，[67]，[21]，[17]，[80]，我们几乎没有看到任何证据表明密码会在短期内被替换。这很重要，因为建立密码猜测的极限——以及更好地理解现实世界中密码的可猜测性——将有助于使基于密码的系统更加安全。而且很及时，因为最近泄露的包含数亿个密码的信息，[20]为攻击者破坏系统，并为系统管理员提供了一个强大的数据来源，以便重新评估密码策略。

组织本文的其余部分组织如下。在第二节中，我们简要概述了深度学习、GANs和密码猜测，并提供了相关的技术状况的总结。第三节讨论了用于实例化PassGAN的GAN的存档和培训选择，以及在我们的评估中使用的超参数。我们报告了对PassGAN的评价，并与国家-的比较最先进的基于规则的技术，在第四节中。我们的结论是在第五节。

II .背景及相关工作

在本节中，我们将简要概述深度学习和GANs。然后，我们回顾了密码猜测的最新进展。

A.深度学习

在90年代中期，机器学习方法，如支持向量机[64]、随机森林[7]和高斯过程[60]，在对大多数不相关的人类工程（手工编码）特征的分类和回归方面显示出了显著的结果。从21世纪000年代中期开始，随着存储和数据可用性的增加，这些方法已经被深度学习所取代。对深度学习的研究表明，特征可以有效地从数据中学习出来，而手工编码的特征往往表现不到学习到的特征。这些增益与相关的特征更相关，其中人类工程的特征可能只编码低维的相关性。

深度学习被广泛应用于解决计算机视觉[39]、图像处理[70]、视频处理[16]，[50]，语音识别[26]、自然语言处理[2]，[12]，[79]或游戏[24]，[36]，[45]，[47].等相关问题最近，在健康相关问题[13]，[19]中使用深度学习也有了显著的改进。

深度学习对数据使用、从训练过的模型中学习到什么以及模型学习比特定任务更多私人信息的能力提出了一些隐私影响。因此，研究人员提出了保护隐私的协作学习技术[65]，以及依赖于差异隐私[1]的技术。然而，最近的研究表明，这些技术并不像最初认为的那样保护隐私。具体来说，训练后的模型容易受到信息泄漏[4]、模型反演攻击[22]、成员攻击[66]、[29]和模型提取攻击[74]的影响，即使模型使用保护隐私的协作学习技术[31]进行训练。

除了从训练的人中提取信息的攻击
最近的研究表明，样本可以进行微妙的修改，使其在人眼中看起来不变，但始终被深度学习算法[52]、[51]、[43]、[8]、[9]、[35]、[41]、[30]错误分类。已经提出了几种对策，[53]，[77]。然而，这仍然是一个开放的研究问题。

B.生成对抗性网络

生成对抗网络（GANs）代表了深度学习领域的一个显著进步。GAN由两个神经网络组成，一个生成深度神经网络G和一个鉴别深度神经网络D。给定一个输入数据集I = {x1; x2; ...; x_n，G的目标是从潜在的概率分布Pr (x)中产生被D所接受的“假”样本。同时，

D的目标是学会区分来自G的假样本和来自I的真实样本。更正式地说，在输入一个简单的噪声分布z时，用GANs求解的优化问题可以总结如下：

$$\min_{\theta_G}\max_{\theta_D}\sum_{i=1}^{n}\log f(\mathbf{x}_i;\theta_D)+\sum_{j=1}^{n}\log(1-f(g(\mathbf{z}_j;\theta_G);\theta_D))$$

哪里的模型试图最小化关于aG，并同时最大化对aD.当D无法区分G产生的假样本和I产生的真实样本时，认为学习阶段是完整的。

自古德费勒等人的原始工作以来。[23]，在GANs上已经有了一些改进。拉德福德等人。[59]引入了DCGAN，它通过使用卷积神经网络而不是多层感知器来改进了[23]。因此，与[23]相比，DCGAN可以产生更真实的图像样本。

关于GANs的其他工作包括开始的[5]，DiscoGAN [33]，有条件的GAN [46]，AdaGAN [73]，InfoGAN [11]，拉普拉斯金字塔GAN [15]，和StackGAN [78]。这些技术对以前的工作进行了改进，比如培训和使用GAN的新方法。

阿约夫斯基等人。介绍瓦瑟斯坦甘氏菌（WGAN）[3]。WGAN通过使用梯度剪切，提高了先验GANs的学习稳定性。这种方法的好处包括减少模态崩溃和有意义的学习曲线，这有助于识别最优超参数。

以上所有的工作都集中在真实图像的生成上。为了解决文本生成的问题，古拉贾尼等人。[27]最近推出了改进的瓦瑟斯坦GAN（IWGAN）。对于IWGAN，G和D都是简单的卷积神经网络（CNNs）。G以一个潜在噪声向量作为输入，通过转发到其卷积层进行转换，输出32个单热字符向量序列。在G的输出处应用软最大非线性，然后转发到D。IWGAN的每个输出字符是通过计算G产生的每个输出向量的argmax得到的每个输出字符。

C.密码管理

在密码猜测攻击中，对手试图通过反复测试多个候选密码来识别一个或多个用户的密码。密码猜测攻击可能和密码本身的[6]一样古老，更正式的研究可以追溯到1979年的[48]。

两种流行的现代密码猜测工具分别是开膛手约翰（JTR）[71]和HashCat [28]。这两种工具都实现了多种类型的密码猜测策略，包括：彻底的暴力攻击；基于字典的攻击；基于规则的攻击，包括从字典单词[63]，[62]的转换中生成密码猜测；以及基于马尔可夫模型的攻击[72]，[56]，其中密码的每个字符都通过一个考虑前一个或多个字符的随机过程进行选择，并在明文密码字典上进行训练。JTR和HashCat在猜测密码方面非常有效。具体来说，那里有几个例子表明，超过90%的
从在线服务泄露的密码已成功恢复到[57]。

纳拉亚南等人首先使用马尔可夫模型来生成密码猜测。[49].他们的方法使用手动定义的密码规则，例如生成的密码的哪一部分是由字母和数字组成的。该技术随后被Weir等人改进。[75]，他展示了如何从密码分发中“学习”这些规则。这一早期的工作随后被Ma等人扩展。[42]和Durmuth等人合著。[18].基于马尔可夫模型的技术也被用于实现实时密码强度估计器，并评估明文数据库中的密码强度(见，e。g.，[14]，[10]).

概率上下文无关语法（PCFGs）[32]，[75]利用密码结构上的手动编码信息来生成新的猜测。这个信息可以是隐式的(e。g.，一个字典单词，后面跟着用户的出生日期)或显式的(e。g.，密码必须包含至少六个字符，一个大写字母和一个数字)。然后随机选择适当的标记，从生成的语法中构建密码。

最近，梅利歇尔等人。[44]介绍了一种基于递归神经网络[25]，[69]的密码猜测方法。通过这种技术，神经网络使用从几个网站泄露的密码进行训练。在密码生成过程中，神经网络一次输出一个密码字符。每个新字符（包括一个特殊的密码末字符）都是根据给定当前密码的概率选择的，类似于基于马尔可夫模型的方法。（该技术也用于[44]进行实时密码强度估计。）[44]中提出的评估表明，在测试大量密码猜测时，他们的技术优于JTR和HasHashCat常用的密码组合规则$10^{}$至10^{25}一系列

III .GAN架构和超参数

为了利用GAN从训练集中有效地估计密码的概率分布，我们对各种参数进行了实验。在本节中，我们将报告我们对特定的GAN体系结构和超参数的选择。

我们使用古拉贾尼等人的瓦瑟斯坦加纳斯（IWGAN）的改进训练，实例化了PassGAN。[27].本文中使用的IWGAN实现依赖于ADAM优化器[34]来最小化训练误差，i。e.，以减少模型的输出与训练数据之间的不匹配。

我们的模型具有以下超参数的特征：

- 批处理大小，它表示密码数
  在优化器的每一步中通过GAN传播的训练集。
- **迭代次数，表示迭代次数**
  乘以GAN调用其前进步骤和反向传播步骤[38]，[37]。在每次迭代中，GAN运行一个生成器迭代和一个或多个鉴别器迭代。
- **每个生成器的鉴别器迭代次数**
  迭代，它表示生成器在每次GAN迭代中执行了多少次迭代。
- **模型的维数，它表示模型的数量**
  每个卷积层的尺寸（权重）。
- **梯度惩罚系数(λ)，它指定了**
  应用于鉴别器相对于其输入[27]的梯度范数的惩罚。增加这个参数会使GAN [27]的训练更稳定。
- 输出序列长度，表示最大值
  由生成器生成的字符串的长度（此后变为G）。
- 输入噪声向量（种子）的大小，它决定了
  有多少随机比特作为输入到G以生成样本。
- 最大数量，表示
  要加载的最大训练项目数（使用PassGAN时为密码）。
- **Adam优化器的超参数：**
  0学习率，我。e.，它的重量能有多快呢
  模型调整
  0系数β1，它指定了衰减速率
  梯度的运行平均值。
  0系数β2，这表示衰减速率
  梯度的平方的运行平均值。

我们用64的批处理大小实例化了我们的模型。我们使用不同数量的迭代来训练GAN，并最终确定了199,000次迭代，因为进一步的迭代提供了匹配数量的收益减少（见第IV-A节的分析）。*每次生成迭代的鉴别器迭代次数设置为10次，这是IWGAN使用的默认值*。我们对生成器和鉴别器都使用了5个残差层，在两个深度神经网络中的每一层都有128维。

我们将梯度惩罚设置为10，并将GAN生成的序列长度从32个字符（IWGAN的默认长度）修改为10个字符，以匹配训练期间使用的最大密码长度（见第IV-A节）。*由GAN加载的最大示例数量被设置为整个训练数据集的大小*。我们将噪声向量的大小设置为128个浮点数。

系数β1和β2优化器分别设置为0.5和0.9，学习率为10^{-4}.这些参数是古拉贾尼等人使用的默认值。[27].

增值评价

在本节中，我们首先介绍我们的培训和测试程序。然后，我们报告了我们的实验结果，并将PassGAN的输出与JTR和HashCat常用的密码生成规则的输出进行了比较。

我们的实验是使用IWGAN的张量流实现运行的。我们使用了张量流的1.2版本。1个用于gpu，Python2.7版本。12.所有实验都是在运行Ubuntu 16.04.2 LTS的工作站上进行的，该工作站有64GB内存，12核2.0 GHz Intel Xeon CPU和NVIDIA GeForce GTX 1080 Ti GPU。

A. GAN培训和测试

为了评估PassGAN的性能，并将其与最先进的密码生成规则进行比较，我们首先对来自RockYou泄漏[61]的GAN、JTR和HashCoat的大量密码进行了训练。这个数据集中的条目代表了常见密码和复杂密码的混合，因为这些密码以明文存储在服务器上，因此所有密码都被恢复。然后，我们计算了每个工具生成的密码的数量在两个独立的测试集中存在：与训练集不同的RockYou的一个子集，以及领英密码数据集[40]。

RockYou数据集包含32,603,388个密码。我们选择了所有长度超过10个字符的密码（29599,680个密码，对应数据集的90.8%），使用80%（23,679,744个密码，9,925,896个唯一密码）来训练每个工具。在测试中，我们使用了剩下的20%（5,919,936个总密码，3,094,199个唯一密码）。

我们还在领英数据集的所有长度不超过10个字符的密码上测试了每个工具。此数据集
总共由60,065,486个唯一密码组成，其中43,354,871个唯一密码的长度不小于10个字符（领英数据集的频率计数不可用）。领英数据集中的密码以哈希的形式被过滤出来，而不是明文。因此，领英数据集只包含诸如JTR和HashCat等工具能够恢复的明文密码。我们在第IV-B节中展示的结果表明，用于恢复领英密码的规则和数据集与本工作中使用的规则和数据集有很大的重叠。

我们的培训和测试程序使我们能够确定：
(1)在相同的密码分布上进行训练和测试时，PassGAN对密码的预测效果如何。e.，当使用RockYou数据集进行训练和测试时)；(2)PassGAN是否跨密码数据集进行推广。e.，它是如何在RockYou数据集上进行训练，并在领英数据集上进行测试时的表现的。

GAN训练过程对其输出的影响。训练一个GAN是一个由大量迭代组成的迭代过程。随着迭代次数的增加，GAN从数据的分布中学习到更多的信息。然而，增加步数也增加了过拟合[23]，[76]的概率。为了评估密码数据的权衡，我们存储了中间训练检查点，并生成了10个^{8}每个检查点的密码。

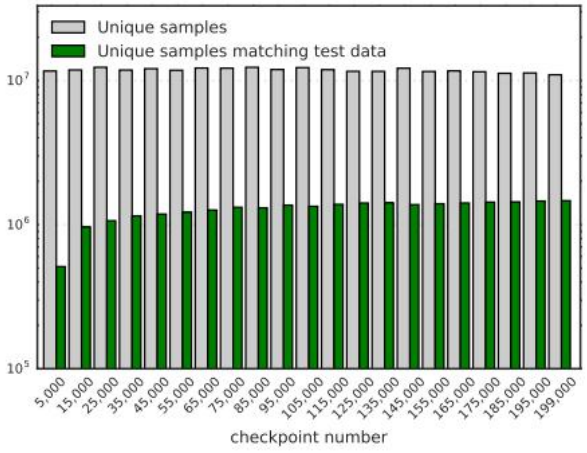

图1：由GAN生成的唯一密码数，以及与RockYou测试集相匹配的密码数。x轴表示PassGAN的训练过程的迭代次数（检查点）。对于每个检查点，PassGAN总共生成了10个^{8}口令

图1显示了GAN在每个检查点上生成的唯一密码的数量，以及这些密码中有多少与RockYou测试集的内容相匹配。由GAN生成的唯一样本的数量与迭代次数（检查点数）保持相当不变。然而，与测试集相匹配的密码的数量会随着迭代次数的增加而稳步增加。这种增长在大约175,000次和199,000次迭代中逐渐减少，其中唯一密码的数量也略有减少。这表明，进一步增加迭代次数可能会导致过拟合，从而降低GAN生成各种极有可能的密码的能力。因此，我们认为这个迭代范围足够适合我们的RockYou训练集。

B. 评估由PassGAN生成的密码

我们生成了多达10个元素^{10}密码使用PassGAN、JTR和HashCat。对于JTR，我们使用了蜘蛛实验室的破坏规则[68]，而对于HashCat，我们使用了最好的64和gen2规则[28]。这些规则通常在密码猜测文献[44]中使用，并且多年来已经在包括RockYou和领英在内的密码数据集上进行了优化。由于这些特定于数据集的优化，我们认为这些规则是使用手动生成的规则可以获得的最佳匹配性能的良好表示。

最好的64代和第2代都能够生成少于10代^{10}密码（分别约998亿和754亿密码，见表1）。对于蜘蛛实验室的混乱规则，我们生成了大约6 • 10个^{10}口令从这个集合中，我们均匀地抽样了528,834,530个唯一的密码。这使得我们可以在JTR和PassGAN之间进行公平的比较，因为后者生成了

表一：使用PassGAN、HashCat和JTR生成的密码的唯一性和新颖性的比较。列(1)显示了使用每个工具生成的密码的总数。在我们的实验中，由于我们使用的训练数据集和规则集，HashCat无法生成像PassGAN和JTR一样多的密码。列(2)显示了每个工具生成的密码的数量是唯一的。列(3)表示每个工具生成的有多少个唯一的密码已经出现在训练集中。

口令 生成工具	(1)合计 口令 生成	(2)唯一密码	(3)匹配的密码 在训练中 （9926278个条目）
帕斯甘	10^6	182,036	27,320 (0.28%)
	10^7	1,357,874	134,647 (1.36%)
	10^8	10,969,748	487,878 (4.92%)
	10^9	80,245,649	1,188,152 (12%)
	$\approx 10^{9.86}$	441,357,719	1,825,915 (18.4%)
	10^{10}	528,834,530	2,177,423 (21.9%)
哈希猫 最佳64条规则	754,315,842	441,357,719	9,898,464 (99.7%)
哈希猫 第二代规则	998,076,164	646,401,854	3,267,236 (32.9%)
JTR间谍实验室规则	$6 \cdot 10^{10}$	528,834,530	2,278,045 (23%)
best64+GAN	10,754,315,842	947,606,924	9,898,807 (99.7%)

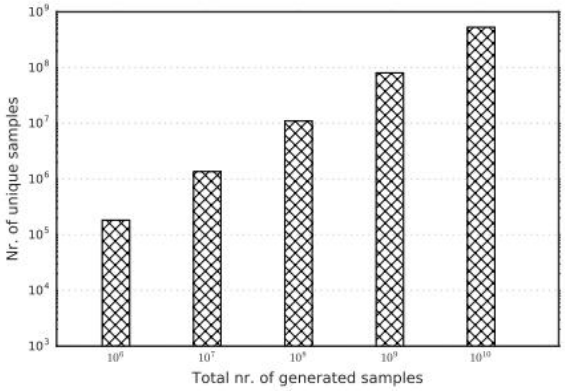

图2：唯一样本数，与PassGAN生成的样本总数相比。

相同数量的唯一密码从一组10个^{10}样品

我们首先确定了增加由PassGAN生成的密码数量是否也增加了唯一密码的数量，以及与训练集和测试集的匹配数量。为此，我们生成了大小在10个密码之间的各种密码集^{6}和10^{10}. 我们观察到，随着密码数量的增加，唯一密码的数量也在增加。该评估的结果见表一第(2)列。如图2所示，当我们增加由PassGAN生成的密码数量时，生成新的唯一密码的速率仅略有下降。同样，如图3所示，匹配数量的增长率随着数量的增加而减少

表二：在RockYou测试集上，使用我们的技术HashCat和JTR生成的密码的比较。列(2)显示了测试集中所生成的密码的数量，该测试集由3,094,199个唯一条目组成。列(3)显示了在测试集中出现的生成密码的数量，通过测试集中对应条目的数量对每个匹配进行加权。（测试集包含5919936个非唯一条目）。列(4)和(5)列报告了在测试集中出现的生成密码的数量，以及不在训练集中的密码(i。*e.*，*新密码*）。第(4)列中的数字将每个匹配视为唯一，而第(5)列中的数字根据测试集中密码的出现次数来计算每个密码匹配的权重。第(4)列的最大匹配数为1,978,367，第(5)列为2,032,728。

口令 生成工具	(1)生成唯一密码	(2)密码 在测试中匹配 集合（唯一）	(3)匹配的密码 在测试集中(与 重复	(4)匹配的密码 在测试中，而不是 在训练集中（唯一）	(5)匹配的密码 在测试集中，而不是在 训练集（重复）
帕斯甘	182,036	17,421 (0.56%)	449,583 (7.59%)	1,850 (0.094%)	2,039 (0.1%)
	1,357,874	76,473 (2.47%)	870,126 (14.7%)	11,398 (0.576%)	12,489 (0.6%)
	10,969,748	236,375 (7.64%)	1,466,336 (24.8%)	54,325 (2.746%)	58,682 (2.88%)
	80,245,649	501,272 (16.2%)	2,133,147 (36%)	162,652 (8.221%)	172,997 (8.51%)
	441,357,719	699,798 (22.6%)	2,373,825 (40.1%)	286,736 (14.49%)	301,416 (14.83%)
	528,834,530	833,434 (26.9%)	2,774,269 (46.9%)	342,439 (17.31%)	359,980 (17.7%)
哈希猫 最佳64条规则	441,357,719	1,744,127 (56.4%)	4,545,600 (76.8%)	630,067 (31.85%)	662,215 (32.577%)
哈希猫 第二代规则	646,401,854	1,288,769 (41.7%)	4,060,366 (68.6%)	448,969 (22.69%)	475,462 (23.39%)
JTR间谍实验室规则	528,834,530	472,417 (15.3%)	1,368,106 (23.1%)	161,807 (8.178%)	170,437 (8.38%)
best64+GAN	**947,606,924**	**1,859,765 (60.1%)**	**4,664,141 (78.8%)**	**745,680 (37.69%)**	**780,705 (38.4%)**

表三：使用我们的技术HashCat和JTR在领英测试集中生成的密码的比较，该测试集包含43,354,871个唯一条目。第(2)列显示了多少生成的密码出现在测试集中，而第列(3)显示了多少生成的密码与测试集匹配，并且不在训练集中（最大匹配数为40,597,129）。

口令 生成工具	(1)生成唯一密码	(2)密码 匹配在 测试集（唯一）	(3)匹配的密码 在测试集中，而不是 在训练集中（唯一）
帕斯甘	182,036	33,794 (0.08%)	12,946 (0.032%)
	1,357,874	185,775 (0.43%)	87,230 (0.215%)
	10,969,748	804,326 (1.86%)	474,861 (1.169%)
	80,245,649	2,341,529 (5.40%)	1,637,122 (4.032%)
	441,357,719	4,185,625 (9.65%)	3,255,417 (8.018%)
	528,834,530	4,996,980 (11.5%)	3,890,043 (9.582%)
哈什猫最好的64条规则	441,357,719	9,930,005 (22.9%)	7,174,986 (17.67%)
哈什卡特 第2代规则	646,401,854	6,271,492 (14.5%)	4,475,775 (11.02%)
颈鼓反射 Spyder实验室规则	528,834,530	2,763,640 (6.37%)	2,023,113 (4.98%)
best64+GAN	**947,606,924**	**11,702,590 (27%)**	**8,947,510 (22.039%)**

生成的密码的数量增加了。这是意料之中的，因为更简单的密码在早期就已经匹配好了，而其余的（更复杂的）密码需要大量的尝试才能被匹配。

我们的实验还表明，与JTR和HashCat规则相比，PassGAN的输出具有更高的密码重复率。因此，在这三种工具中，PassGAN是唯一一个试图生成与训练集相同分布的密码的工具——训练集的特征是有大量重复的密码。这允许PassGAN比不可能更早输出非常可能的密码

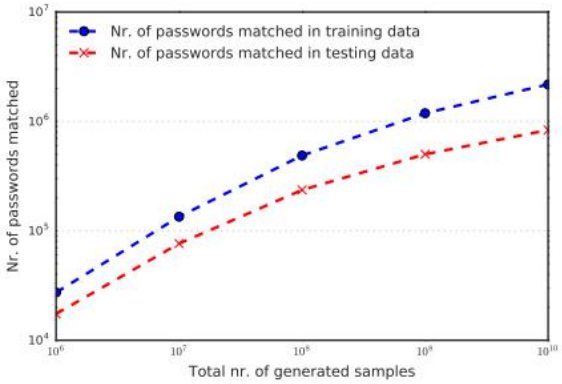

图3：随着GAN输出中的样本数量的增加，出现在训练集和测试集中的GAN生成的密码数量。

密码，因此可能会减少在实践中尝试猜测的次数。

对于每个工具，我们都会报告生成的同样出现在训练集中的密码的数量。我们不认为这些密码是有价值的，因为它们导致的任何匹配都可以通过基于训练集的字典攻击来获得。表一，列(3)，显示了每个工具的结果，而图3（蓝线）提供了与PassGAN的输出相关的进一步细节。在我们的实验中，与JTR和HashCat相比，PassGAN生成的匹配其训练集的密码要少得多。

接下来，我们计算了每个工具生成的密码数量，它们与RockYou和领英测试集相匹配。我们对RockYou的研究结果见表二所示

表四：由GAN生成的与测试集不匹配的密码的示例。

love42743	ilovey2b93	paolo9630	意大利它
悲伤的恶心	usa2598	s13trumpy	trumpart3
ttybaby5	dark1106	vamperiosa	~dracula
悲伤的泪囊	luvengland	阿尔巴尼亚。	bananabake
苍白的年轻	@crepess	emily1015	enemy20
goku476	coolarse18	很酷	serious003
nyc1234	thepotus12	伟大的运行	babybad528
圣塔区	apple8487	1loveyoung	bitchin706
toshibaod	tweet1997b	103眼泪	1holys01

领英网站的检测结果见表三。表二的第(2)和(3)列显示了每个工具生成的RockYou测试集中的密码数量——分别不计数重复匹配和计数重复匹配。表II的第(4)和(5)列报告了当我们排除出现在训练集和测试集中的密码（分别没有匹配和有重复匹配）时的匹配数量。在我们的实验中，最好的64条规则能够比GAN多匹配54%的密码，最好的64条规则比我们的技术多匹配大约1.7 -2倍。××在相同的实验中，PassGAN显著且持续地优于JTR中使用的蜘蛛实验室规则。

在所有指标下，将GAN的输出与最佳64的输出相结合，从而在计算唯一密码时（见表二时匹配115613个额外密码，列(4)），在计算重复时匹配118490个额外密码（见表二，列(5)）。在这两种情况下，这都相当于HashCat的64提高了18%。我们认为这是一个重要的结果，因为它表明GAN能够成功地仅根据密码样本来模拟用户如何选择密码——在某些情况下，比编写HashCat密码生成规则的专家要好。

表三的第(2)和(3)列显示了与领英测试集的匹配数量。列(2)包含所有匹配项，而列(3)不包含同样出现在RockYou训练集中的密码的匹配项。在我们的实验中，HashCat规则和PassGAN之间的差距要小于RockYou数据集。这说明GAN比人工生成的规则更容易进行一般化。此外，将帕斯根与最佳64强结合起来，再次获得了最高的比赛数量。这些结果表明，当没有来自目标分布的样本时，PassGAN提供了实质性的好处，这通常是密码猜测的情况。此外，当我们将PassGAN的输出与HashCat的最佳64条规则的输出结合起来时，我们能够匹配的密码多1772524个密码
增长约24%。

我们检查了由PassGAN生成的与任何测试集都不匹配的密码列表，并确定其中许多密码是合理的候选密码。因此，我们推测，由GAN生成的大量密码可能与我们的测试集不匹配，但可能仍然与来自RockYou以外的服务的用户帐户相匹配。我们在表四中列出了这些密码的一个小示例。

V.结论

本文介绍了第一种基于GANs的密码猜测技术PassGAN。我们的结果表明，当对泄露的密码数据进行训练时，字符级的gan非常适合生成密码猜测。此外，我们的结果表明，当在密码数据集上进行训练，并在不同分布的数据集上进行测试时，GANs可以很好地泛化。

在我们的实验中，当GAN在RockYou的不同子集上进行训练时，我们能够在从RockYou密码数据集中提取的测试集中匹配超过46%的密码。与基于规则的密码生成技术相比，我们的实验表明，PassGAN具有很强的竞争力。尽管HashCat最好的64代和第2代规则优于PassGAN，但我们的方法能够匹配的密码是JTR的间谍实验室规则的两倍。此外，通过将GAN的输出与最佳64条规则的输出相结合，我们能够匹配RockYou数据集中超过78%的密码，以及来自领英数据集的近27%的密码——分别增加了约18%和24%。这是值得注意的，因为它表明，GAN可以生成大量目前最先进的密码生成工具无法获取的密码。此外，当我们在一个不同于训练集（LockYou）的数据集（LinkedIn）上评估每个密码猜测工具时，GAN匹配率的下降就不那么明显。

我们相信，我们的密码猜测方法是革命性的，因为与当前基于规则的工具不同，PassGAN能够在不需要用户干预的情况下生成密码——因此不需要关于密码的域知识，也不需要对密码数据库泄漏的手动分析。此外，我们对训练性能的评估表明，当提供足够大的泄露密码集时，PassGAN的性能可以超过最好的基于规则的密码生成技术。

参考文献

[1] M.阿巴迪。朱，我。好伙伴，H。B.麦克马汉，我。米罗诺夫，K.Talwar和L。张，“具有不同隐私的深度学习”，在*2016年ACM SIGSAC计算机和通信安全会议论文集*。ACM，2016年，页。308 - 318.

[2] A.阿卜杜勒卡德，A。拉克希米拉坦和J。张。（2016）简介深度文本：脸书的文本理解引擎。在线可用：https://tinyurl.com/jj359dv

[3] M.Arjovsky, S.钦塔拉和L。瓦瑟斯坦根，《科尔》，第1卷。abs/1701.07875, 2017.

[4] G.Ateniese, L.V.曼奇尼。斯波格纳迪，A。维拉尼，D。维塔利和G.*费力西，“用智能机器入侵智能机器：如何从机器学习分类器中提取有意义的数据”，《国际安全与网络杂志》，第1卷*。10，没有。3, pp.137 - 150, 2015.

[5] D.Berthelot, T.舒姆和L。“开始：边界平衡生成式对抗网络，”arXiv预印本arXiv：1703.10717，2017.

[6] H.《信息安全威胁、漏洞、信息安全漏洞手册》，预防、检测和管理，第3卷，“2006年”。

[7] L.*《随机森林》，机器学习，第1卷*。45岁，没有。1, pp. 5 - 32, 2001.

[8] N.卡里尼和D。瓦格纳说，“防御性蒸馏对对抗性的例子，”arXiv预印本arXiv：1607.04311,2016。

[9]一一说，“敌对的例子不容易被发现：绕过十个
检测方法”，“arXiv预印本arXiv：1705.07263,2017年。

[10] C.Castelluccia, M.ЏDrmuth和D。“自适应密码”
来自马尔可夫模型的强度计。”2012年，在NDSS网站上。

[11] X.陈，Y。段，R。Houthooft, J.舒尔曼，我。萨特斯克弗和
P.艾文，“信息研究：信息最大化的可解释表示学习”，《神经信息处理系统的进展》，2016年，页。2172－2180.

[12] R.科洛伯特，J。韦斯顿，L。Bottou。卡伦K。卡武库格鲁和
P.库克萨，“自然语言处理（几乎）”，《机器学习研究杂志》，第1卷。12日，没有。8月，页。2493－2537, 2011.

[13] A.A.Cruz-RoaJ。E.A.Ovalle。马达布希和F。A.G.
“一种用于图像表示、视觉可解释性和自动基底细胞癌癌症检测的深度学习架构”
国际医学图像计算和计算机辅助干预会议。施普林格，柏林，海德堡，2013年，第二页。403－410.

[14] M.Dell'Amico, P.米奇亚迪和Y。“密码强度： An
“实证分析”，发表在信息通，2010年论文集。电器和电子工程师学会2010, pp.1－9.

[15] E.L.丹顿，S。钦塔拉，R。费格斯等人。，“深度生成图像
模型使用？《对抗性网络的拉普拉斯金字塔》，《神经信息处理系统的进展》，2015年，页。1486－ 1494.

[16] J.多纳休，L。A.亨德里克斯。瓜达拉马，M。罗巴赫，S。维努戈帕兰，K。Saenco和T。达雷尔，“视觉识别与描述的长期循环卷积网络”，2015年IEEE计算机视觉与模式识别会议（CVPR），第页。2625－2634，
2015.

[17] B.Duc, S.费舍尔和J。“用gabor infor-
《IEEE图像处理事务》，第1卷。8、没有。4, pp.504－516, 1999.

[18] M.ЏDrmuth, F.Angelstorf, C.Castelluccia, D.外围设备和C。Abdelberi,
预兆：使用有序的马尔可夫枚举器可以更快地猜测密码。在ESSoS中。施普林格，2015，页。119－ 132.

[19] R.Fakoor, F.拉达克，A。纳粹和M。“利用深度学习来实现
加强癌症的诊断和分类，”在第30届国际会议上说
机器学习会议（ICML2013年），小麦研讨会，2013年。

[20] S.菲格曼。（2017年）雅虎称有5亿个账户被盗。
在线可用： http://money.cnn。com/2016/09/22/技术/yahoo-data-breach/index.html

[21] M.弗兰克，R。Biedert, E.妈妈，我。马蒂诺维奇和D。Song，《触觉分析：
关于触摸屏输入作为连续认证的行为生物特征的适用性，“IEEE的信息取证与安全事务，卷。8、没有。1, pp.136－ 148, 2013.

[22] M.Fredrikson, S.Jha和T。“模型反演攻击”
利用信心信息和基本对策
第22届ACM SIGSAC计算机和通信安全会议论文集。ACM，2015，pp。1322－ 1333.

[23] I.古德费罗，J。Pouget-Abadie，M。米尔扎，B。徐，D。沃德法利，S.奥扎尔。库尔维尔和Y。“生成对抗网》，《神经信息处理系统的进展》，2014年，页。2672－ 2680.

[24]谷歌深度思维。（2016）阿尔帕戈，第一个计算机程序
曾经在围棋游戏中击败过职业球员。在线可用：
https://deepmind.com/alpha-go

[25] A.格雷夫斯，“用递归神经网络生成序列”，
arXiv预印本，arXiv：1308.0850,2013年。

[26] A.坟墓，Ar。.-穆罕默德和G。辛顿说，“语音识别与……有关”
“深度递归神经网络”，2013年IEEE声学、语音和信号处理国际会议。IEEE，2013，页。6645－
6649.

[27] I.Gulrajani, F.艾哈迈德，M。Arjovsky, V.杜莫林和A。C.考维尔
“瓦瑟斯坦的改进训练”，科尔，卷。abs/1704.00028,
2017.

[28] HashCat.(2017).在线可用： https://hashcat.net

[29] J.海耶斯，L。梅利斯，G。Danezis和E。D.克里斯托法罗，“洛根：Evalu-
利用生成对抗网络处理生成模型的隐私泄漏。abs/1705.07663, 2017.

[30] W.他，J。魏，X。陈，N。卡里尼和D。歌曲，“对抗性的例子
防御：弱防御的集合并不强大，”arXiv预印本arXiv：1706.04701,2017。

[31] B.Hitaj, G.阿特尼斯和F。Perez-Cruz，“GAN下的深度模型：
来自协作性深度学习的信息泄露，”CCS‘17,2017。

[32] P.G.凯利，S。Komanduri, M.L.马祖里克，R。谢伊，T。维达斯，L。鲍尔
N.Christin, L.F.克兰诺和J。洛佩兹，“反复猜测（一次又一次）：通过模拟密码破解算法来衡量密码强度”，发表在安全与隐私（SP），2012年IEEE研讨会上。IEEE，2012，页。523－537.

[33] T.金，M。Cha, H.金，J。李和J。Kim，学习发现
与生成对抗网络的跨领域关系”，arXiv预印本arXiv：1703.05192,2017。

[34] D.金玛和J。《亚当：随机优化方法》
arXiv预印本，arXiv：1412.6980,2014年。

[35] J.Kos和D。歌曲，“深入研究对深层政策的对抗性攻击”，
arXiv预印本，arXiv：1705.06452,2017年。

[36] M.赖，《长颈鹿：使用深度强化学习来下棋》，
arXiv预印本，arXiv：1509.01549,2015年。

[37] Y.莱村，B。博瑟，J。丹克，D。亨德森，R。霍华德，W。槽和L。摘要，“手写数字识别”，在神经信息处理系统的进展2，NIPS 1989。摩根·考夫曼出版社，1990年，第3页。396－404.

[38] Y.莱村，B。博瑟，J。S.丹克，D。亨德森，R。E.霍华德
W.哈伯德和L。D.“反向传播应用于手写邮政编码识别”，神经计算，第1卷。1、没有。4, pp.541－551, 1989.

[39] Y.莱村，K。Kavukcuoglu, C.法拉贝等人。卷积网络
以及在视觉中的应用程序。”在ISCAS，2010，页。253－256.

[40]领英。林克丁。在线可用：https://哈希。org/public.php

[41] Y.刘，X。陈，C。刘和D。歌曲，“深入到转移-
能够对抗的例子和黑盒攻击，”arXiv预印本arXiv：1611.02770,2016。

[42] J.妈妈，W。杨，M。罗和N。李彦，“对概率密码的研究”
模型”，在安全与隐私（SP），2014年IEEE研讨会。IEEE，2014，页。689－704.

[43] P.麦克丹尼尔。Papernot和Z。B.Celik，机器学习
对抗性设置，“IEEE安全与隐私，卷。14日，没有。3, pp.68－ 72, 2016.

[44] W.梅利切尔，B。Ur, S.M.Segreti。Komanduri, L.鲍尔
N.克里斯汀和L。F.鹤，“快速、精益、准确：使用神经网络建模密码的可猜测性”，第25届USENIX安全研讨会（USENIX安全16）。奥斯汀，德克萨斯州： USENIX协会，2016年，页。175－ 191.在线可用性：https://www.usenix.org/
conference/usenixsecurity16/technicalsessions/presentation/melicher

[45] C.梅斯（2016）谷歌的GO胜利只是一个一瞥
强大的ai将是。在线可用：https://tinyol。com/l6ddhg9

[46] M.米尔扎和S。“条件生成对抗网”，
arXiv预印本，arXiv：1411.1784,2014年。

[47] V.Mnih, K.Kavukcuoglu, D.银，A。格雷夫斯，我。Antonoglou, D.维尔
斯特拉和M。A.瑞德米勒，《用深度强化学习演奏雅式》，第1卷。abs/1312.5602, 2013.

[48] R.莫里斯和K。汤普森，《密码安全：一个案例的历史》，
ACM的通信，vol。22日，没有。11, pp.594－597, 1979.

[49] A.纳拉亚南和V。“快速字典攻击密码”
他在第12届ACM计算机和通信安全会议论文集中说。ACM，2005，pp。364－372.

[50] Y.平移，T。Mei, T.姚，H。李和Y。瑞伊，“联合建模嵌入”

“翻译为视频与语言的桥梁”，“2016 IEEE计算机视觉与模式识别会议（CVPR），页。4594 - 4602，
2016.

[51] N.Papernot, P.麦克丹尼尔和我。古德费罗，《马-》中的可转移性》中国学习：从现象到黑盒攻击，”arXiv预印本arXiv：1605.07277,2016。

[52] N.Papernot, P.麦克丹尼尔。Jha, M.Fredrikson, Z.B.Celik和A.Swami，“对抗性环境下深度学习的局限性”，发表在第一届IEEE欧洲安全与隐私研讨会论文集，2015年。

[53] N.Papernot, P.麦克丹尼尔，X。吴，S。Jha和A。斯瓦米，“蒸馏作为对针对深度神经网络的对抗性扰动的防御，”在第37届IEEE安全与隐私研讨会的论文集中，2015.

[54] C.珀西瓦尔和S。“基于密码的密钥派生功能，“技术。众议员，2016年。

[55] S.佩雷斯（2017）谷歌计划推出密码-在年底前免费登录到安卓应用程序。在线可用性：https://techcrunch.com/2016/05/23/ google-plans-to-bring-password-free-logins-to-android-apps-by-year-end/

[56] H.P.位置马尔可夫链。(2017).在线可用： https://www。trustwave.com/Resources/SpiderLabs-Blog/每个位置的马尔可夫链/

[57] T.P.项目(2017).在线可用的内容是：http://密码项目。com/leaked_密码_清单_和_词典

[58] N.普罗沃斯和D。《隐窝算法》，发表在USENIX，1999年。

[59] A.拉德福德，L。Metz和S。“无监督的代表”*深度卷积生成对抗网络学习，第四届国际学习表征会议，2016。*

[60] C.E.拉斯穆森和C。K.*机器的高斯过程学习*麻省理工学院出版社，剑桥出版社，2006年，第1卷。1.

[61] RockYou.(2010) Rockyou.在线可用：http://下载。skullsecurity.org/passwords/rockyou.txt.bz2

[62] H.规则。(2017).在线可用性： https://github.com/hashcat/hashcat/tree/master/rules

[63] J.T.R.K.规则。(2017).在线可用：http://比赛-2010。korelogic.com/rules.html

[64] B.Schlkopf和。J.*Smola，用内核学习：支持向量机器，正则化，优化，等等*。麻省理工学院出版社，2002年。

[65] R.Shokri和V。《保护隐私的深度学习》，在*第22届ACM SIGSAC计算机和通信安全会议记录*。ACM，2015，pp。1310 - 1321.

[66] R.Shokri，M。斯特罗纳蒂，C。宋和V。什马蒂科夫，“会员针对机器学习模型的推理攻击”，在安全与隐私（SP），2017年IEEE研讨会上继续讨论。IEEE，2017，页。3 - 18.

[67] Z.Sitov, J.ednka, Q.杨，G。彭，G。周，P。Gasti和K。S.“面向智能手机用户持续认证的新行为生物特征”，IEEE《信息取证与安全学报》，第1卷。11日，没有。5, pp.877 - 892, 2016.

[68] T.蜘蛛实验室。（2012）韩国的规则。在线可用： https://github.com/SpiderLabs/KoreLogic-Rules

[69] I.苏特斯克弗，J。马滕斯和G。E.“生成文本”*《递归神经网络》，第28届机器学习国际会议论文集（ICML-11），2011年，第3页*。1017 - 1024.

[70] Y.泰格曼，M。杨，M。兰扎托和L。沃尔夫，“深面：在2014年IEEE计算机视觉与模式识别会议论文集上，ser。CVPR ’14.美国华盛顿特区： IEEE计算机学会，2014年，页。1701 - 1708.在线可用：http://dx.doi.org/10.1109/CVPR.2014.220

[71] J.开膛手。(2017).在线可用： http://www。openwall.com/john/

[72] J.开膛手马尔可夫发生器。(2017).在线可用：http://openwall.info/wiki/john/markov

[73] I.Tolstikhin, S.果冻，哦。Bousquet，c.j。西蒙-加布里埃尔和B.“阿达根：增强生成模型”，arXiv预印本arXiv：1701.02386,2017年。

[74] F.Tramr, F.张，A。朱尔斯，M。K.Reiter和T。“偷窃”通过预测apis的机器学习模型。”2016年，在USENIX网站上。

[75] M.Weir，S。AggarwalB。De Medeiros和B。Glodek，密码使用概率上下文无关语法破解”，“安全与隐私，2009年第30届IEEE研讨会。IEEE，2009，页。391 - 405.

[76] Y.吴，Y。Burda，R。萨拉丘蒂诺夫和R。格罗斯，“在泉上”基于解码器的生成模型的标题分析”，“arXiv预印本arXiv：1611.04273,2016年。

[77] V.桑特德斯基，密歇根州。.-尼科拉和A。拉瓦特，“有效的防御”对抗性攻击，”arXiv预印本arXiv：1707.06728,2017。

[78] H.张，T。徐，H。李，S。张，X。黄，X。王和D。Metaxas,“斯塔克根：文本到具有堆叠生成对抗网络的逼真图像合成，”arXiv预印本arXiv：1612.03242,2016。

[79] X.张和Y。A.勒存，“从头开始的文本理解”，arXiv*预印本，arXiv：1502.01710v5,2016年*。

[80] Y.钟，Y。邓和A。K.Jain，“针对用户的击键动态”在计算机视觉和模式识别研讨会中*（CVPRW），2012年IEEE计算机协会会议上*。IEEE，2012，页。117 - 123.