中国科学院大学网络空间安全学院专业普及课

2021-2022学年秋季学期

Web安全技术 Web Security

授课团队:刘奇旭、刘潮歌

助 教: 陈艳辉、杨毅宇、李寅

授课团队名单:刘奇旭、刘潮歌

Web安全技术

Web Security

课程大作业

刘奇旭

liuqixu@iie.ac.cn 中科院信工所 第六研究室



课程成绩

□ CTF之Web安全中期考核: 10%

□ CTF之Web安全期末考核: 10%

□ 大作业: 30%

□ 期末考试: 50%

课程成绩

大作业

口 大作业: 30%

• 大作业一个

• 分组进行

• 要求: 认真思考, 按时提交

• 布置时间: 待选课结束之后

PASSWORD VS. CIPHER

- A *password*, sometimes called a passcode, is a memorized secret used to confirm the identity of a user. Using the terminology of the NIST Digital Identity Guidelines, the secret is memorized by a party called the claimant while the party verifying the identity of the claimant is called the verifier. When the claimant successfully demonstrates knowledge of the password to the verifier through an established authentication protocol, the verifier is able to infer the claimant's identity.
- In cryptography, a *cipher* (or cypher) is an algorithm for performing encryption or decryption—a series of well-defined steps that can be followed as a procedure. An alternative, less common term is encipherment. To encipher or encode is to convert information into cipher or code. In common parlance, "cipher" is synonymous with "code", as they are both a set of steps that encrypt a message; however, the concepts are distinct in cryptography, especially classical cryptography.



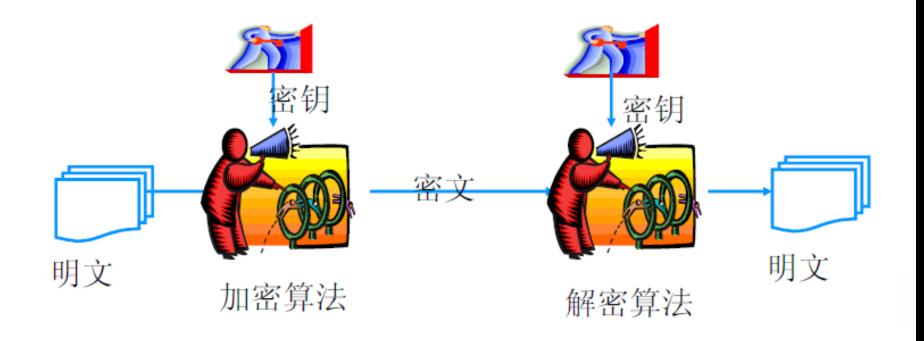
密码学要解决的最基本问题

- □ 信息的保密传输和存储问题
- □ 信息的认证问题
- □ 我收到你写给我的信, 我会问:
 - 信的内容是否被改动?
 - 是否真是你写的信?
 - 是否真是写给我信?
 - 有没有人看过这封信?



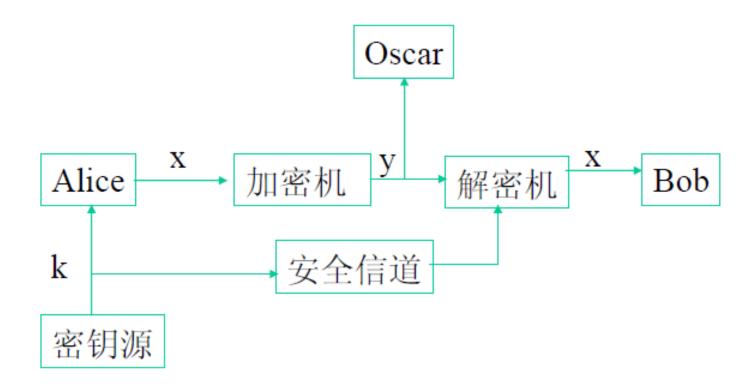
密码学的基本概念

- □ 明文(Plaintext):被隐蔽的消息称作明文。
- □ 密文(Ciphertext): 将明文隐蔽后的结果称作密文或密报。
- □ 加密(Encryption):将明文变换成密文的过程。
- □ 解密(Decryption): 由密文恢复出明文的过程。
- □ 密钥(Key): 控制或参与密码变换的可变参数称为密钥



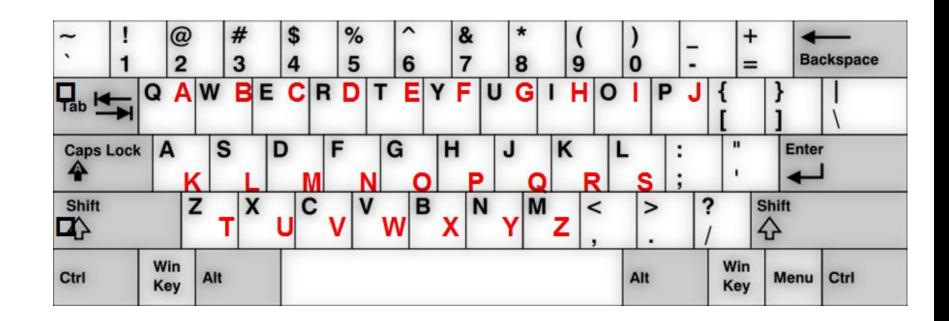
加密通信的模型

□ 密码学的目的: Alice和Bob两个人在不安全的信道上进行通信,而破译者Oscar不能理解他们通信的内容。



加密技术两个基本模块: 代换

- □ 代换法:将明文字母替换成其他字母、数字或符号的方法。
- 如果把明文看做是二进制序列的话,那么代换就是用密文位串来代换明文位串。



加密技术两个基本模块: 置换

- 口 代换是将明文字母代换为密文字母,而与之极不相同的一种加密方法 是置换。置换对明文进行置换。
- □ 最简单的例子就是栅栏技术,按照对角线的顺序写出明文,而按行的 顺序读出作为密文。
- □ 例如深度为2的栅栏技术加密信息 "meet me after the toga party" . 可写为:

□ 那么加密后的信息为: MEMATRHTGPRYETEFETEOAAT





大作业内容: 方案一

Password



- 口令(Password)常用于进入: 计算机、加密文件、邮箱、社交网站、网银、电子 商务、即时通工具等。
- □ 每个人一般拥有2~6个口令,用得最多的口令包括: password, 123456, iloveyou, qwert, football等。



用户名 "admin"、密码 "123456",乌克兰军方系统安全问题被曝光

△ Andy.i ② 2018-09-27 共10416人围观,发现 1 个不明物体 资讯

"123456、admin"在2017年弱密码TOP 100中,分别位列第一位和第十一位。大多数账户系统在注册时基本禁止使用这种"网红弱密码",你很难想象这竟然会成为一个国家军 方系统的用户名和密码。



9月25日,乌克兰记者 @alexdubinskyi 爆料称,乌克兰武装部队的第聂伯罗军队自动化控制系统的服务器用户名和密码分别为"admin"、"123456"。据了解,该系统主要用于 协调顿巴斯地区的军事行动,也就是从2014年至今乌克兰和俄罗斯频繁交火的地区。

最先发现这个问题的是一名数据专家 Vlasyuk Dmitry,在5月22日对该系统进行网络测试的时候发现,很多服务器可以通过简易的用户名(admin)和密码(admin或者123456)访 问,他及时汇报了这个安全隐患,但一段时间之后这个问题并没有得到改善。

5月25日,Vlasyuk 在邮件服务器上发现了类似的情况,基本上不需要技术很高深的黑客就能够轻松访问交换机、路由器、服务器、打印机和扫描仪等设备,能够分析出武装部队大量 的机密信息甚至掌握整个夏天乌克兰军队在顿巴斯地区的一切计划。

口令泄漏事件



□ 2011-12-21: 有网友爆料称,今天有黑客在网上公开了知名网 站CSDN的用户数据库,这是一次严重的暴库泄密事件,涉及 到的账户总量高达600万。

```
0, \dots, 1, \dots, 4, 0, \dots, 7, \dots, 2, 0, \dots, 1, \dots, 3, 0, \dots, 1, \dots, 4, 0, \dots, 7, \dots, 5, 0, \dots, 1, \dots, 1
                              lg # 12344321 # zdg@ dn.net
                              oZheng # 6702033137
                                                                                                                                               / # chengmeng zheng@163.com
                              tao # 730413 # fsta tom.com
                             wolf # 2535263 # hu ve@263.ne
                                                                                                                                               edcil@21c .com
                            dcjl # KIC43dk6! #
                            tsky # s12345 # son hail@21cn. om
                           .chael # apple # app .lp@neteas .com
                           .clj # 1j7202 # junl peoplemai .com.cn
                             nbuhuan # 12345 # j buhuan@16 .net
                            .e # hebeibdh # fwg@ fw.com
                            inroad # 8398518 # inroad@pu lic.cta.cq.cn
11
                              .ver99 # priverhe #
                                                                                                                                         river1999@etease.com
                            igong # 12345 # lei ng@21cn.com
                            azyDragon # kingdom chzhy102 3.net
                            henzhuangyuan # wangene # chzhy @263.net csdn 1
                            s # today # rss@tjm .1.com
                            .an # 6crx99tj # ala x@21cn.cc csdn 1
                             ruila # smart1010 # ruila@21cm com
                            nion # 980527 # onic hm@sina.com
                             ıva # zzzz # zava@16 net
```



口令泄漏事件——爆发

- 口 21日16时左右, 网友"潇洒小姐燕子"在微博上列出一长串名 单,除了CSDN,有同样问题的大网站还有CNBETA、CNZZ、 ENET硅谷动力、百合网、珍爱网、开心网、YY、酷6等,并上 传了一个类似数据库的东西作为证据。
- □ 截至22日17时,被点名的网站均没有对此回应











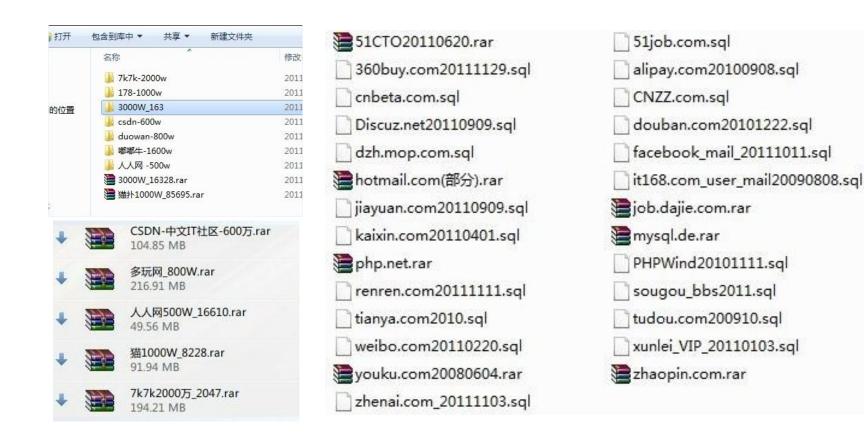
163 网易免费邮

中文邮箱第一品牌





口令泄漏事件——后果

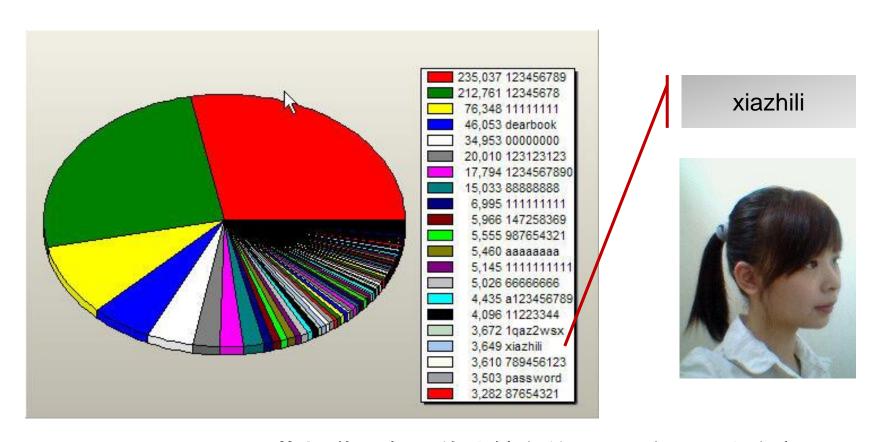




0839X6M05006H. 刘奇旭. Web安全技术: 大作业

口令背后那点事儿

——CSDN最常用口令



xiazhili 是csdn下载频道09年上传比较多的一MM会员,众多程序员yy的对象,故xiazhili作为密码不足为奇。

中国科学院大学 University of Chinese Academy of Sciences

CSDN——谁是你的最爱

□ CSDN杯我最喜爱的CSDN密码活动,统计了一下包含ilove*** 的密码,其中被爱得死去活来的前20位如下:

排名	ilove***	口令数量
1	you	3567
2	you1314	110
3	U	92
4	csdn	68
5	you123	57
6	china	49
7	you520	43
8	thisgame	41
9	java	31
10	chu	20

排名	ilove***	口令数量
11	123	30
12	2hcx	29
13	you1	26
14	myself	22
15	you521	22
16	r	19
17	jay	19
18	qsr	18
19	520	18
20	1314	18

CSDN——爱的就是你

排在最前的国家是: China

排在最前的网站是: csdn

排在最前的编程语言是: Java

排在最前的操作系统是: I inux

排在最前的电脑品牌是: apple

排在最前的明星是: 周杰伦

排在最前的动物分别是:

- 1. 猪
- 2. 猫
- 3. 猪头
- 4. 狗

排在最前的mm名字是

- yan(燕)
- mei(梅)
- jing(静)
- ling (玲)
- 5. ping (萍)
- 6. dan (丹)
- fang(芳)
- 8. wei(薇)
- 9. juan (娟)
- 10. nana (娜娜)



CSDN——爱的就是你



暗恋了她好几年,一直不敢表白。 后来,他下载到一份CSDN泄露的 用户名密码名单,习惯性的查找她 的邮箱、发现她使用的是ilove加自 己名字的拼音作为密码,正感动得 无以复加时电话响起, 只听她在电 话那头颤抖着说:傻瓜,我看到了 你的密码。——程序猿伤不起 啊。。故事好感人啊。。

翻译自中文

2017/9/27 上午9:43



CSDN—EDU.CN

排名	域名	数量	归属
1	mail.ustc.edu.cn	2035	中国科学技术大学
2	sjtu.edu.cn	1874	上海交通大学
3	bjtu.edu.cn	1322	北京交通大学
4	fudan.edu.cn	981	复旦大学
5	stu.xjtu.edu.cn	929	西安交通大学
6	zju.edu.cn	872	浙江大学
7	mails.tsinghua.edu.cn	718	清华大学
8	bit.edu.cn	691	北京理工大学
9	mail.nankai.edu.cn	640	南开大学
10	stu.edu.cn	554	汕头大学



CSDN——口令生成的艺术

□ CSDN的600万用户数据被泄露后,有人发现其中几个最牛的密码:

鱼和熊掌不可兼得

hold?fish:palm

hanshansi.location()!∈[gusucity]

FLZX3000cY4yhx9day

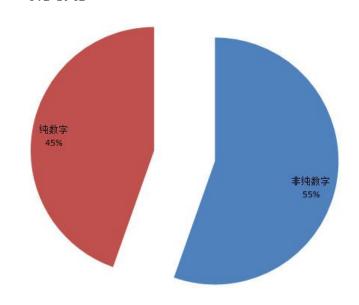
姑苏城外寒山寺

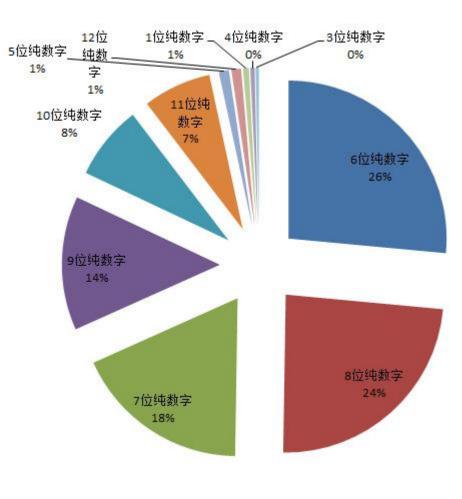
飞流直下三千尺,疑似银河下九天



全部泄露口令——纯数字的比例

45%之巨(大约1700万)的 被统计用户选择了纯数字作 为他们的密码,这反应出国 内大多数网民安全意识还相 当淡薄

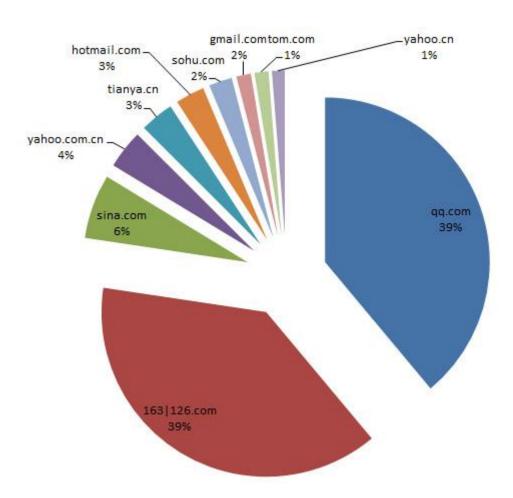






全部泄露口令——纯数字的比例

口 网易和腾讯几乎瓜分了 80%的市场份额,在本 次统计中两者差距在10 万以内





规律性?

- □ 为了便于记忆,□令结构和设置方法一般都具有规律性
 - 口令结构规律
 - 口令设置方法规律
 - 例如: password123! 规律L₈D₃S₄
 - MySpace口令统计规律
 - L₆D₁是最常见的口令结构,占9.98%,例如loveme8
 - 。1是使用最多的数字,占所有数字的27.47%
 - 。!是使用最多的特殊字符,占38.46%



口令安全研究

用户脆弱口令行为

用户倾向性构造模式选择

口令重用

基于个人信息构造口令

口令猜解攻击算法

"奇思妙想"

概率模型科学化算法

PCFG: 概率上下文无关方法 (probabilistic context-free grammars)

Markov-Chain: 马尔科夫链

Natural Language Processing:自然语言处理技术



口令猜解攻击

漫步攻击

漫步攻击(trawling attacking)是指攻击者不关心具体的攻击对象是谁,其唯一的目标是在允许的猜测次数下,猜测出越多的口令越好。

漫步攻击算法:

启发式算法

PCFG

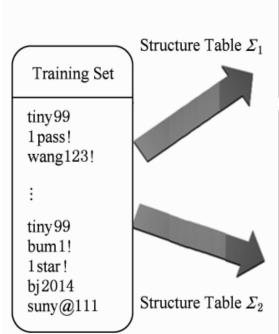
Markov

NLP



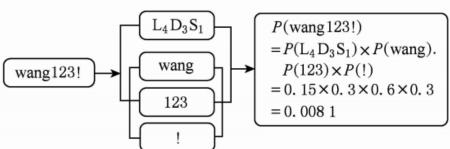
举例: PCFG

- □ 2009年,第一个完全自动化的、建立在严密的概率上下文无关文法(Probabilistic Context Free Grammar, PCFG)基础之上的漫步口令猜测算法。
- □ 该算法的核心假设是口令的字母段L、数字段D和特殊字符段S是相互独立的,然后进行训练和猜测集生成。算法过程中最关键的是统计出口令模式频率表和字符组件(语义)频率表。



S→	L_4D_2	0. 20
S→	$D_1L_4S_1$	0.15
S→	$L_4D_3S_1$	0.15
:	:	::

$D_3 \rightarrow$	123	0.6
$D_3 \rightarrow$	111	0.4
$D_2 \rightarrow$	99	0.7
$D_2 \rightarrow$	11	0.3
$D_1 \rightarrow$	1	0.8
$D_1 \rightarrow$	2	0.2
$S_1 \rightarrow$	@	0.7
$S_1 \rightarrow$!	0.3
$L_4 \rightarrow$	wang	0.3
$L_4 \rightarrow$	suny	0.2
:	:	:



这样就能获得每个字符(猜测)的概率,按照概率递减排序即可获得一个猜测集。



参考文献

计算机研究与发展 Journal of Computer Research and Development DOI:10.7544/issn1000-1239.2016.20160483 53(10): 2173-2188, 2016

口令安全研究进展

王平1,3 汪定1 黄欣沂2

- 1(北京大学信息科学技术学院 北京 100871)
- 2(福建师范大学数学与计算机科学学院 福州 350117)
- ³(北京大学软件与微电子学院 北京 102600) (wangdingg@pku.edu.cn)

Advances in Password Security

Wang Ping^{1,3}, Wang Ding ¹, and Huang Xinyi²

- ¹(School of Electronics Engineering and Computer Science, Peking University, Beijing 100871)
- ²(School of Mathematics and Computer Science, Fujian Normal University, Fuzhou 350117)
- ³ (School of Software and Microelectronics, Peking University, Beijing 102600)



口令猜解攻击

定向攻击

定向攻击(targeted attacking)中攻击者会利用与攻击对象相关的个人信息(personal information,PI),以增强猜测的争对性。

个人信息有很多种,比如人口学相关信息(姓名、生日、职业等), 用户在该网站的过期口令<u>(旧口令),用户在其他网站泄露的口令。</u>

当前定向口令猜测研究尚在起步阶段。

定向攻击算法:

Targeted-Markov

Personal-PCFG

其核心是训练之前先将PI分类(用户名 A、邮箱 E、姓名N、生日B、手机号P、身份证G等),并将其看作是L、D、S一样的组件,猜测生成阶段仿照传统Markov和PCFG方法即可。

网络与系统安全顶级会议

S&P (Oakland)	IEEE Symposium on Security and Privacy
CCS	ACM Conference on Computer and Communications Security
Security	Usenix Security Symposium
NDSS	ISOC Network and Distributed System Security Symposium



Safer Sign-Ons

Harbor DEF

Session Chair: Tadayoshi Kohno, *Microsoft Research and University of Washington*

Password Managers: Attacks and Defenses

David Silver, Suman Jana, and Dan Boneh, *Stanford University;* Eric Chen and Collin Jackson, *Carnegie Mellon University*

The Emperor's New Password Manager: Security Analysis of Web-based Password Managers

Zhiwei Li, Warren He, Devdatta Akhawe, and Dawn Song, University of California, Berkeley

SpanDex: Secure Password Tracking for Android

Landon P. Cox , Peter Gilbert, Geoffrey Lawler, Valentin Pistol, and Ali Razeen, Bi Wu, and Sai Cheemalapati, Duke University

SSOScan: Automated Testing of Web Applications for Single Sign-On Vulnerabilities

Yuchen Zhou and David Evans, University of Virginia

Passwords

Harbor DEF

Session Chair: David Wagner, *University of California, Berkeley*

A Large-Scale Empirical Analysis of Chinese Web Passwords

Zhigong Li and Weili Han, Fudan University; Wenyuan Xu, Zhejiang University

Password Portfolios and the Finite-Effort User: Sustainably Managing Large Numbers of Accounts

Dinei Florêncio and Cormac Herley, *Microsoft Research;* Paul C. van Oorschot, *Carleton University*

Telepathwords: Preventing Weak Passwords by Reading Users' Minds

Saranga Komanduri, Richard Shay, and Lorrie Faith Cranor, *Carnegie Mellon University;* Cormac Herley and Stuart Schechter, *Microsoft Research*

Towards Reliable Storage of 56-bit Secrets in Human Memory

Joseph Bonneau, *Princeton University;* Stuart Schechter, *Microsoft Research*

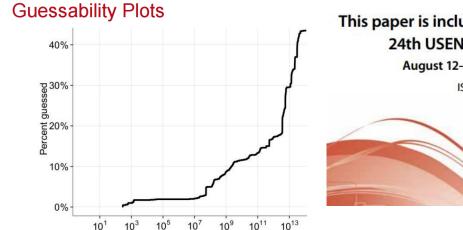




Measuring Real-World Accuracies and Biases in Modeling Password Guessability

Blase Ur, Sean M. Segreti, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor,
Saranga Komanduri, and Darya Kurilova, *Carnegie Mellon University*; Michelle L. Mazurek, *University of Maryland*; William Melicher and Richard Shay, *Carnegie Mellon University*

https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/ur



This paper is included in the Proceedings of the 24th USENIX Security Symposium

August 12-14, 2015 • Washington, D.C.

ISBN 978-1-931971-232

Open access to the Proceedings of the 24th USENIX Security Symposium is sponsored by USENIX



Guesses



Fast, Lean, and Accurate: Modeling Password **Guessability Using Neural Networks**

William Melicher, Blase Ur, Sean M. Segreti, Saranga Komanduri, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor, Carnegie Mellon University

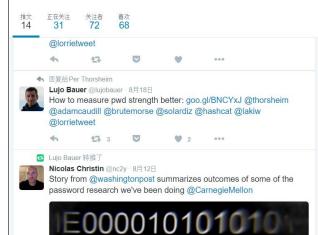
https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/melicher

This paper is included in the Proceedings of the 25th USENIX Security Symposium

August 10-12, 2016 • Austin, TX

ISBN 978-1-931971-32-4







1000PASSWOR

On the Accuracy of Password Strength Meters

Maximilian Golla Ruhr University Bochum Bochum, Germany maximilian.golla@rub.de Markus Dürmuth Ruhr University Bochum Bochum, Germany markus.duermuth@rub.de

A METER COMPARISON

In the following, we list the full results of our data collection. We separated the five categories *Academic Proposals*, *Password Managers*, *Operating Systems*, *Websites*, and *Previous Work* into two tables. A colorful version that allows easier comparison can be found online [29].

A.1 Academic Proposals, Password Managers, and Operating Systems

Table 5: We computed the weighted Spearman correlation as the best similarity score (cf. Section 5). The table lists the online use case on the left, the offline use case on the right. We highlighted if and on how many bins a meter quantized its output. Additionally, we list whether a meter runs on client- or server-side and how the meter visualizes the strength to the user.

					Online Attacker		Offline Attacker		cker	
ID	Meter	Type	Quant.	Visu.	RockYou	LinkedIn	000Webhost	RockYou	LinkedIn	000Webhost
					Academic Proposals					
1A	Comprehensive8 [61]	С	-	-	-0.652	-0.589	0.251	-0.476	-0.616	0.441
1B	Comprehensive8 [61]	C	Q5	Text	-0.331	-0.084	0.409	-0.128	-0.123	0.421
2	Eleven [22]	C	-	-	0.670	0.912	0.492	0.755	0.951	0.733
3	LPSE [32]	C	Q3	-	0.584	0.669	0.508	0.544	0.718	0.693
4A	Markov (OMEN) [13]	S	-	-	0.721	0.697	0.410	0.701	0.669	0.660
4B	Markov (Single) [27]	S	-	-	0.718	0.998	0.817	0.828	0.991	0.872
4C	Markov (Multi) [27]	S	-	-	0.721	0.998	0.902	0.997	0.995	0.777
5A	NIST [11]	C	-	-	0.670	0.912	0.492	0.755	0.951	0.733
5B	NIST (w. Dict.) [11]	C	-	-	0.669	0.910	0.472	0.756	0.953	0.816
6	PCFG (fuzzyPSM) [65]	S	-	-	1.000	0.994	0.963	0.998	0.999	0.899
7A	RNN Generic [46]	C	-	-	0.632	0.542	0.427	0.535	0.520	0.800
7B	RNN Generic (Web) [59]	C	-	-	0.473	0.649	0.421	0.449	0.688	0.777
7C	RNN Target [46]	C	-	-	0.951	0.913	0.965	0.896	0.860	0.885
7D	RNN Target (w. Bloom) [46]	C	-	-	0.951	0.913	0.965	0.896	0.860	0.882
8A	zxcvbn (Guess Number) [71]	С	-	-	0.989	0.990	0.554	0.989	0.999	0.868
8B	zxcvbn (Score) [71]	С	Q5	-	0.341	0.490	0.359	0.373	0.567	0.817



30th USENIX Security Symposium



Reducing Bias in Modeling Real-world Password **Strength via Deep Learning and Dynamic Dictionaries**

Dario Pasquini, Sapienza University of Rome, Institute of Applied Computing CNR; Marco Cianfriglia, Institute of Applied Computing CNR; Giuseppe Ateniese, Stevens Institute of Technology; Massimo Bernaschi, Institute of Applied Computing CNR

https://www.usenix.org/conference/usenixsecurity21/presentation/pasquini

This paper is included in the Proceedings of the 30th USENIX Security Symposium.

August 11-13, 2021

978-1-939133-24-3



Open access to the Proceedings of the 30th USENIX Security Symposium is sponsored by USENIX.



Password Leakage



English



Chinese



Over 100 million plaintext passwords

作业要求

分析目标

国内



600W

国外



73W

快讯 | 雅虎承认其30亿用户信息全部被黑

Akane ② 2017-10-05 共14246人围观

咨讯.

近日,已被Verizon电信收购的雅虎公司宣布2013年8月发生的大规模数据泄露事件,影响范围包括所有雅虎用户,这意味着全球30亿雅虎账户无一幸免。据悉,雅虎最初只披露了10亿用户账号被泄露,后续也有陆续增加,但昨天的公告表明,如果你有一个雅虎账号,那么很不幸,你也被曝光了。

去年12月15日,雅虎承认,其系统曾在2013年遭到黑客攻击,约10亿账号的**用户姓名、生日、邮箱地址、密码、电话、安全问题和答案**全被泄露。除2013年的重大泄露事件以外,雅虎2014年也有至少5亿用户资料落入黑客之手。



作业要求

分组: 每组1-10人

- 1、基础分析(选择下列内容中的三种进行分析):
 - ① 密码构成元素分析(数字、字符、字母等)和结构分析,得到 密码中这些基本元素常用的组合方法;
 - ② <mark>键盘密码的模式分析</mark>,键盘密码就是基于键位变化的一类密码 ,比如asdfgh;
 - ③ 日期密码及其格式分析,难点在于识别日期密码,区分不同的 日期格式,日期与其他字符混排的组合方式统计
 - ④ 拼音的使用统计, Top10, 大小写等, 难点是如何识别
 - ⑤ 英文单词的使用统计, Top10, 大小写等, 难点是如何识别
 - 6 其他,可以自己想
 - ⑦ 其他,可以自己想
 - ⑧ 其他,可以自己想
- 2、基于分析结果,实现一种口令猜测攻击算法,并生成口令字典。 (推荐编程语言: Python)



分组: 每组1-10人



口令相关最新顶会论文



大作业内容: 方案三

人员组成:每组1-10人

超大自选



网络空间安全领域更值得交流分享的内容

(选题需征得授课老师同意)



大作业时间节点

2021

		十月(Oct)				十一月(Nov)			
	5	6	7	8	9	10	11	12	13
星期一	27	4	11	18	25	1	8	15	22
星期二	28	5	12	19	26	2	9	16	23
星期三	29	6	13	20	27	3	10	17	24
星期四	30	7	14	21	28	4	11	18	25
星期五	1 体	8	15	22	29	5	12	19	26
星期六	2	9	16	23	30	6	13	20	27
星期日	3	10	17	24	31	7	14	21	28









作业提交

□ 确定分组、各组自行指定组长

• 2021年10月10日之前,组长联系教师助教-陈艳辉,将组员名单发给助教,分组名单

包括姓名、学号。

□ 提交时间: 2021年11月06日23: 00

□ 提交内容:

• 【必须】分组名单-指定组长

【必须】课堂交流PPT

• 【必须】相关程序源代码、程序使用说明

• 【可选】附属材料(相关视频、参考资料等)

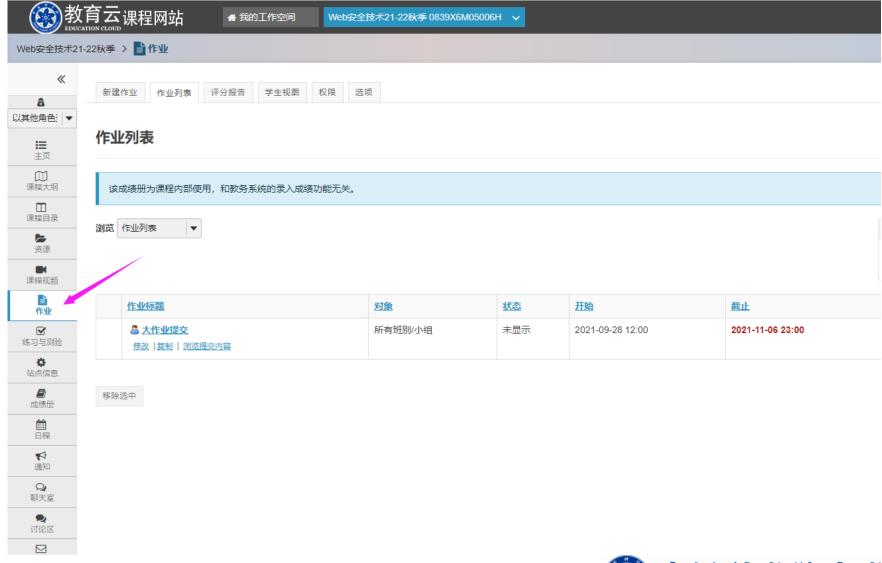
□ 格式要求

• <u>各组组长</u>以压缩包的形式上传到课程网站,组员无需上传。若压缩包太大,可以单独 联系教师助教陈艳辉,通过其他方式提交大作业。



作业提交

课程网站



谢谢大家

刘奇旭 liuqixu@iie.ac.cn 中科院信工所 第六研究室

