Observing passwords in their natural habitat

Sarah Pearman, Jeremy Thomas, Pardis Emami Naeini, Hana Habib,
Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, Alain Forget


12345
123456789
password
iloveyou
princess
1234567
rockyou
12345678
abc123
nicole
daniel
babygirl
monkey
lovely
jessica
654321
michael
ashley
qwerty
111111
000000
michelle
tigger
sunshine
chocolate
password1
soccer
anthony
friends
butterfly
purple
angel
jordan
liverpool
justin
loveme
fuckyou
123123
football
secret
andrea
carlos
jennifer
joshua
bubbles
1234567890
superman
hannah
amanda
loveyou
pretty
basketball
andrew
angels
tweety
flower
playboy
hello
elizabeth
hottie
tinkerbell
charlie
samantha
barbie
chelsea
lovers
teamo
jasmine
brandon
666666
shadow
melissa
eminem
matthew
robert
danielle
forever
family
jonathan
987654321
computer
whatever
dragon
vanessa
cookie
naruto
summer
sweety
spongebob
joseph
junior
softball
taylor
yellow
daniela
lauren
mickey
princesa
alexandra
alexis
jesus
estrella
miguel
william
thomas
beautiful
mylove
angela
poohbear
patrick
iloveu
sakura
adrian
alexander
destiny
christian
121212
sayang
america
dancer
monica
richard
112233
princess1
555555
diamond
carolina
steven
rangers
louise
orange
789456
999999
shorty
11111
nathan
snoopy
gabriel
hunter
cherry
killer
sandra
alejandro
buster
george
brittany
alejandra
patricia
rachel
tequiero
7777777
cheese
159753
arsenal
dolphin
antonio
heather
david
ginger
stephanie
peanut
blink182
sweetie
222222
beauty
987654
victoria
honey
00000
fernando
pokemon
maggie
corazon
chicken
pepper
cristina
rainbow
kisses
manuel
myspace
rebelde
angel1
ricardo
babygurl
heaven
55555
baseball
martin
greenday
november
alyssa
madison
mother
123321
123abc
mahalkita
batman
september
december
morgan
mariposa
maria
gabriela
iloveyou2
bailey
jeremy
pamela
kimberly
gemini
shannon
pictures
asshole
sophie
jessie
hellokitty
claudia
babygirl1
angelica
austin
mahalko
victor
horses
tiffany
mariana
eduardo
andres
courtney
booboo
kissme
harley
ronaldo
iloveyou1
precious
october
inuyasha
peaches
veronica
chris
888888
adriana
cutie
james
banana
prince
friend
jesus1
crystal
celtic
zxcvbnm
edward
oliver
diana
samsung
freedom
angelo
kenneth
master
scooby
carmen
456789
sebastian
rebecca
jackie
spiderman
christopher
karina
johnny
hotmail
0123456789
school
barcelona
august
orlando
samuel
cameron
slipknot
cutie1
monkey1
50cent
bonita
kevin
bitch
maganda
babyboy
casper
brenda
adidas
kitten
karen
mustang
isabel
natalie
cuteako
javier
789456123
123654
sarah
bowwow
portugal
laura
777777
marvin
denise
tigers
volleyball
jasper
rockstar
january
fuckoff
alicia
nicholas
flowers
cristian
tintin
bianca
chrisbrown
chester
101010
smokey
silver
internet
sweet
strawberry
garfield
dennis
panget
francis
cassie
benfica
love123
asdfgh
lollipop
olivia
cancer
camila
qwertyuiop
superstar
harrypotter
ihateyou
charles
monique
midnight
vincent
christine
apples
scorpio
jordan23
lorena
andreea
mercedes
katherine
charmed
abigail
rafael
icecream
mexico
brianna
nirvana
aaliyah
pookie
johncena
lovelove
fucker
abcdef
benjamin
131313
gangsta
brooke
333333
hiphop
aaaaaa
mybaby
sergio
welcome
metallica
julian
travis
myspace1
babyblue
sabrina
michael1
jeffrey
stephen
love
dakota
catherine
badboy
fernanda
westlife
blondie
sasuke
smiley
jackson
simple
melanie
steaua
dolphins
roberto
fluffy
teresa
piglet
ronald
slideshow
asdfghjkl
minnie
newyork
jason
raymond
santiago
jayson
88888888
5201314
jerome
gandako
muffin
gatita
babyko
246810
sweetheart
chivas
ladybug
kitty
popcorn
alberto
valeria
cookies
leslie
jenny
nicole1
12345678910
leonardo
jayjay
liliana
dexter
sexygirl
232323
amores
rockon
christ
babydoll
anthony1
marcus
bitch1
fatima
miamor
lover
chris1
single
eeyore
lalala
252525
scooter
natasha
skittles
brooklyn
colombia
159357
teddybear
winnie
happy
manutd
123456a
britney
katrina
christina
pasaword
cocacola
mahal
grace
linda
albert
tatiana
london
cantik
0123456
lakers
marie
teiubesc
147258369
charlotte
natalia
francisco
amorcito
smile
paola
angelito
manchester
hahaha
elephant
mommy1
shelby
147258
kelsey
genesis
amigos
snickers
xavier
turtle
marlon
linkinpark
claire
stupid
147852
marina
garcia
fuckyou1
diego
brandy
letmein
hockey
444444
sharon
bonnie
spider
iverson
andrei
justine
frankie
pimpin
disney
rabbit
54321
fashion
soccer1
red123
bestfriend
england
hermosa
456123
qazwsx
bandit
danny
allison
emily
102030
lucky1
sporting
miranda
dallas
hearts
camille
wilson
potter
pumpkin
iloveu2
number1
katie
guitar
212121
truelove
jayden
savannah
hottie1
phoenix
monster
player
ganda
people
scotland
nelson
jasmin
timothy
onelove
ilovehim
shakira
estrellita
bubble
smiles
brandon1
sparky
barney
sweets
parola
evelyn
familia
love12
nikki
motorola
florida
omarion
monkeys
loverboy
elijah
joanna
canada
ronnie
mamita
emmanuel
thunder
999999999
broken
rodrigo
maryjane
westside
california
lucky
mauricio
yankees
jamaica
justin1
amigas
preciosa
shopping
flores
mariah
matrix
isabella
tennis
trinity
jorge
sunflower
kathleen
bradley
cupcake
hector
martinez
elaine
robbie
friendster
cheche
gracie
connor
hello1
valentina
melvin
mahalkoh
sasha
felipe
liberty
nicolas
1q2w3e4r
yahoo
paris
jessica1
spider
abcdefg
mustang
jesus
fuck you
lollipop
carlitos

How do people manage all their passwords?

Risks of password reuse

CRACKED PASSWORDS

UserID	Password
jane	iloveyou89
john	godoggo!
jur	monkey1
kar	pa$$wo
katie	princ3s

jur monkey1 → Online Store

jur monkey1 → Bank

jur monkey2 → Employer

jur monkey1 → Employer

Difficult to observe password use across all of a user's accounts

Most past work relies on indirect or incomplete data sources

- Interviews and surveys
  - Self reports
- Leaked password databases
  - Allow examination of only one password per user
- Passwords created specifically for studies

Few studies of entire portfolio

Notable exception: Wash et al. (SOUPS 2016)

- Comprehensive study of real password portfolios and in-situ use
- Captured passwords of 134 participants over 6 weeks
- Found high levels of password reuse

But more research needed to examine:

- Partial reuse
- Correlation with other security behaviors
- Correlation with accurate password strength estimates
- Diverse participants

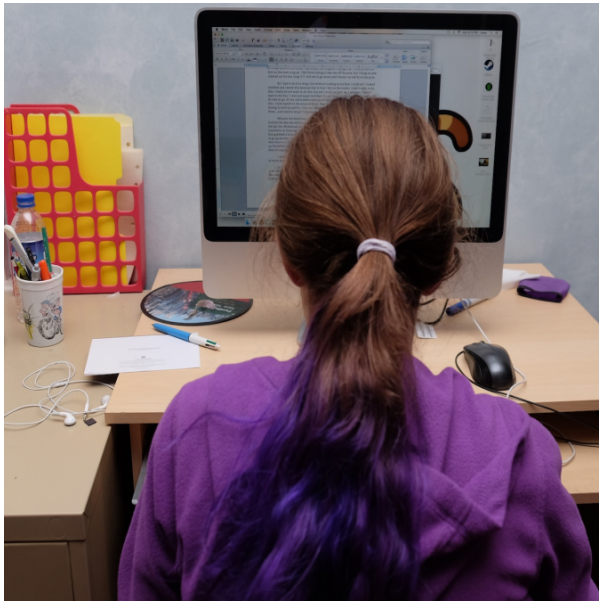
Security Behavior Observatory (SBO)

- Instrumented home Windows computers
- Collecting data since 2014
- Total participants over time: ~500
- Current active participants: ~200
- Conducting a wide range of studies of security and privacy behavior

The SBO is funded by the NSA Science of Security Lablet

SBO allows for empirical observation, scientific analysis of behavior

- Ecological validity of *in-situ* observation of home users
- Diverse participants
- Longitudinal observation
- Ability to observe most of a user's passwords across a wide range



Password data collection

Deployed new SBO browser extensions in January 2017 to collect

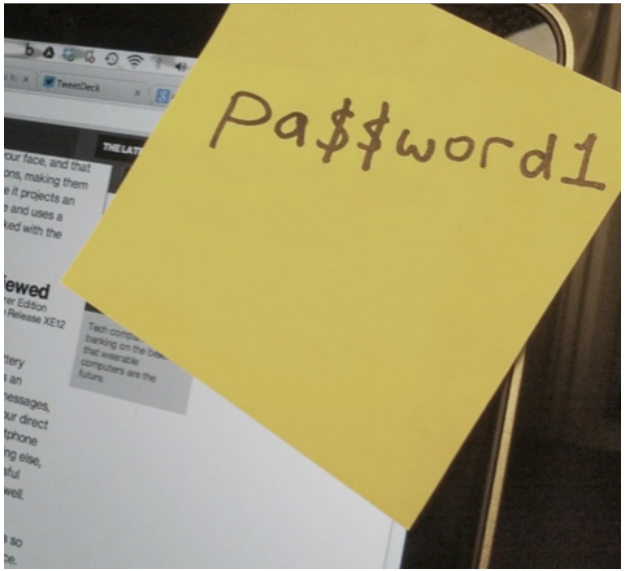
- Hashes of passwords and 4+ character substrings
- Length, strength, characters in each class (uppercase, lowercase, digits, special characters)
- Browsing metadata (including URL)

Accurate strength measurement

- Neural network guesser (Melicher et al. 2016)
- Accurate measurement of guessability by sophisticated attacker
- Runs in browser plugin (no need to transmit password)
- No noticeable delay to user

Terminology for understanding password reuse

- Types of reuse
  - **Exact** reuse
  - **Partial** reuse
- Counting a user's passwords
  - Users' total number of credentials (domain-password pairs)



Domain	Password	Reuse Type
nytimes.com	**Usabl3!!**	
yahoo.com	**s3curity**	
google.com	**security123**	
cmu.edu	**p4$$w0rd**	
facebook.com	**p4$$w0rd**	
amazon.com	**security?**	
twitter.com	**security?**	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password	Reuse Type
nytimes.com	**Usabl3!!**	
yahoo.com	**s3curity**	
google.com	**security123**	
cmu.edu	**p4$$w0rd**	
facebook.com	**p4$$w0rd**	
amazon.com	**security?**	
twitter.com	**security?**	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password	Reuse Type
nytimes.com	`Usabl3!!`	Not reused
yahoo.com	`s3curity`	
google.com	`security123`	
cmu.edu	`p4$$w0rd`	
facebook.com	`p4$$w0rd`	
amazon.com	`security?`	
twitter.com	`security?`	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password	Reuse Type
nytimes.com	Usabl3!!	Not reused
yahoo.com	s3curity	
google.com	security123	
cmu.edu	p4$$w0rd	
facebook.com	p4$$w0rd	
amazon.com	security?	
twitter.com	security?	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password	Reuse Type
nytimes.com	Usabl3!!	Not reused
yahoo.com	s3curity	Partially reused
google.com	security123	
cmu.edu	p4$$w0rd	
facebook.com	p4$$w0rd	
amazon.com	security?	
twitter.com	security?	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password	Reuse Type
nytimes.com	Usabl3!!	Not reused
yahoo.com	s3curity	Partially reused
google.com	security123	
cmu.edu	p4$$w0rd ←	
facebook.com	p4$$w0rd ←	
amazon.com	security?	
twitter.com	security?	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password	Reuse Type
nytimes.com	Usabl3!!	Not reused
yahoo.com	s3curity	Partially reused
google.com	security123	
cmu.edu	p4$$w0rd	Exactly reused
facebook.com	p4$$w0rd	
amazon.com	security?	
twitter.com	security?	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password	Reuse Type
nytimes.com	Usabl3!!	Not reused
yahoo.com	s3curity	Partially reused
google.com	security123	
cmu.edu	p4$$w0rd	Exactly reused
facebook.com	p4$$w0rd	
amazon.com	security? ←	
twitter.com	security? ←	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password	Reuse Type
nytimes.com	Usabl3!!	Not reused
yahoo.com	s3curity	Partially reused
google.com	security123	
cmu.edu	p4$$w0rd	Exactly reused
facebook.com	p4$$w0rd	
amazon.com	security?	
twitter.com	security?	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password	Reuse Type
nytimes.com	Usabl3!!	Not reused
yahoo.com	s3curity	Partially reused
google.com	security123	
cmu.edu	p4$$w0rd	Exactly reused
facebook.com	p4$$w0rd	
amazon.com	security?	Partially AND exactly reused
twitter.com	security?	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password	Reuse Type
nytimes.com	`Usabl3!!`	Not reused
yahoo.com	`s3curity`	Partially reused
google.com	`security123`	
cmu.edu	`p4$$w0rd`	Exactly reused
facebook.com	`p4$$w0rd`	
amazon.com	`security?`	Partially AND exactly reused
twitter.com	`security?`	

*These are fictitious passwords; we do not record actual plaintext passwords

Domain	Password		Reuse Type
nytimes.com	`Usabl3!!`	1	Not reused
yahoo.com	`s3curity`	2	Partially reused
google.com	`security123`	3	
cmu.edu	`p4$$w0rd`	4	Exactly reused
facebook.com	`p4$$w0rd`		
amazon.com	`security?`	5	Partially AND exactly reused
twitter.com	`security?`		

*These are fictitious passwords; we do not record actual plaintext passwords

Overview of data collected

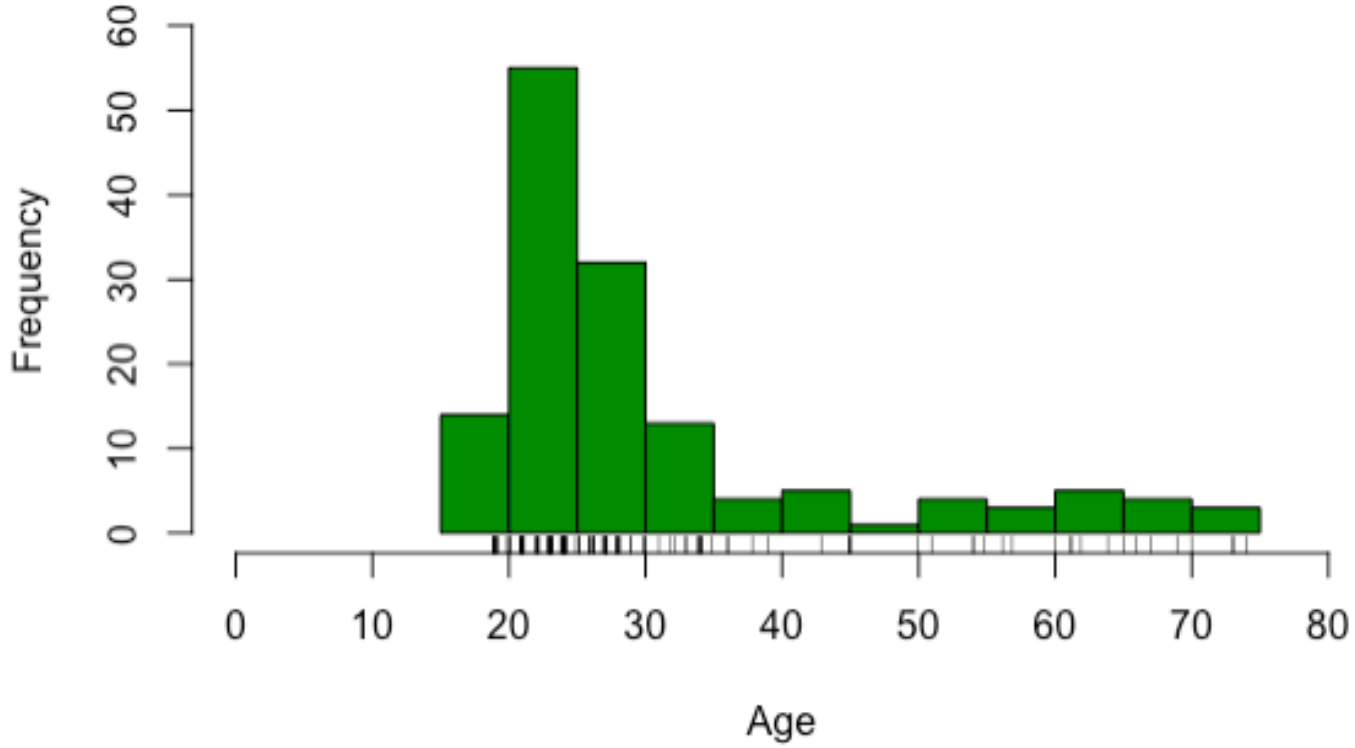
- 143 participants with >2 weeks of data
- 1,302 distinct correct passwords
  - From user's primary web browser
- 3,253 accounts
- 1,619 domains
- 131 days
- ~22,000 password entry events

Data continues to be collected

Results outline

- Participant demographics
- Password reuse
- Dominant reuse strategies
- Predicting password reuse
- Other security and usage behaviors

Median participant age: 26



Skews female and educated

- Gender

  63% female
- Education

  58% have at least a bachelor’s degree

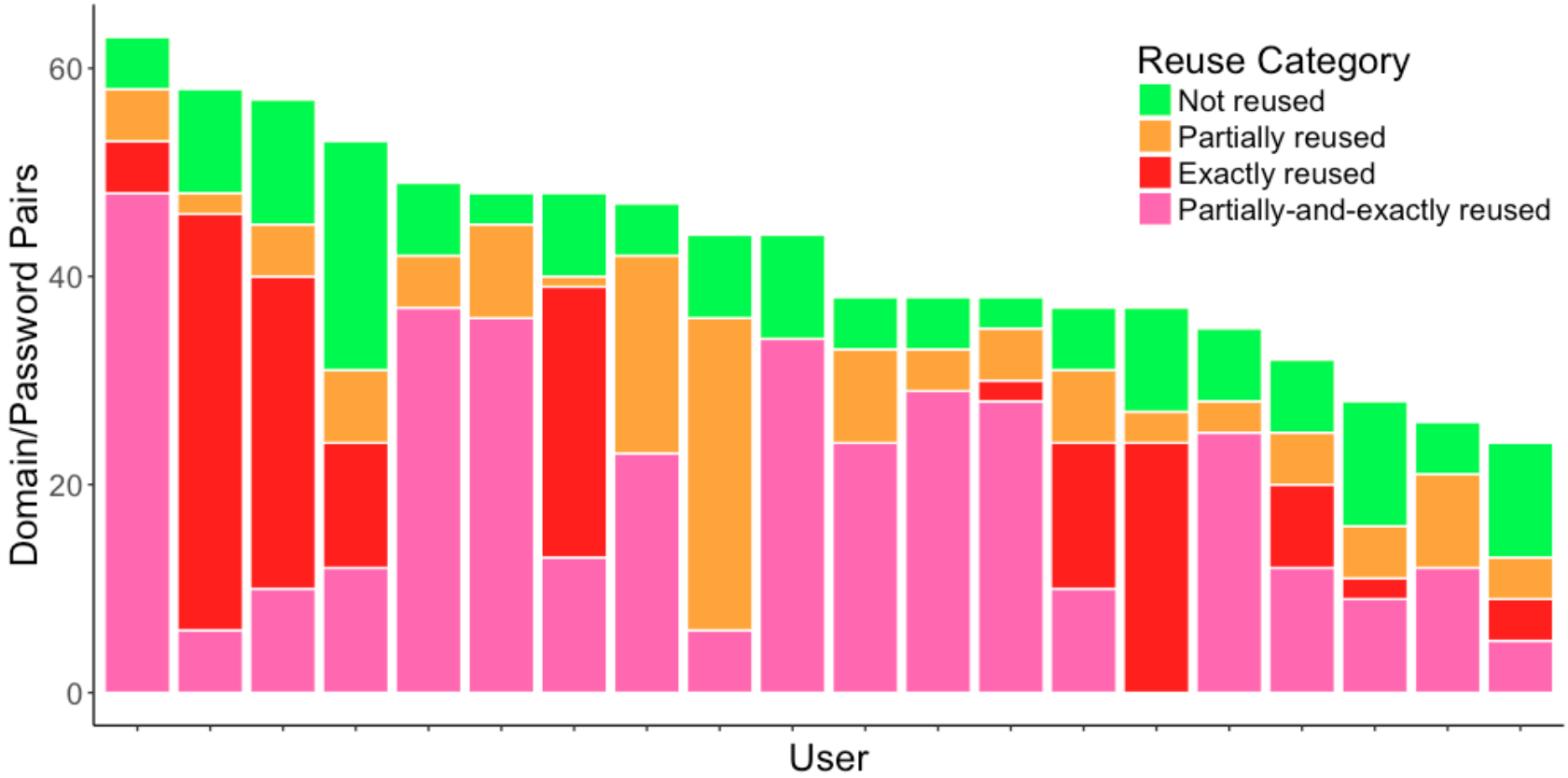
Password reuse per participant

- 22.75 different accounts (domain/password pairs)
- 9.10 distinct passwords

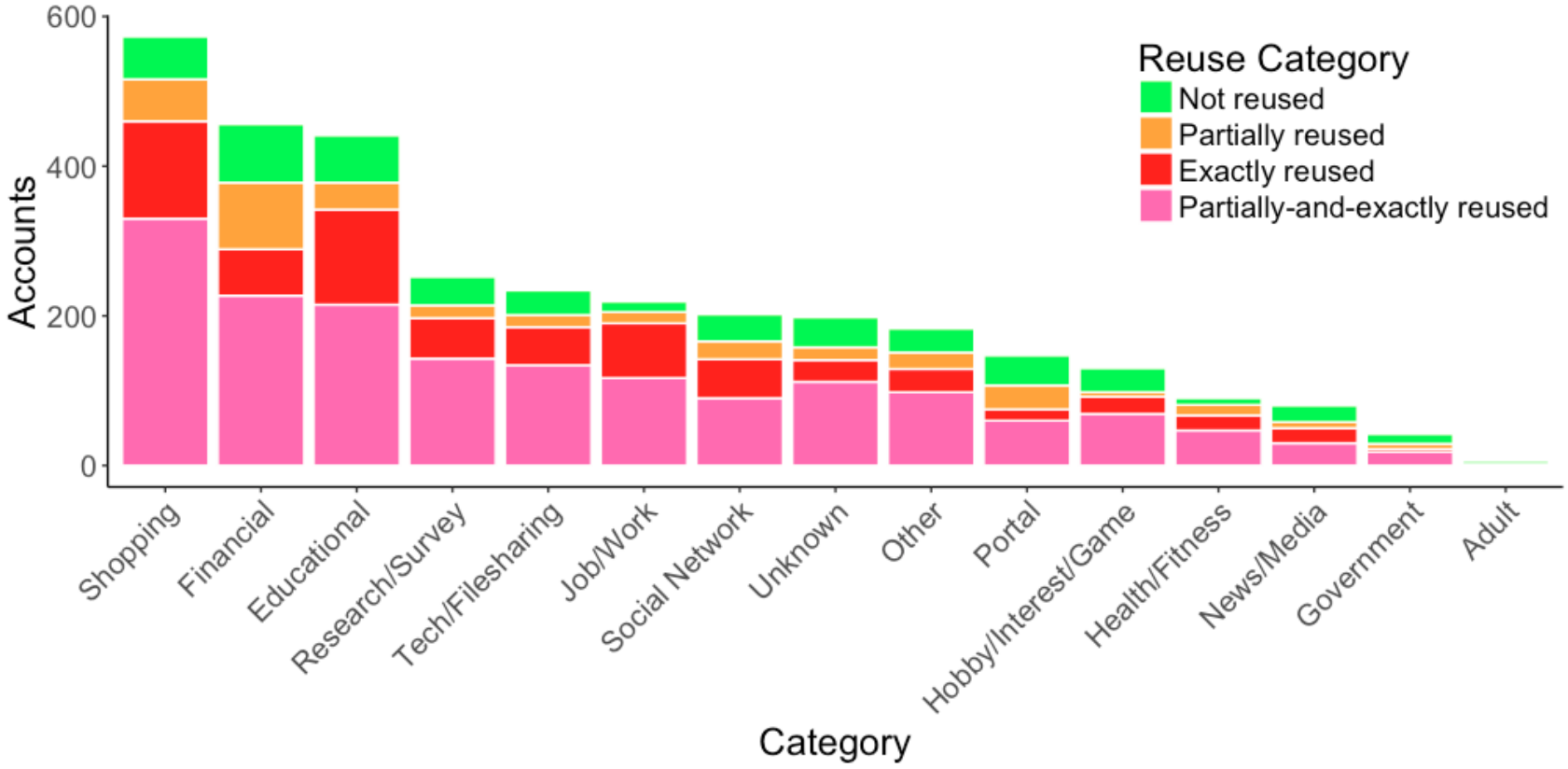
Percentages of accounts:

- With non-reused passwords: 16%
- With partially-reused passwords: 11%
- With exactly-reused passwords: 21%
- With exactly-and-partially-reused passwords: 52%

Password Reuse Types: Top 20 Users



Lots of reuse across all types of sites



Substring reuse patterns

- 83% of partially-reused passwords have reused substrings as *prefixes*
  - `password123`
  - `password456`

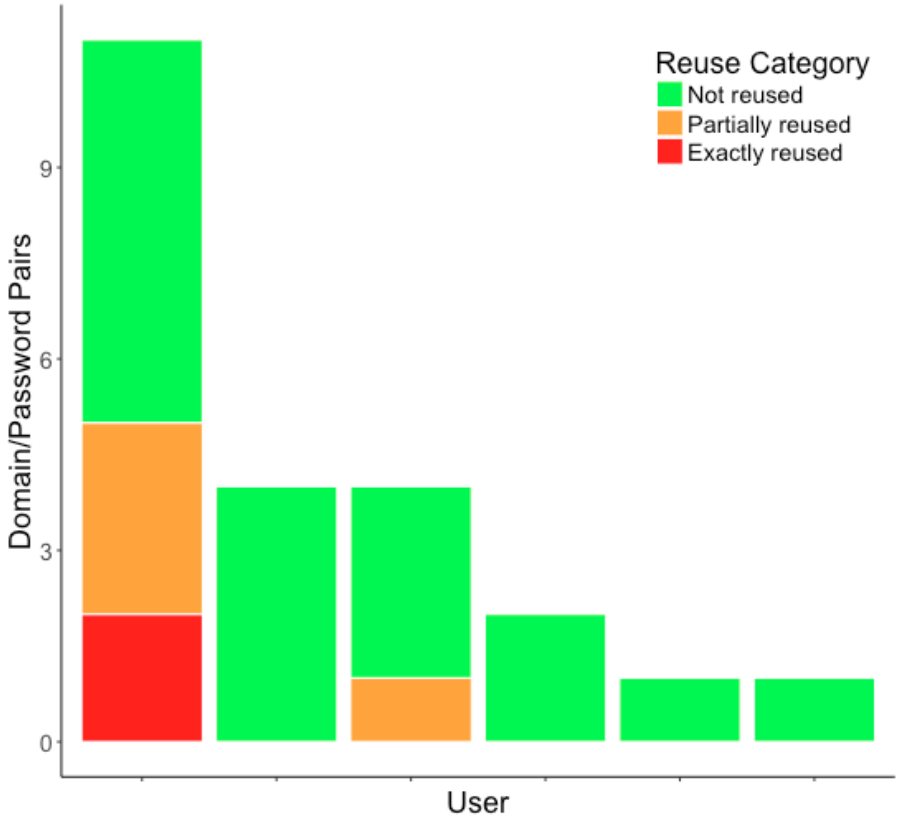
- 56% have reused substrings as *suffixes*
  - `123password`
  - `456password`

*These are fictitious passwords; we do not record actual plaintext passwords

Five dominant reuse strategies

Group 1: Unique password creators

- 6 users
- 4 with no reuse at all
- Average accounts per user: 3.8
- Average distinct passwords per user: 3.7



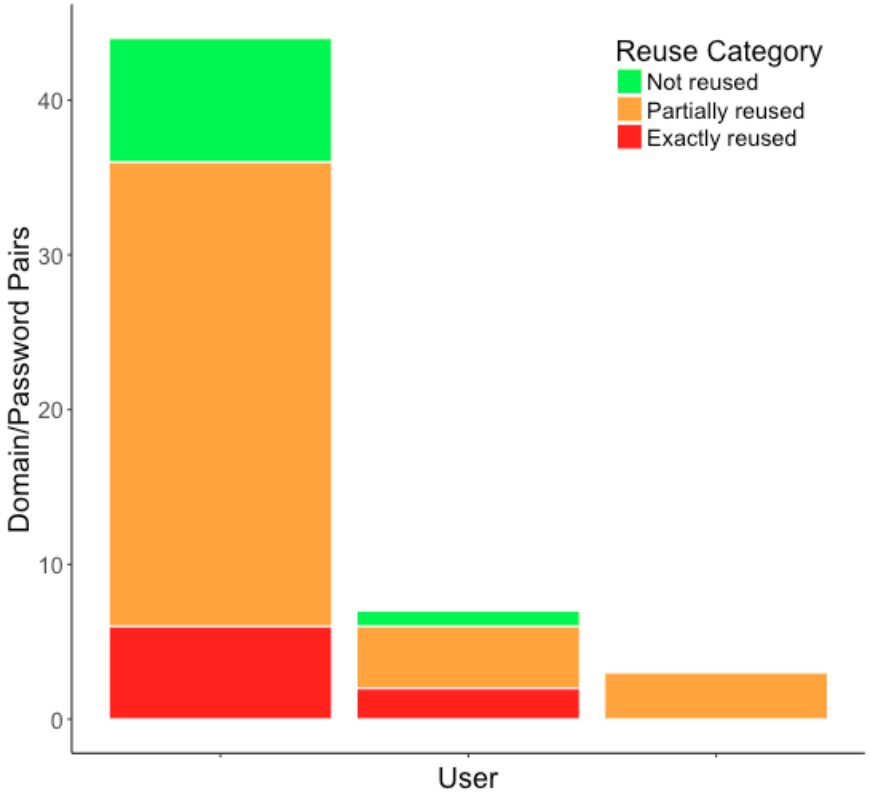
Example “Unique password creator” portfolio

- example1.com: password2345
- yahoo.com: s3curity
- facebook.com: 23ji2gdslk32
- pnc.com: H;$u.eq5uAFg

*These are fictitious passwords; we do not record actual plaintext passwords

Group 2: Partial password reusers

- 3 users
- Average accounts per user: 18
- Average distinct passwords per user: 16.7



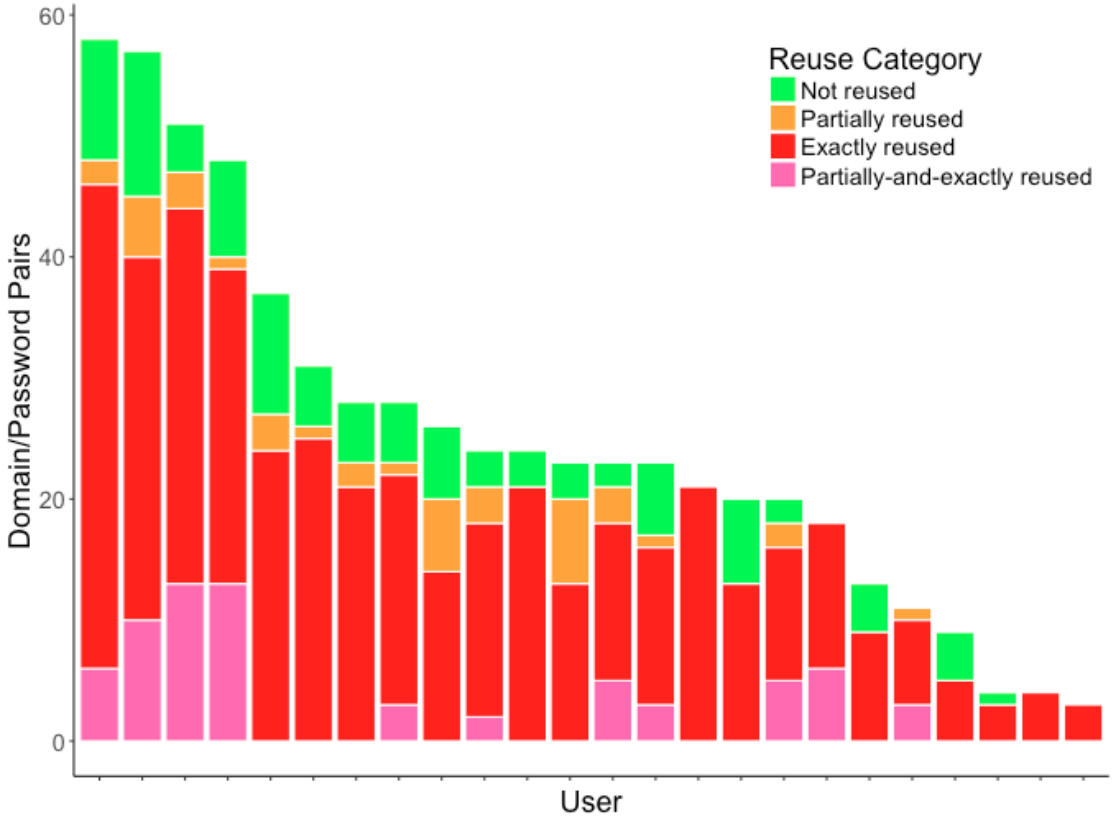
Example “Partial reuser” portfolio

- example1.com: password2345
- yahoo.com: s3curity
- google.com: applesauce
- cmu.edu: usableCMU123!
- example.com: usable54321
- facebook.com: usabl3!!!!!!!!!!
- pnc.com: usable123
- wellsfargo.com: usabl3
- twitter.com: usable1
- verizon.com: usable1!

*These are fictitious passwords; we do not record actual plaintext passwords

Group 3: Exact password reusers

- 24 users
- Average accounts per user: 25.2
- Average distinct passwords per user: 8.7



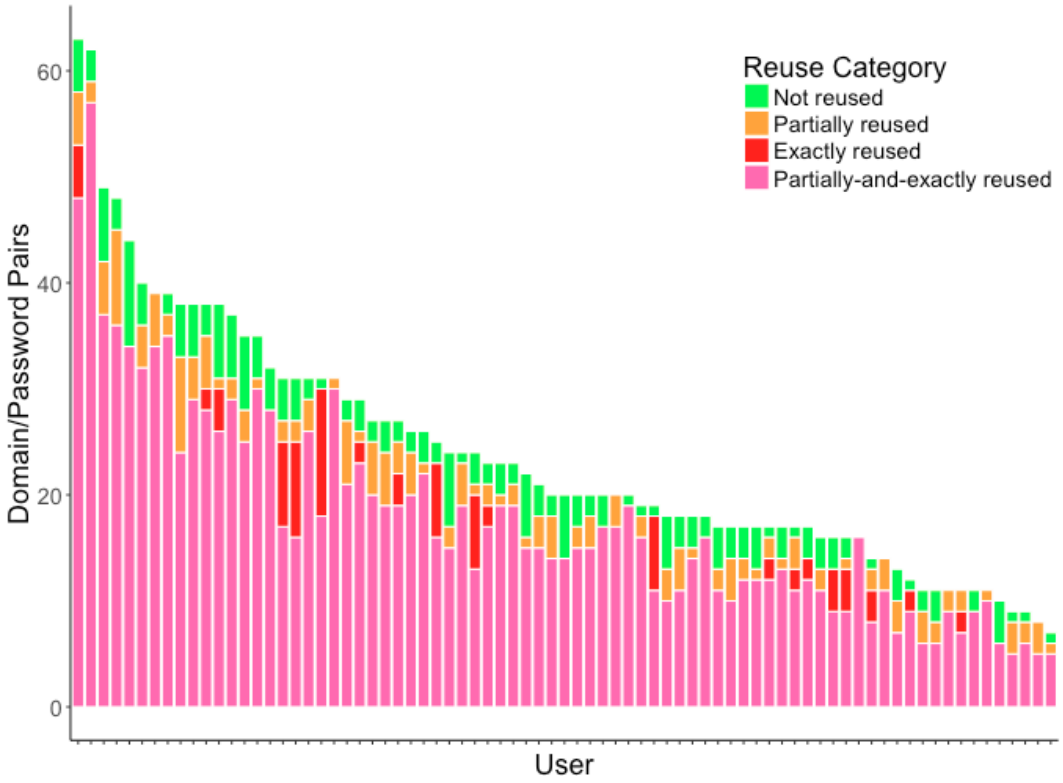
Example “Exact reuser” portfolio

- example1.com: Appl3sauce!
- yahoo.com: s3curity
- google.com: Appl3sauce!
- cmu.edu: p4$$w0rd
- example.com: Appl3sauce!
- facebook.com: Appl3sauce!
- pnc.com: p4$$w0rd
- wellsfargo.com: Appl3sauce!
- twitter.com: p4$$w0rd
- verizon.com: Appl3sauce!

*These are fictitious passwords; we do not record actual plaintext passwords

Group 4: Exact-and-partial password reusers

- 77 users
- Average accounts per user: 23.9
- Average distinct passwords per user: 8.3



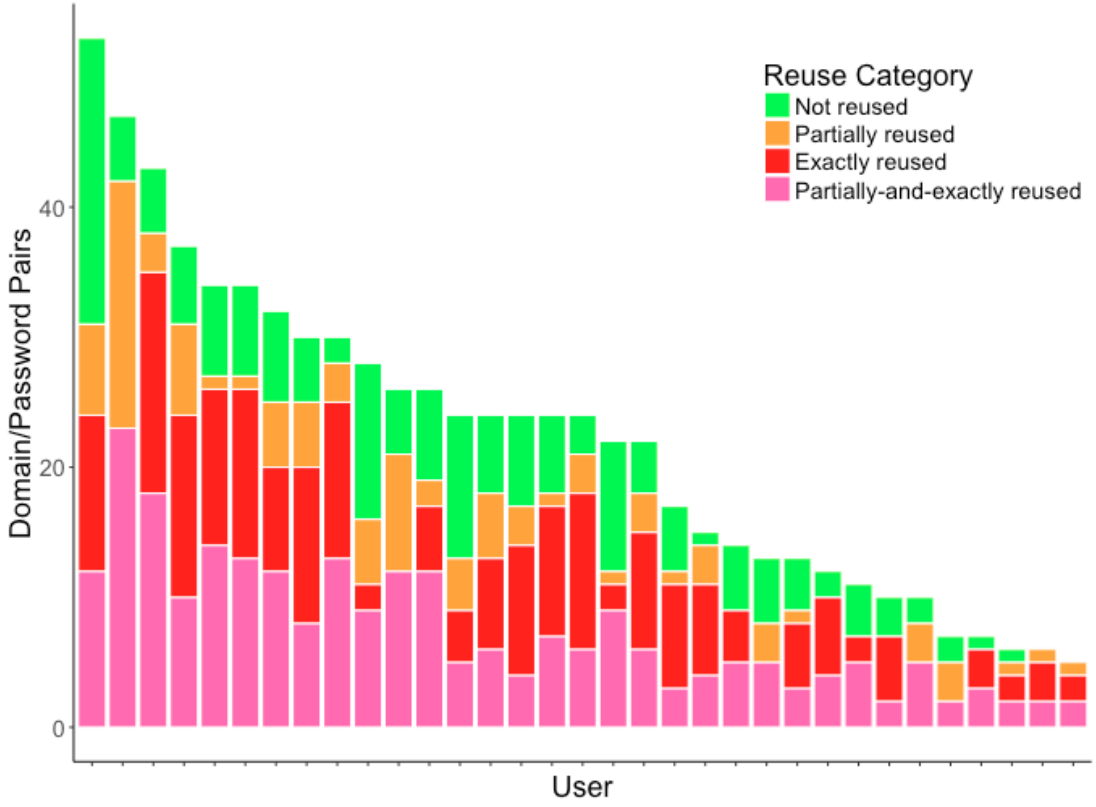
Example “Exact-and-partial” portfolio

- example1.com: Appl3sauce!
- yahoo.com: s3curity
- google.com: appl3sauce
- cmu.edu: appl3sauce123
- example.com: Appl3sauce!
- example11.com: banana
- example12.com: pennsylv4nia
- example13.com: pennsylvania
- facebook.com: Appl3sauce!
- pnc.com: appl3sauce123
- wellsfargo.com: appl3sauce
- twitter.com: p3nnsylv4n!a!!!
- verizon.com: B!CYCL3
- example16.com: pennsylv4nia
- example17.com: Appl3sauce!
- example18.com: appl3sauce123

*These are fictitious passwords; we do not record actual plaintext passwords

Group 5: Mixed-strategy users

- 33 users
- Average accounts per user: 22.1
- Average distinct passwords per user: 11.7



Example “Mixed-strategy user” portfolio

- example1.com: password2345
- yahoo.com: Pennsylvania!
- google.com: appl3sauce123
- cmu.edu: usableCMU123!
- example.com: usable54321
- example11.com: banana
- example12.com: p3nnsylv4n!a
- example13.com: pennsylvania
- facebook.com: Appl3sauce!
- pnc.com: asparagus
- wellsfargo.com: B!CYCL3
- twitter.com: p4$$w0rd
- verizon.com: p4$$w0rd
- example16.com: p4$$w0rd
- example17.com: Appl3sauce!
- example18.com: appl3sauce123

*These are fictitious passwords; we do not record actual plaintext passwords

Regression models

Based on password metrics, website categories, and user/demographic variables

- Logistic multi-level model predicting if a password is reused or not
- Linear multi-level model predicting how many domains reused passwords are reused on

Predictors of password reuse

- Contains digits
- Contains special characters
- Weaker
- Used on job/work site
- Used on shopping site
- Not used on government site

Predictors of reuse on more domains

- Contains digits
- Entered less frequently
- Weaker
- Not used on educational, financial, government, or portal* site

**Portal sites: multipurpose sites with functionality including but not limited to email, such as Google.com or Yahoo.com*

Regression models based on other security/usage behaviors

Presence of password managers, presence of security- or privacy-related browser extensions, dangerous downloads detected, malware detected, average page visits per day

- Logistic multi-level model predicting if a password is reused
- Linear multi-level model predicting how many domains a reused password will be reused on
- Linear multi-level model predicting password strength

Security behaviors as predictors of reuse

- Malware detected on computer → passwords more likely to be reused

- Higher numbers of daily page visits → reused passwords reused on more domains

Security behaviors as predictors of password *strength*

- Security/privacy browser extensions → stronger passwords
- Dangerous downloads detected → stronger passwords
- Password manager installed → weaker passwords

Ongoing work

- More data (>200 users), more time/observations (>120 days of data)
  - With more data we can analyze change over time
  - Can also examine effects of major password breaches
  - Have already increased to >160 users, >120 days of data on average
- Surveys and/or interviews to understand intentions surrounding password (re)use
- Better understanding of *use* of password managers
- Statistical analyses to cluster users

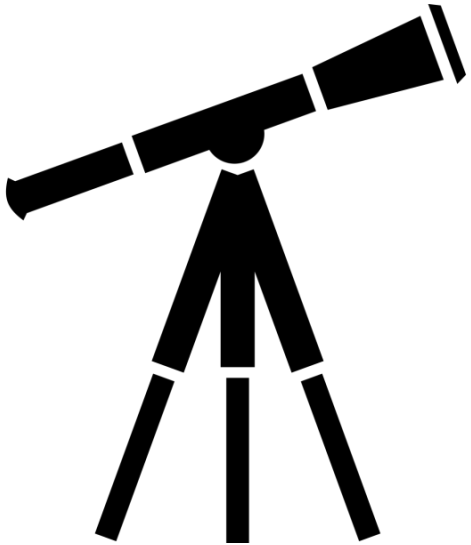
Reuse is rampant

- SBO provides unique opportunity to study how users manage large numbers of passwords
- Users seem to cope with password demands through reuse strategies
- Many people use a mixture of reuse strategies
- Password managers may not be helping very much

Carnegie Mellon University

ICSI
INTERNATIONAL COMPUTER SCIENCE INSTITUTE

SBO project:
http://cups.cs.cmu.edu/sbo.html



Password strength distribution clusters

Cluster	Part.	Mean	Min	25%	Median	75%	Max	SD
1	18	13.10	11.50	12.30	12.90	13.80	15.90	1.050
2	11	10.90	10.40	10.60	10.90	11.10	11.50	0.376
3	29	12.50	10.90	12.10	12.50	12.90	14.30	0.788
4	17	14.50	12.80	14.00	14.40	15.40	16.10	0.966
5	26	10.40	8.52	9.14	10.60	11.20	13.60	1.280
6	3	9.52	9.24	9.31	9.37	9.66	9.96	0.380
7	2	8.63	8.23	8.43	8.63	8.83	9.03	0.564
8	2	18.20	17.80	18.00	18.20	18.40	18.60	0.599
9	5	8.44	6.93	8.34	8.51	9.11	9.29	0.931
10	1	6.06	6.06	6.06	6.06	6.06	6.06	

Table 3: Strength of cluster participants' average passwords

*will re-draw this nicely if we decide we want to include a version of this table

Cluster 1

- Range of strength values 0-10^{30}
- Median: 12.90
- Multimodal
- Participants who mostly use passwords of fairly high or fairly low strength
- We hypothesize this indicates a strategy of using strong passwords for high-value accounts and weak, “throw-away” passwords for low-value accounts

Clusters 2 and 5

- Relatively weak passwords overall
- Cluster 2: almost never pick strong passwords
- Cluster 5: broader range of password strengths, with weaker passwords on average, but with more peaks, including some higher-strength passwords

Cluster 3

- Similar summary statistics to cluster 1 (median strength 12.50), but differently-shaped distribution
- Single mode: one peak
- Distribution is more normal
- Passwords fairly weak overall

Cluster 4

- Single mode
- Median strength much stronger than the other distributions
- Median strength: 14.40
- These users tend to choose stronger passwords overall

Examining password strength distributions per user

- Kolmogorov-Smirnov test to calculate distances between distributions of password strength for different users
- Hierarchical clustering to group users into clusters with similar distributions
- 10 clusters
- Most participants fell within clusters 1-5

	Cluster traits	Mean strength	Median strength
1	Multimodal: some weak “throw-away” passwords and some strong passwords	$10^{13.10}$	$10^{12.90}$
2	Weak passwords overall; almost no strong passwords	$10^{10.90}$	$10^{10.90}$
3	Single mode; normally distributed; fairly weak passwords	$10^{12.50}$	$10^{12.50}$
4	Single mode; stronger passwords than the other clusters	$10^{14.50}$	$10^{14.40}$
5	Weak passwords on average, but more range than Cluster 2	$10^{10.40}$	$10^{10.60}$

K-means clustering of passwords by reuse characteristics

Factors:

- Fraction of accounts that password is exactly reused on
- Fraction of accounts that password is partially reused on
- Average entries per day
- Within-category reuse
- Other-category reuse
- Span of category reuse
- Days site visited

K-means clusters

- Clusters 3, 5, 10: Weakest passwords, more exact reuse
- Cluster 6: Average strength, zero reuse of any kind
- Cluster 2: Moderate strength, less reuse, very little partial reuse, used on less-frequently-visited websites
- Cluster 1: Strongest passwords, sites visited often, passwords entered often
- Cluster 7: Strongest passwords, sites visited often, passwords entered less often

Exact reuse

`security123` `security123`

Partial reuse

security123

security1!?!