Improving the Systematic Generation of Secure and Memorable Passphrases by MASCARA

Kousshik Raj (17CS30022)
Supervisor:- Dr. Mainack Mondal

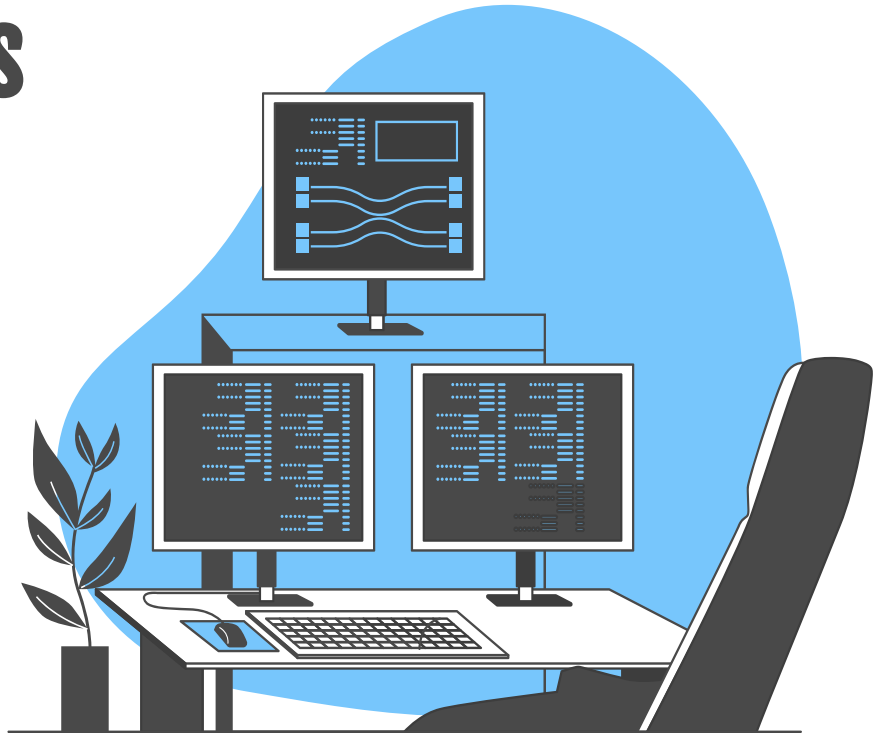


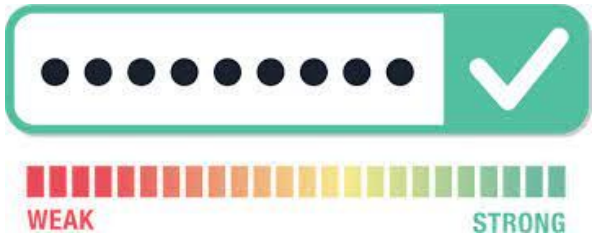
Table of Contents

01

INTRODUCTION

INTRODUCTION

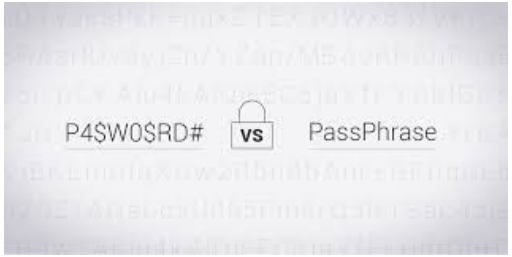
...



Passwords Everywhere

Passwords are the most common authentication mechanism, but they are not memorable and securable often

...



Passphrase as Alternative

We can use passphrase as alternative, either as passwords or context to generate passwords. Easy to remember, hard to guess

...

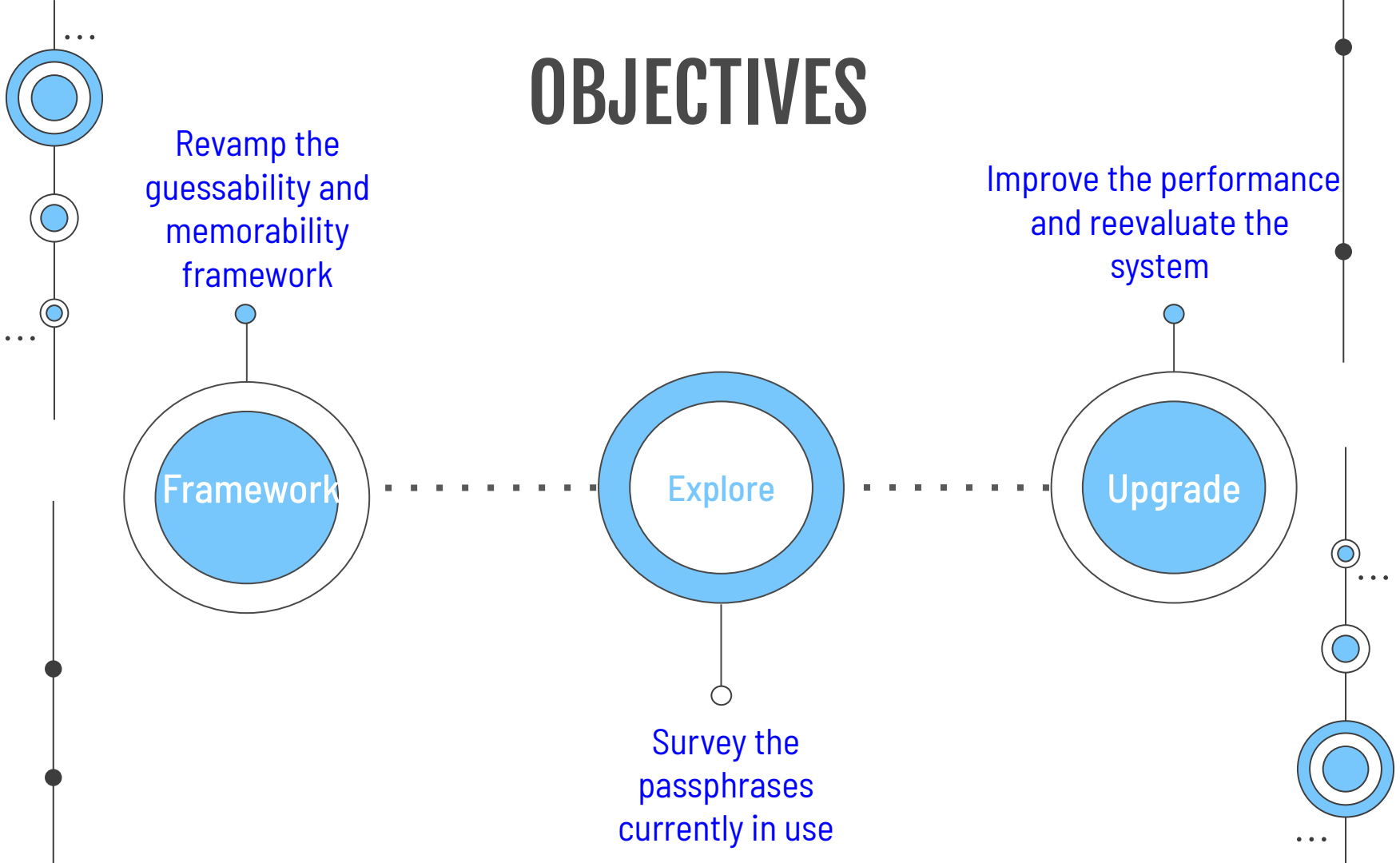
CURRENT APPROACHES

User: Easy to remember and easy to guess

Machine: Hard to guess and hard to remember

Mascara: Achieves a tradeoff between remembering and guessing

OBJECTIVES


Revamp the guessability and memorability framework
Framework
Explore
Survey the passphrases currently in use
Improve the performance and reevaluate the system
Upgrade

02

GUESSABILITY & MEMORABILITY

GUESSABILITY

THREAT MODEL

Adversary has access to core of the system like training data, algorithm, parameters, etc.

GUESSABILITY METRIC

Guessrank: the number of guesses adversary makes before identifying correctly. High guessrank, high security.

MIN-AUTO APPROACH

Different cracking algorithm, different guessrank. Min-auto approach is minimum of multiple algorithms, close to real world scenario guessrank.

MEMORABILITY

Unigram Probability

Sum of log probabilities of occurrences of words in corpus.

$$CER(s) = c_1 L_1(s) + c_2 L_2(s) + c_3 \sigma(s)$$

Fitted to data using linear regression

Bigram Probability

Sum of log bigram probabilities of frequency of word pairs in corpus.

Standard Deviation

Standard deviation of the length of the words

03

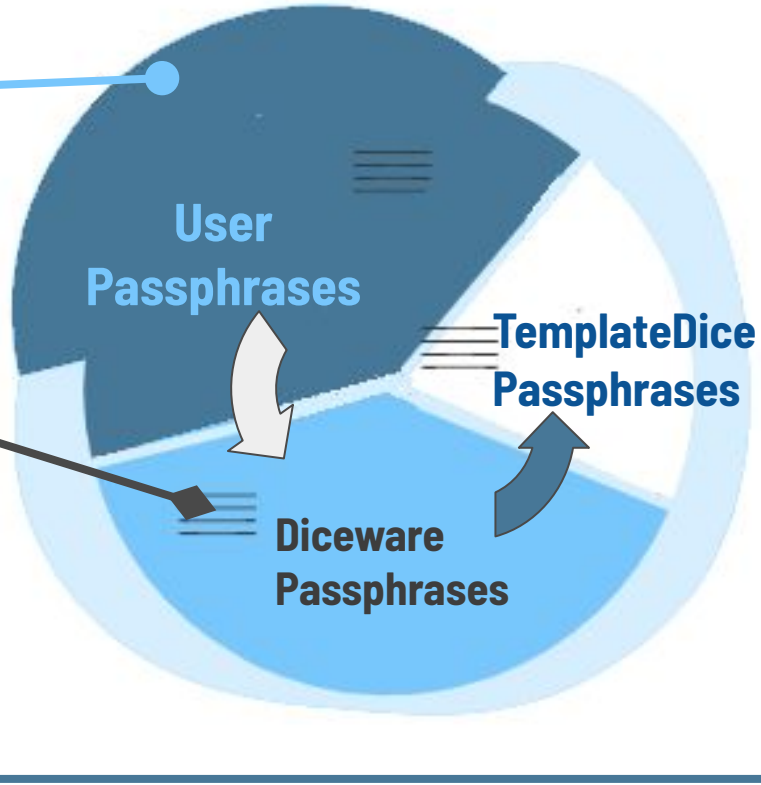
PASSPHRASES IN WILD

CLASSES OF PASSPHRASES

Chosen by user. Extracted from password leak.

System passphrase, word chosen randomly from wordlist

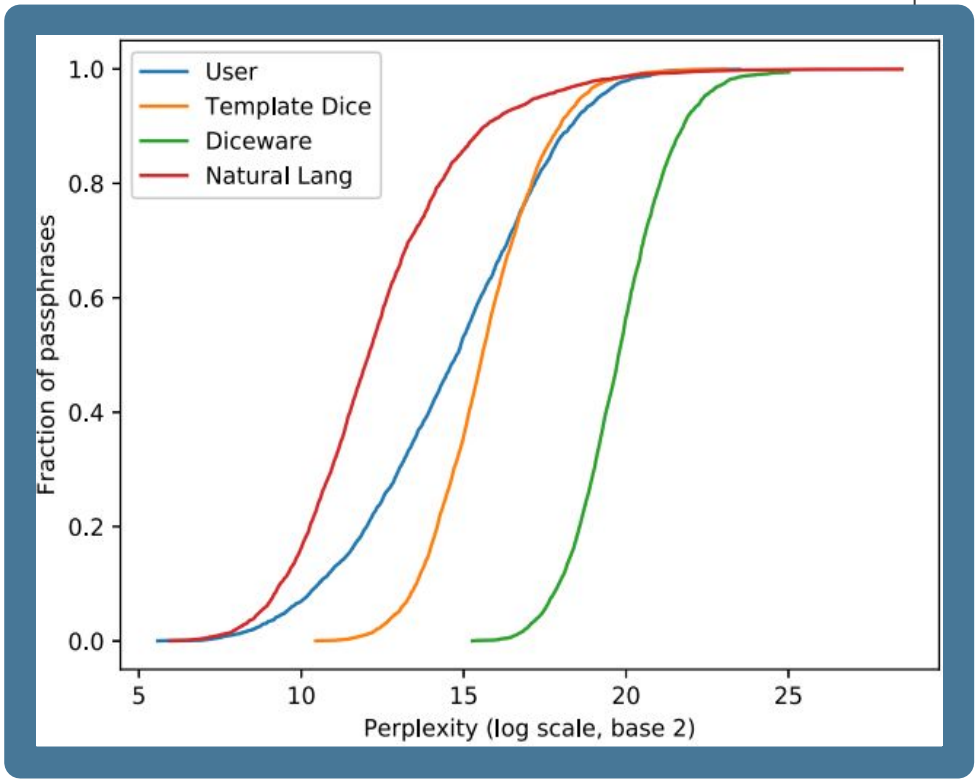
Improved Diceware,use linguistic templates and partitioned wordlist.



LINGUISTIC PROPERTIES

Perplexity of passphrases, w.r.t GPT2 model, is calculated. Defines how accurately the model predicts phrases.

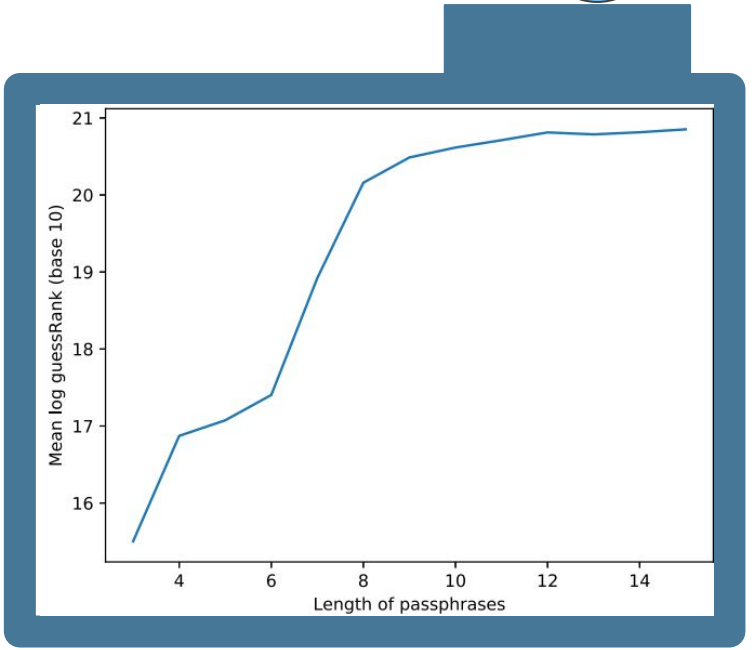
All phrases are similar length. Natural language is reference. User closest, next TemplateDice. Diceware worst.



TEMPLATEDICE: ISSUES

Not scalable, all information hardcoded.

Guessrank saturates after length 8.



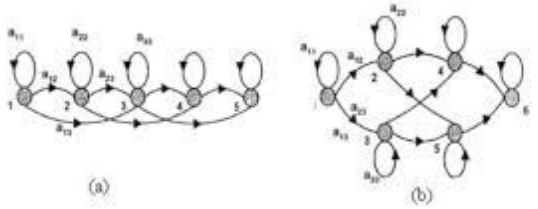
Arbitrary length not possible, restricted by templates.

04

MASCARA

TRAINING DATA

...



Recent Wikipedia Articles

Recent top 5% viewed articles, extracted, cleaned. Have 3.3m sentences, 450k unique words (Wiki-5)

...

Bigram Markov Model

Tokenize Wiki-5, add start, end tokens, and create bigram Markov model. Store unigram and bigram frequency.

...

ALGORITHM

1. For partially generated passphrase $s_i = p_1 \ldots p_i$, calculate score as $C(s_i) = c_1 L_1(p_i) + c_2 L_2(p_{i-1} . p_i) + c_3 \sigma(p_1 \ldots p_i)$

2. Choose p_{i+1} from Markov graph, to generate s_{i+1}, such that $C(s_{i+1}) < \alpha_1$ & $L_2(p_i . p_{i+1}) < \alpha_2$

3. If no word choice found, restart process. Stop algorithm when required length is reached.

4. Pass the final passphrase to *PhraseMachine*. Reject and restart, if passphrase not correct. Else return.

05

PROPOSED MODIFICATIONS

PHRASE MACHINE

PhraseMachine heuristic to check for linguistically sound passphrases.

Rejects over 96% generated passphrases, even 'almost' correct ones.

Slows down generation significantly, not much to show quantitatively

Remove PhraseMachine from the pipeline of generation.

Add word heuristics (checking stopwords), stopping full rejections.

ALGORITHM COMPLEXITY

FILTERING

For every word, full vocabulary taken and words filtered in linear time

Filtering words in preprocessing, dependent on bigrams, in linear time

SAMPLING

After filtering, weight sampling by frequency in linear time is done to choose next word

In preprocessing, for each word, calculate CDF of freq. in linear time

Random no. from [0, 1). Use it with CDF to choose word in logarithmic time

PARAMETER TUNING

Range of values determined through choice and equation. Each value is set and manually evaluated.

Not scalable (new parameters, precision) and reliable (bias, inaccuracies) due to human intervention.

New metric μ incorporating CER & Guessrank for system evaluation introduced

Similar range of values, perform grid search and evaluate using μ value.

Manually evaluate only top k systems. Makes it scalable and more accurate.

06

SYSTEM EVALUATION

PHRASE MACHINE: RUNTIME

...

Length	Runtime$_a$ (in s)	Rejection$_a$	Runtime$_b$ (in s)	Rejection$_b$
4	0.08	4.34	8.7	68.4
5	0.09	5.85	18.2	91.8
6	0.10	7.24	28.1	112.4
7	0.13	11.62	43.1	201.7

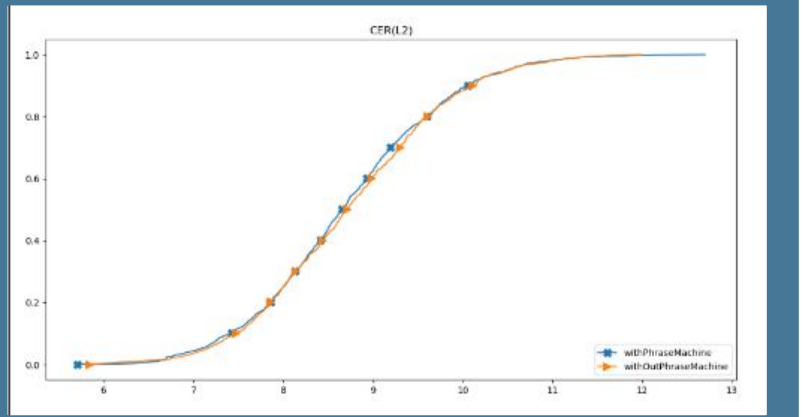
Runtime and rejection count averaged over 100 passphrases before and after removing *PhraseMachine*.

Significant increase in performance.

...

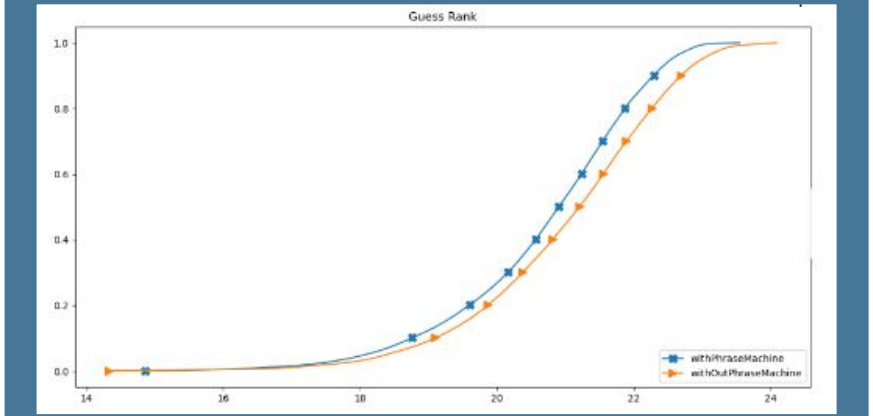
...

PHRASE MACHINE: CER & GUESSRANK



CER

Negligible change



GUESSRANK

Slight Increase

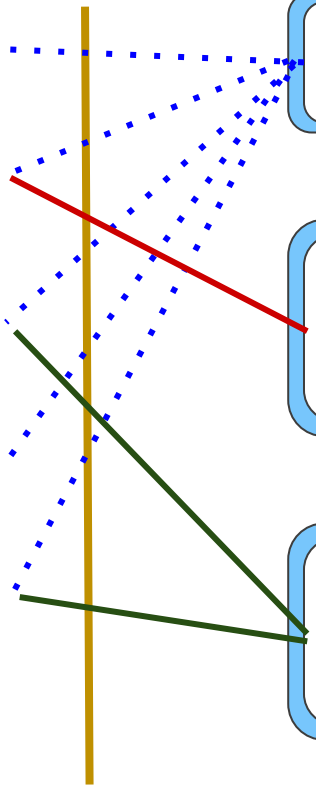
Improvement in System

OVERALL RUNTIME

Diceware System	DICEWARE	0.33s
TemplateDice System	TEMPLATE-DICE	7.85s
Mascara version: Constraints verified at end	$\text{MASCARA}_{\text{END}}$	930s
Mascara version: Original	$\text{MASCARA}_{\text{LIN}}$	86.4s
Mascara version: Logarithmic algorithm complexity	$\text{MASCARA}_{\text{LOG}}$	0.08s

BENCHMARKS & ALGORITHMS

1. **DICEWARE**
2. **TEMPLATEDICE**
3. **MARKOV**
4. **USER**
5. **MASCARA**



Markov (2,3)-word & (4,5,6)-char

Template-Based: Exhaustive guessing pivoted around templates

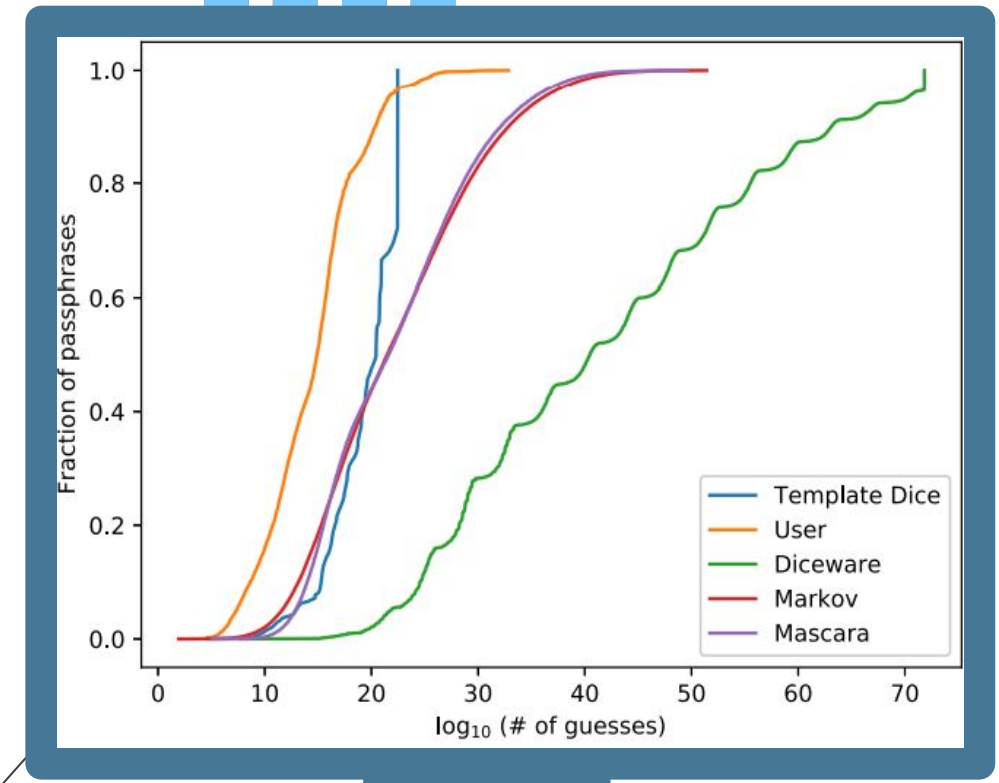
Wiki-5 Bigram: Bigram Markov model trained on Wiki-5

SAMPLE PASSPHRASES

Type	Samples
Diceware	valedictory silliest illeism niglichen dreamscape manchuria dervish verbally whinging ferries portmore permanency downcast epilogue guarding richemont feeney stoppers scalable deuteronomic kudo
UserPP	little bitty pretty one dont forget the password just another happy ending ride with the wind hot and sexy chicken wing
Markov	murrays situation spores have protractable and the problem standard kit during ultraviolet signals beamed there improvements achieved worldwide with coitus are shipwrecked man
MMAP	a flea will republish your gladiator how does its 1 shire milk a jack the frothy thing faxed the aphorism the rivalry was misspelling prior to our eyeballs can Carole wrestle the gaiter
Mascara	woolwich the fleet including queen many local fiscal rights lee sung before god jackal with prayer requests drakes book nearly too quickly

...

GUESSRANK



Highest: Diceware, 50% pp over 10^{40}

Lowest: User, 50% pp within 10^{14}

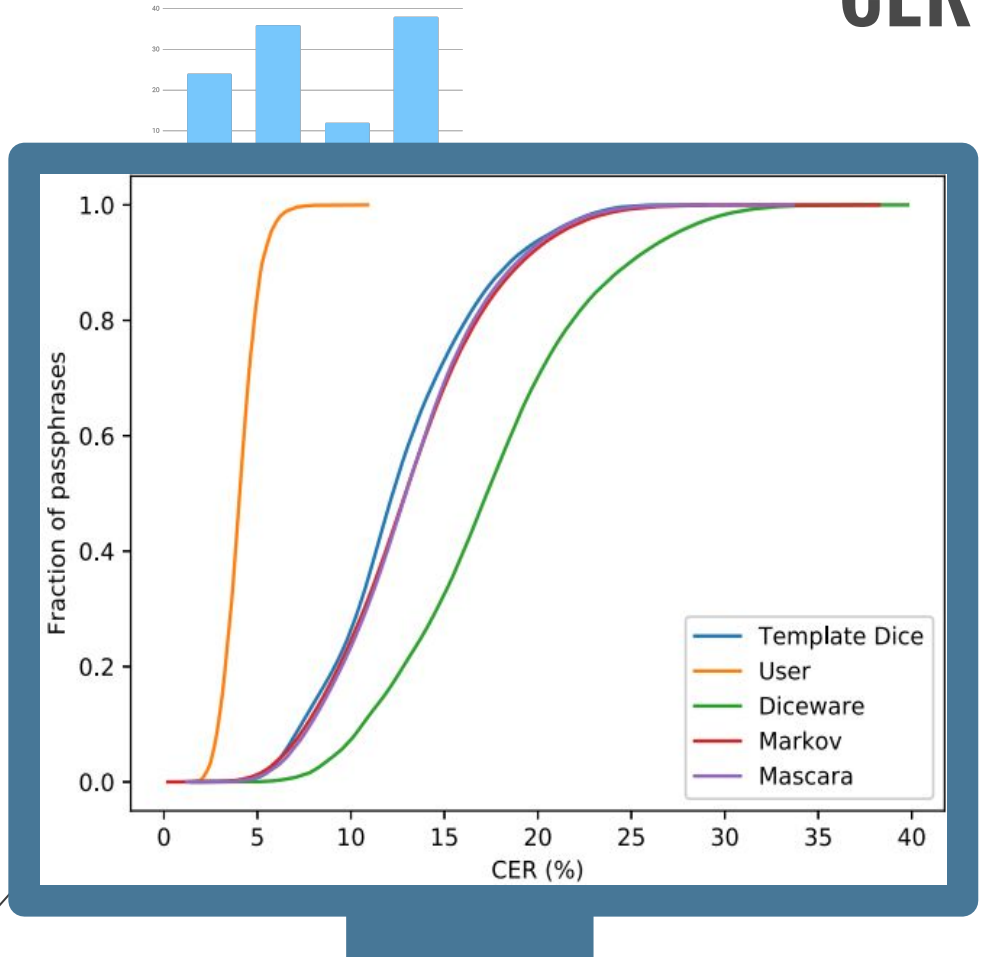
Mascara *vs* TemplateDice

Below 40% pp: TemplateDice better

Above 40% pp: Mascara better,
TemplateDice saturates

Top 20% pp: Mascara(10^{30}) > TemplateDice(10^{22})

CER



50% threshold

Highest CER: Diceware, mistake every 6 char

Lowest CER: User, mistake every 20 char

Rest: Very similar, mistake every 9 char

USER STUDY

Model	Recall	Mean CER	Median CER
Mascara	26.23%	34.78%	35.85%
TemplateDice	17.46%	35.44%	36.58%
Markov	21.95%	37.84%	41.27%
Diceware	24.00%	38.49%	42.57%

- Two part authentication survey. 3 similar passphrase choice.
- 12 passphrase classes: 3 lengths (<7, 7-11, >12) x 4 systems.
- Highest Recall after 2 days: MASCARA.
- Diceware comparable due to lower return rate. High CER among returnees.

THANK YOU!