CERIAS

The Center for Education and Research in Information Assurance and Security

A Study of Probabilistic Password Models

Jerry Ma, Weining Yang, Min Luo, Ninghui Li

To appear in 2014 IEEE Symposium on Security and Privacy (Oakland)

Password

- Most widely used method for user authentication
- Easy to understand and use, easy to implement

→ IMPORTANT

Users tend to choose WEAK passwords
- easy to guess

Probabilistic Password Models

- A probabilistic password model assigns a probability value to each string
- Goal: To approximate as accurately as possible an unknown password distribution

Template-based Model

- Divide a password into several segments, often by grouping consecutive characters of the same category (e.g., lower-case letters, digits)
- Probabilistic Context Free Grammar (PCFG)

Whole-string Model

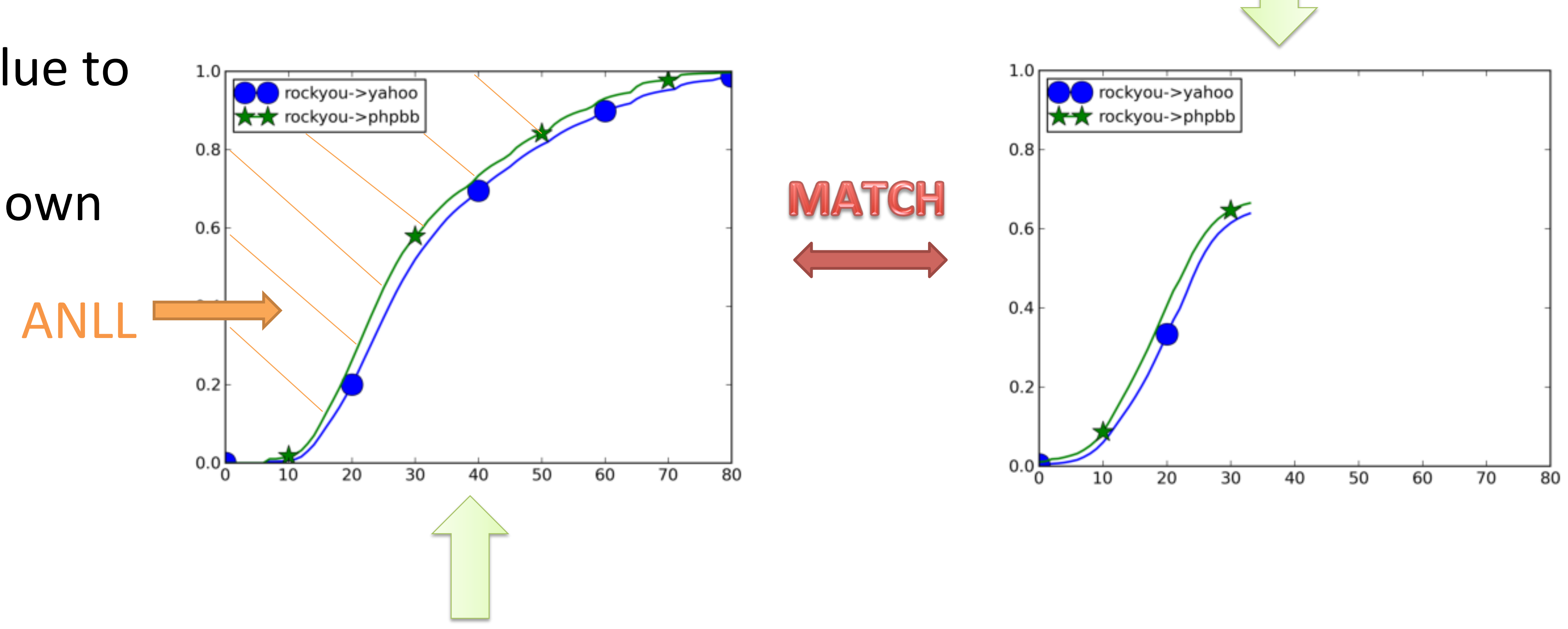
- Does not divide a password into segments
- Markov chain models (aka n-gram models)
- Fixed order with smoothing (ws-mc), Variable-order using backoff (ws-mc-b)

Research Topics on Passwords

- What makes users choose more (or less) secure passwords?
- How to find the best password models?

Current Approach: Guess Number Graph

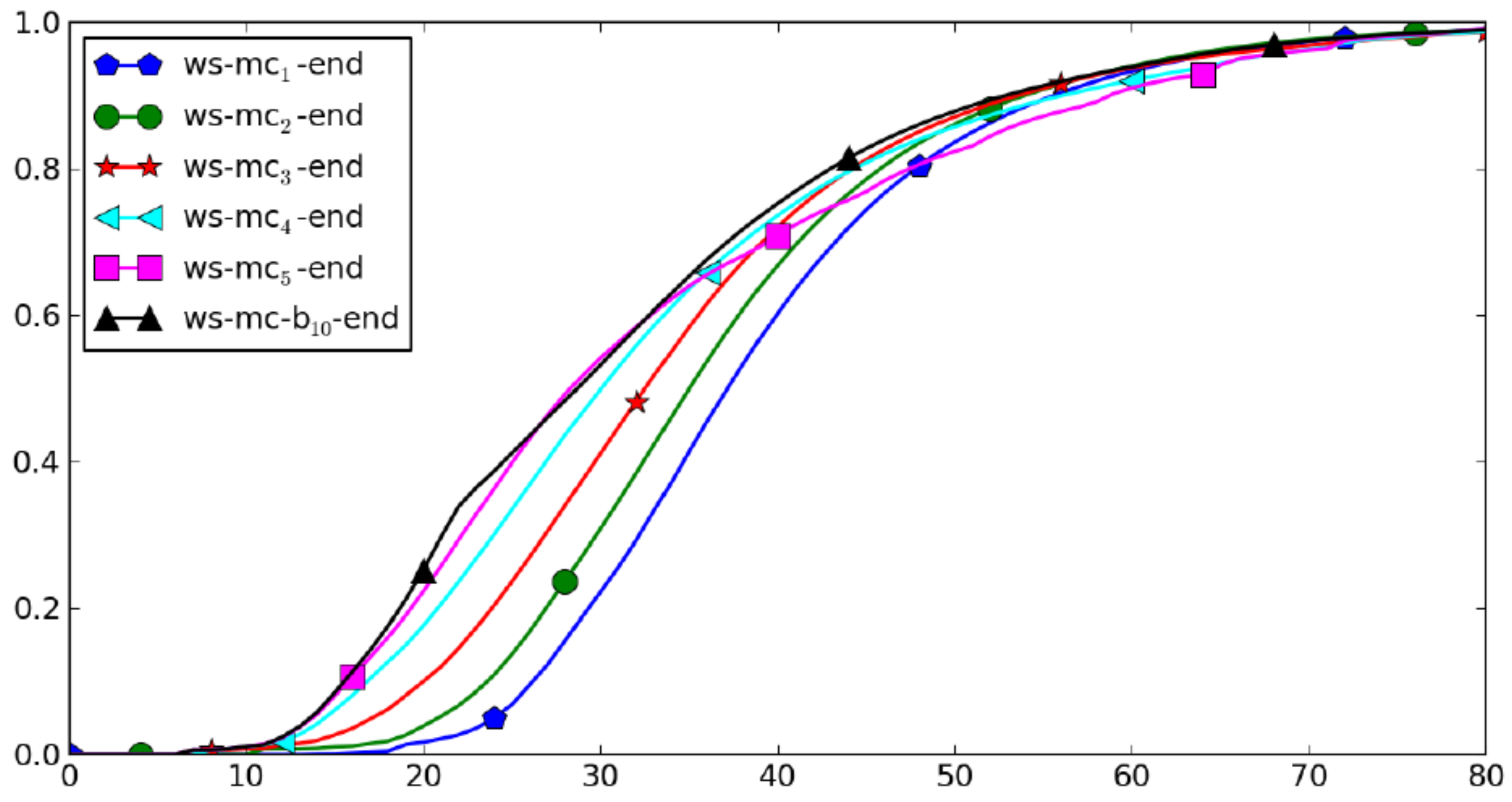
- The number of guesses in log scale vs. the percentage of passwords cracked in the dataset.
- Computational expensive: need to generate a very large number of password guesses

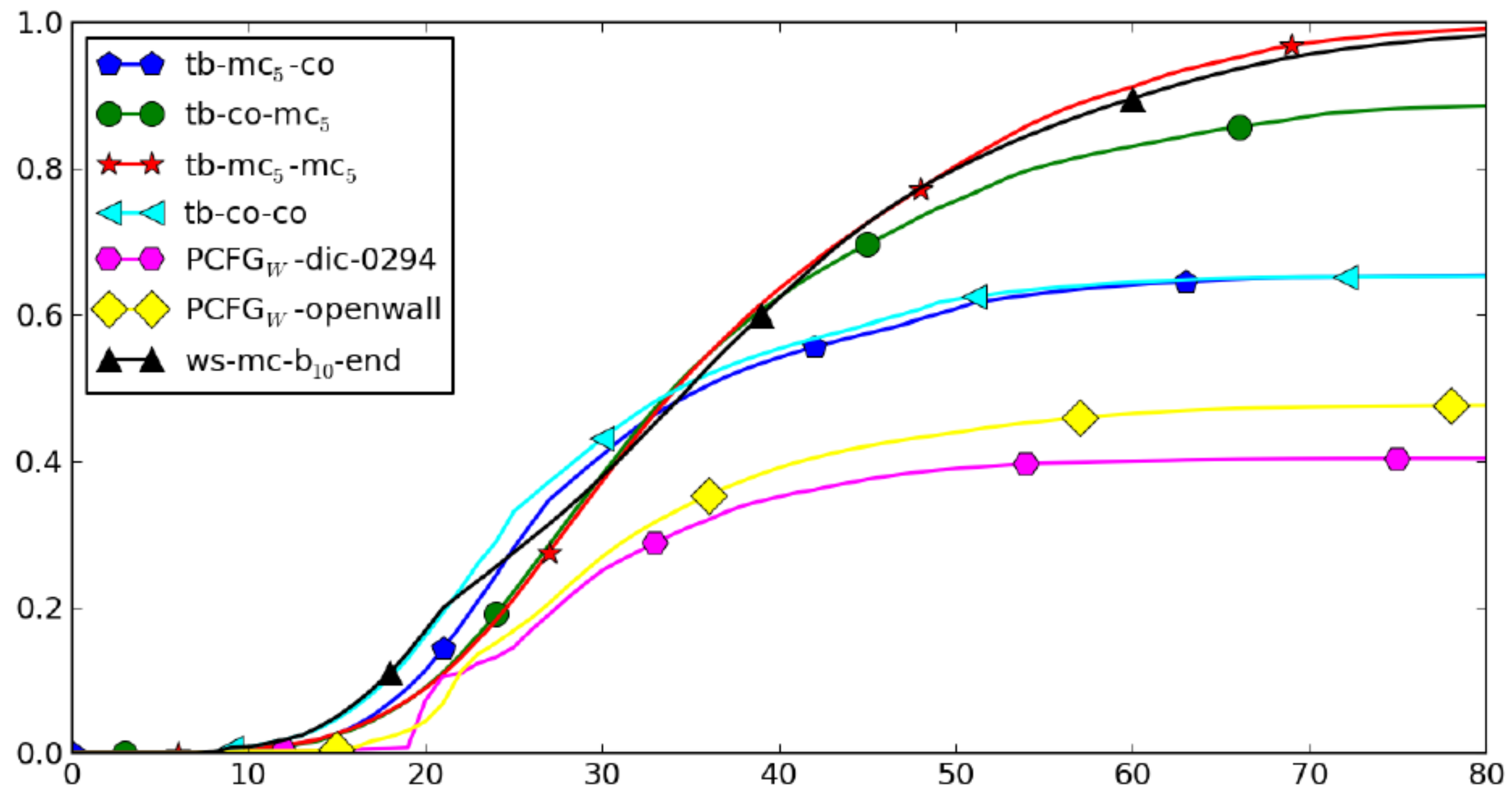


We propose: Probability-threshold Graph & ANLL

- The probability threshold in log scale vs. the percentage of passwords above the threshold.
- Only need to compute the probabilities the model assigns to each password in the testing dataset.
- ANLL (Average-Negative-Log-Likelihood) equals the area to the left of the probability-threshold curve
- ANLL$_\theta$: the area to the left of the curve below θ

Alg	ANLL$_{0.8}$
ws-mc$_1$	28.4
ws-mc$_2$	26.9
ws-mc$_3$	25.2
ws-mc$_4$	23.9
ws-mc$_5$	23.5
ws-mc-b$_{10}$	22.9





CERIAS

PURDUE
UNIVERSITY