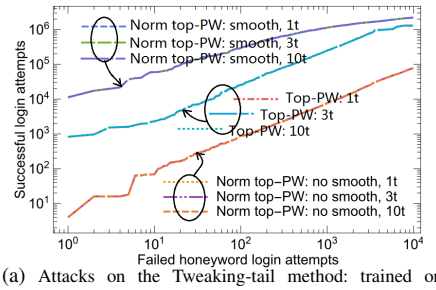
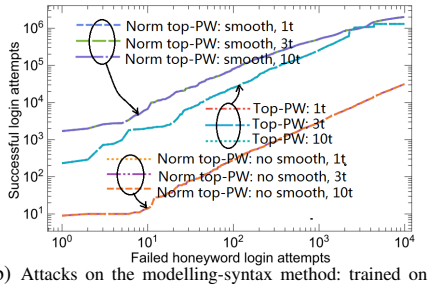
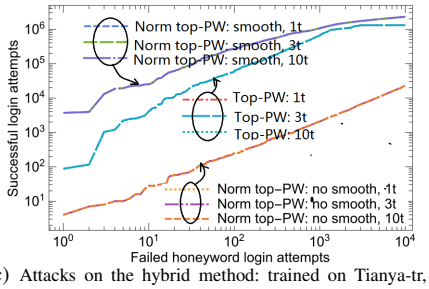


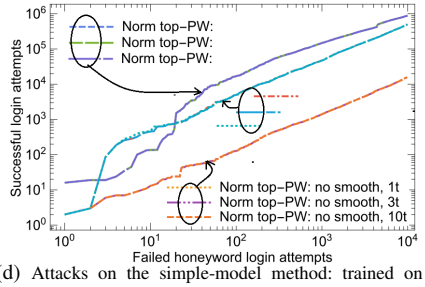
(a) Attacks on the Tweaking-tail method: trained on Tianya-tr, tested on Tianya-ts.



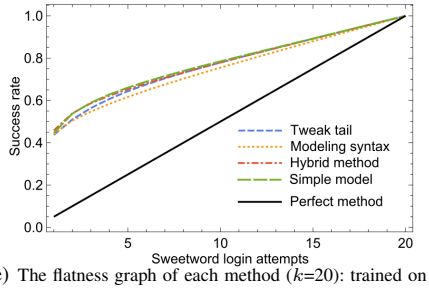
(b) Attacks on the modelling-syntax method: trained on Tianya-tr, tested on Tianya-ts.



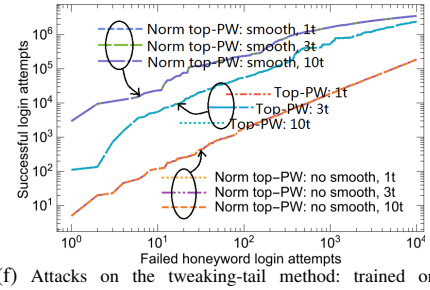
(c) Attacks on the hybrid method: trained on Tianya-tr, tested on Tianya-ts.



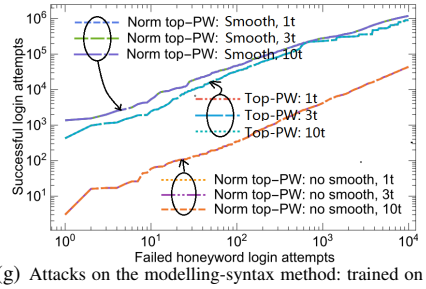
(d) Attacks on the simple-model method: trained on Tianya-tr, tested on Tianya-ts.



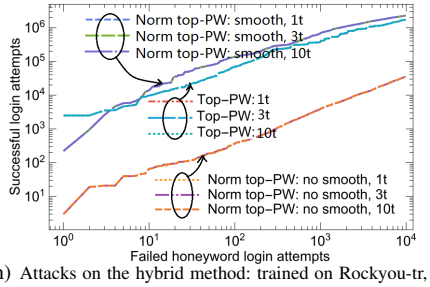
(e) The flatness graph of each method (k=20): trained on Tianya-tr, tested on Tianya-ts.



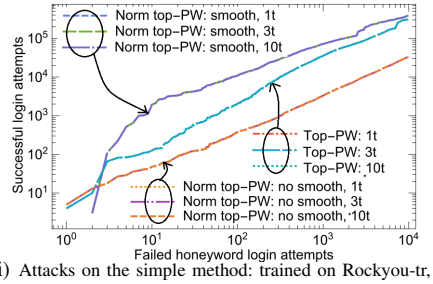
(f) Attacks on the tweaking-tail method: trained on Rockyou-tr, tested on Rockyou-ts.



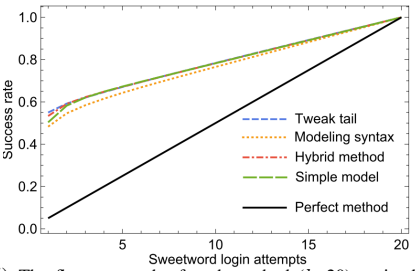
(g) Attacks on the modelling-syntax method: trained on Rockyou-tr, tested on Rockyou-ts.



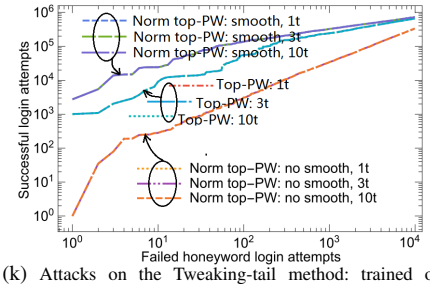
(h) Attacks on the hybrid method: trained on Rockyou-tr, tested on Rockyou-ts.



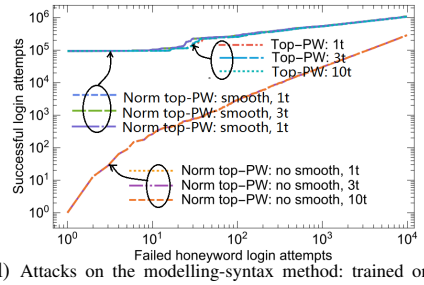
(i) Attacks on the simple method: trained on Rockyou-tr, tested on Rockyou-ts.



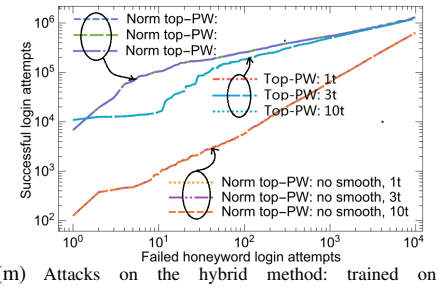
(j) The flatness graph of each method (k=20): trained on Rockyou-tr, tested on Rockyou-ts.



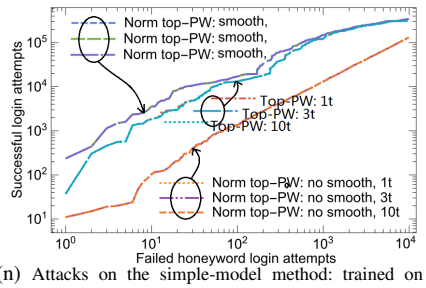
(k) Attacks on the Tweaking-tail method: trained on 000webhost-tr, tested on 000webhost-ts.



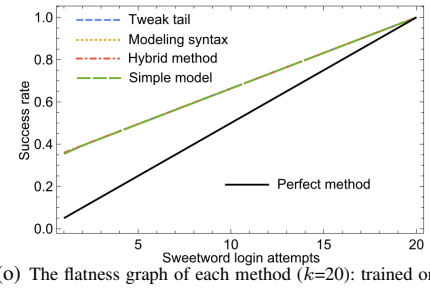
(l) Attacks on the modelling-syntax method: trained on 000webhost-tr, tested on 000webhost-ts.



(m) Attacks on the hybrid method: trained on 000webhost-tr, tested on 000webhost-ts.



(n) Attacks on the simple-model method: trained on T000webhost-tr, tested on 000webhost-ts.



(o) The flatness graph of each method (k=20): trained on 000webhost-tr, tested on 000webhost-ts.

Fig. 9. Further experiments for evaluating the four methods in [23] in terms of success-number and flatness, using datasets Tianya, Rockyou and 000webhost. Each method is evaluated by 9 attacking strategies, trained on 50% of a dataset (i.e., xxx-tr) and tested on the remaining 50% (i.e., xxx-ts). Results show that the "norm top-PW" attacking strategy with smoothing can distinguish 5.37%∼8.44% of the real PWs against these 4 methods when allowed $\mathcal{T}_2$=10^4 honeyword logins, where $1t$ means $\mathcal{T}_1 = 1$, $3t$ means $\mathcal{T}_1 = 3$ and so on; Sub-figures 9(e), 9(o) and 9(j) show that, all 4 methods are 0.3437^{+}-flat, about 7 times weaker than expected in [23].