6.885
L20.5
Nov. 22, 2004

Randomized algorithm

- replace each length with label between 1 & n' such that two lengths are equal $\Leftrightarrow$ labels match
  ($O(n)$ time w.h.p. using hashing)
- substring corresponds to base-$(n'+2)$ number
- fingerprint of substring = number modulo random prime p
- easy to maintain fingerprint when adding/removing character to/from either end of substring, in $O(1)$ time
  (add/subtract character and/or multiply/divide by p).
  [based on Karp-Rabin string matching]
- choose p (once) randomly between $2, \ldots, n^{3c}$ $(c \geq 3)$
- search for allowable fold by working in from ends (as before), shifting substrings appropriately & maintaining fingerprints
- assume fingerprint match means real match
- at the end, check for real matches; if wrong, start over
- $\Pr[\text{failure in any of } n \text{ steps}] = O(c \lg n / n^{c-2})$
  $\Rightarrow O(n)$ time w.h.p. (& in expectation)