Safra's Construction

Given a nondeterministic Büchi automaton $B = (Q, \Sigma, \Delta, s, F)$ with n states, we show how to construct an equivalent deterministic Rabin automaton R with $2^{O(n \log n)}$ states and n pairs in the acceptance condition.

Say we have a collection of *colors*, each with an associated *bell* and *buzzer*. There are several tokens of each color which will be placed on the states of B at various times, moved around, and sometimes removed. A *stack* is a pile of tokens on a state. The *height* of a stack σ is the number of tokens in σ and is denoted $|\sigma|$.

A token is *in play* at time t if it is in some stack on some state at time t. A color is *in play* at time t if it is the color of a token in play at time t. A color is *visible* at time t if it is the color of the top token of some stack at time t.

The colors in play at time t are ordered by *age*, which is the time they last came into play. All tokens of any one color in play will always come into play at the same time, therefore will have the same age. When we bring a new token into play, we always place it on top of a stack; when we remove a token from play, we always remove all the tokens above it; and when we move tokens around, we always move an entire stack at once. Thus it is an invariant of the simulation that the tokens on any stack are always ordered by age from the oldest on the bottom to the youngest on top.

The stacks are linearly ordered at time t as follows: $\sigma \ll_t \tau$ if either

- σ is a proper extension of τ (i.e., τ can be obtained by removing tokens from the top of σ); or

- neither σ nor τ is an extension of the other and σ is lexicographically older than τ; that is, starting from the bottom and moving up, at the first position where σ and τ differ in color, the color of σ is older.

The ages of colors and the order $\ll_t$ are time-dependent, because colors can go in and out of play.

The simulation starts with a single white token on the start state s. Now assume we have stacks of colored tokens on the states of B at time t. To construct the configuration at time $t + 1$, execute the following three steps. (In this construction, the intermediate configurations are transitory and do not count as states of the simulating automaton.)

Move. Suppose the next input symbol is a. For each state q, remove the stack currently on q. For each p such that $(q, a, p) \in \Delta$, clone the stack that was on q and try to put it on p. If we try to put more than one stack on p, resolve in favor of the $\ll_t$-least stack. If any color completely disappears from the board in this process, ring its buzzer.

Cover. For each final state $q \in F$, place a token of an unused color on top of q's stack. For this purpose we bring k unused colors into play, where k is the number of distinct visible colors on states $q \in F$. If two stacks on two different final states have the same visible color, then we cover them with the same new color.

This is the only way new colors can come into play; thus it is an invariant of the simulation that if a token of color c in play is directly over a token of color d, then all tokens color c in play are directly over a token of color d.

Audio Check. For every invisible color c in play, ring its bell and remove all tokens above any token of color c. Buzz the buzzers of the removed tokens. Note that if any token is removed in this process, then all tokens of that color are removed.

After these three steps are executed, there are at most n colors remaining in play; otherwise there would be an invisible one, contradicting the Audio Check step.

Claim 0.1 *B accepts x iff there is a color whose bell rings infinitely often but whose buzzer buzzes only finitely often.*

Proof. Suppose there is a color, say cyan, whose bell rings infinitely often but whose buzzer buzzes only finitely often. Let $t_0, t_1, \ldots$ be the times at which cyan's bell rings *after* its buzzer has already buzzed for the last time. From t_0 on, cyan is continuously in play, otherwise its buzzer would have buzzed. At the times t_i, all cyan tokens are visible. Between t_i and t_{i+1}, each cyan token gets covered with another token of a different color, because the only way cyan's bell can ring at time t_{i+1} is if cyan becomes invisible. Thus for every state q with a cyan token at time t_{i+1}, there is a segment of a run from some state with a visible cyan token at time t through a state of F to q. As above, König's lemma can be used to construct a run with infinitely many occurrences of states in F.

Conversely, suppose there is an accepting run ρ of B. Let σ_t be the stack on the state of ρ at time t. Let $m = \liminf |\sigma_t|$; that is, m is the maximum height such that from some point on, the stacks along ρ reach height m and then never go below height m again. Note that $m \geq 1$ since white (the oldest color) is always in play, and $m \leq n$ since there are at most n colors in play. From some point on, the stack is of height at least m, and infinitely often exactly m. After that point, the colors on the stack at height m and below may change due to being replaced by a $\ll_t$-lesser stack in the Move step, but this can happen only finitely often because lexicographic order on the age space ω^m is well-founded. So from some point on, the colors at height m and below do not change. Say the color at height m at this point is magenta. Then infinitely often after that point, the run goes through a state of F, so the stack acquires a new token; but sometime after that the stack must shrink back to height m again, and the only way that can happen is if magenta's bell rings. Thus magenta rings infinitely often and buzzes only finitely often. $\square$

A state of the simulating Rabin automaton R will consist of a stack of tokens (possibly empty) on each state of B, the current ordering $\ll_t$, and an indication of which bells rang. Each token configuration can be specified by a map $Q \to \{\text{colors}\}$ giving the top color of the stack on each state and a map $\{\text{colors}\} \to \{\text{colors}\} \cup \{\text{none}\}$ telling for each color in play the color below it. There are at most $n!$ orderings $\ll_t$ and 2^n ways the bells can ring. Thus there are at most $n^n \cdot n^n \cdot n! \cdot 2^n = 2^{O(n \log n)}$ states in all. The number of Rabin pairs is n, the number of colors.