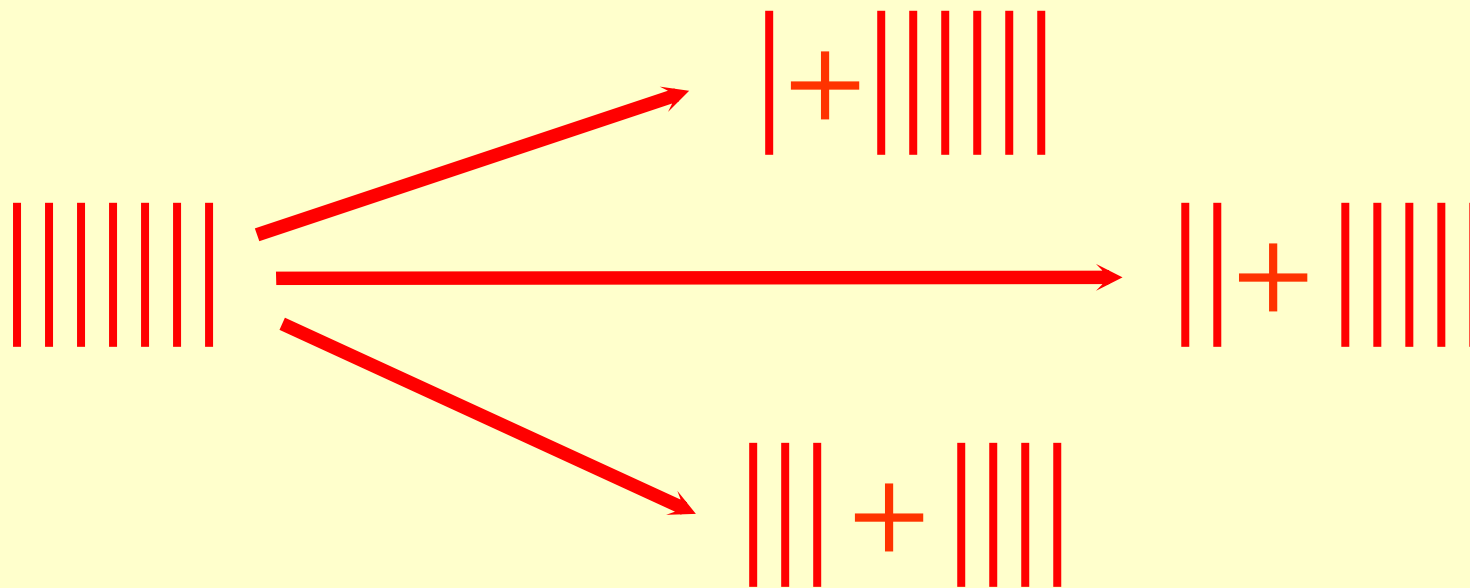


Implementasi Reinforcement Learning pada game NIM

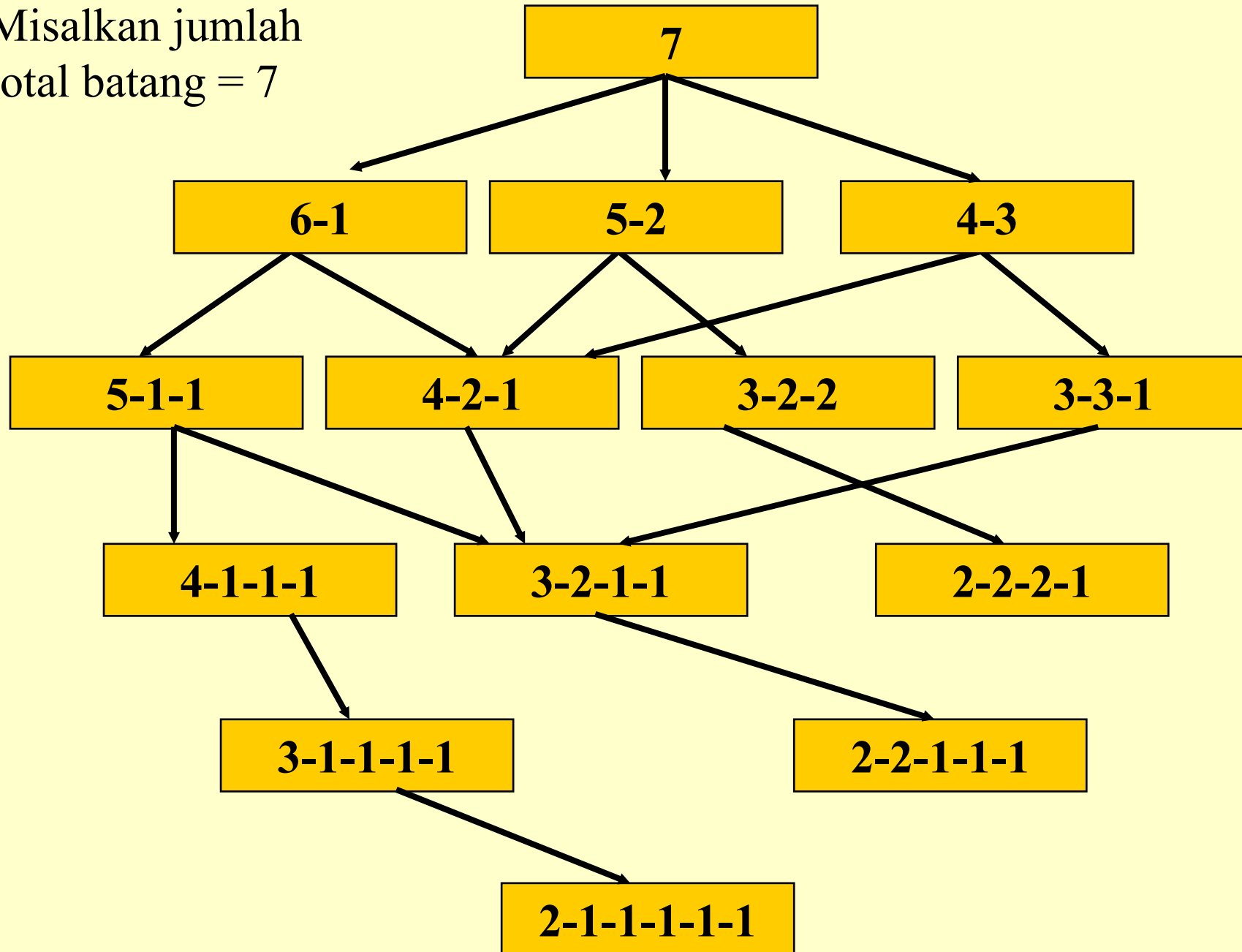
Ali Ridho Barakbah

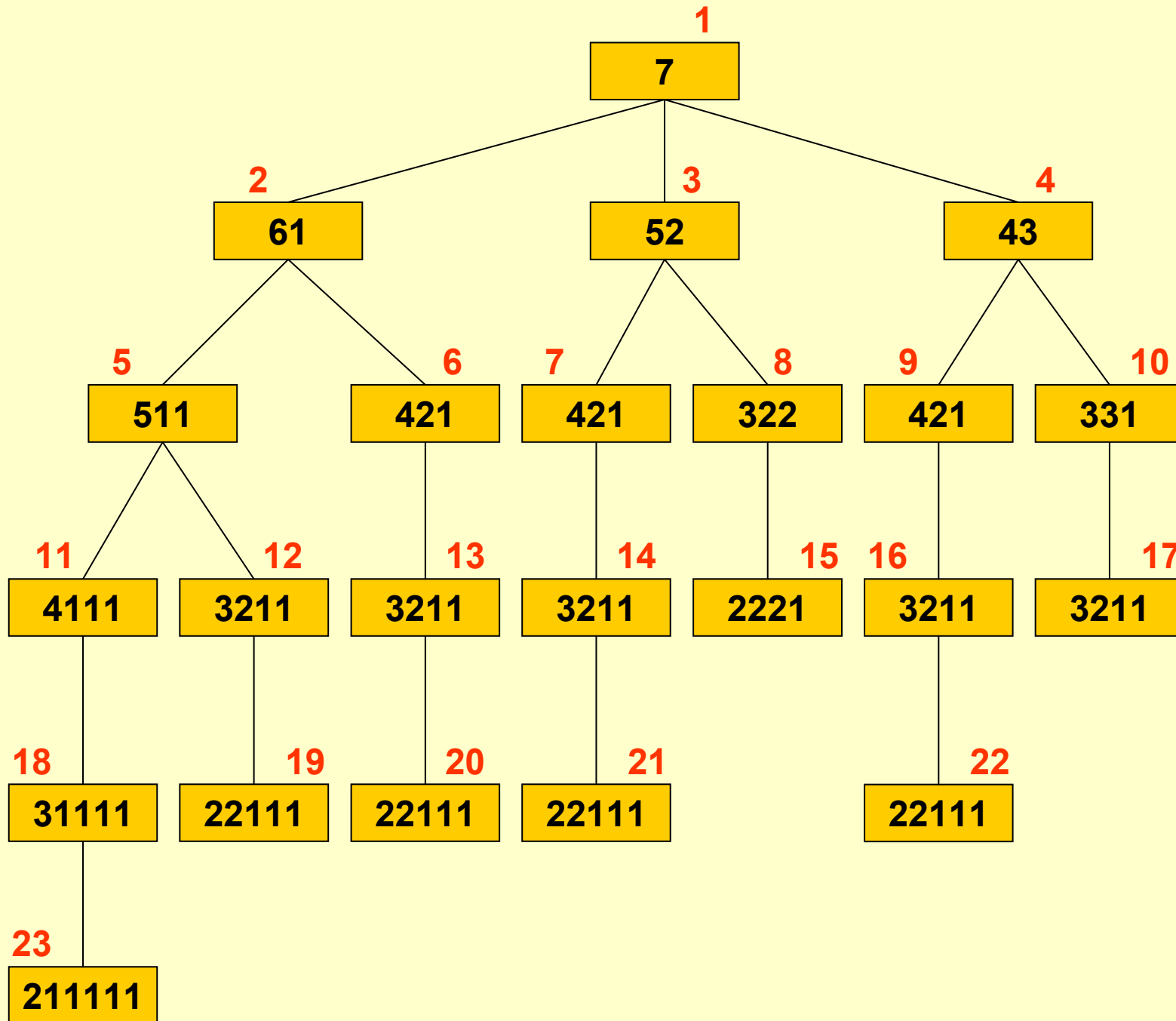
Deskripsi

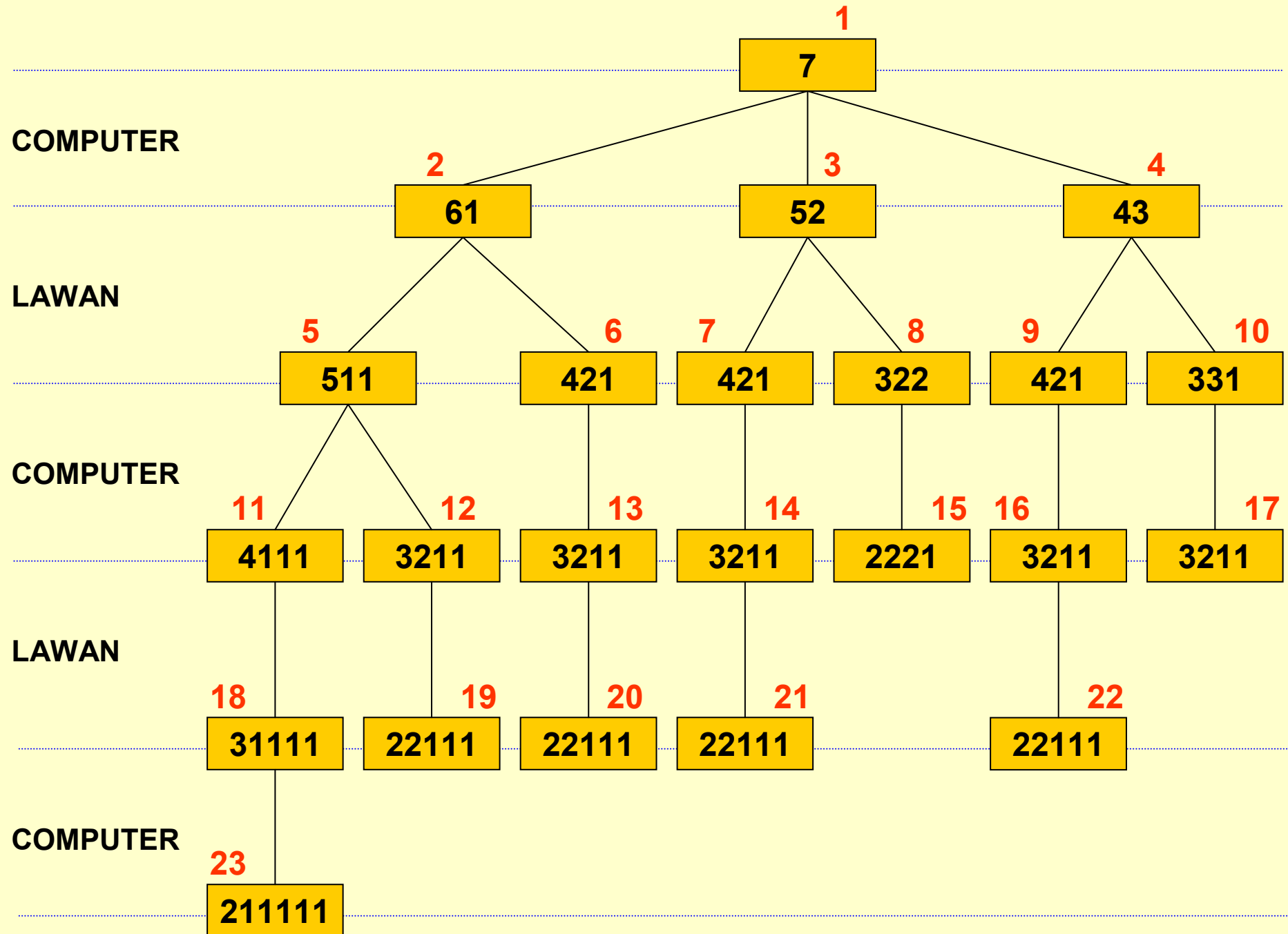
- Diawali serangkaian batang
- Setiap pemain harus memecah serangkaian batang menjadi 2 kumpulan dimana jumlah batang di tiap kumpulan tidak boleh sama dan tidak boleh kosong

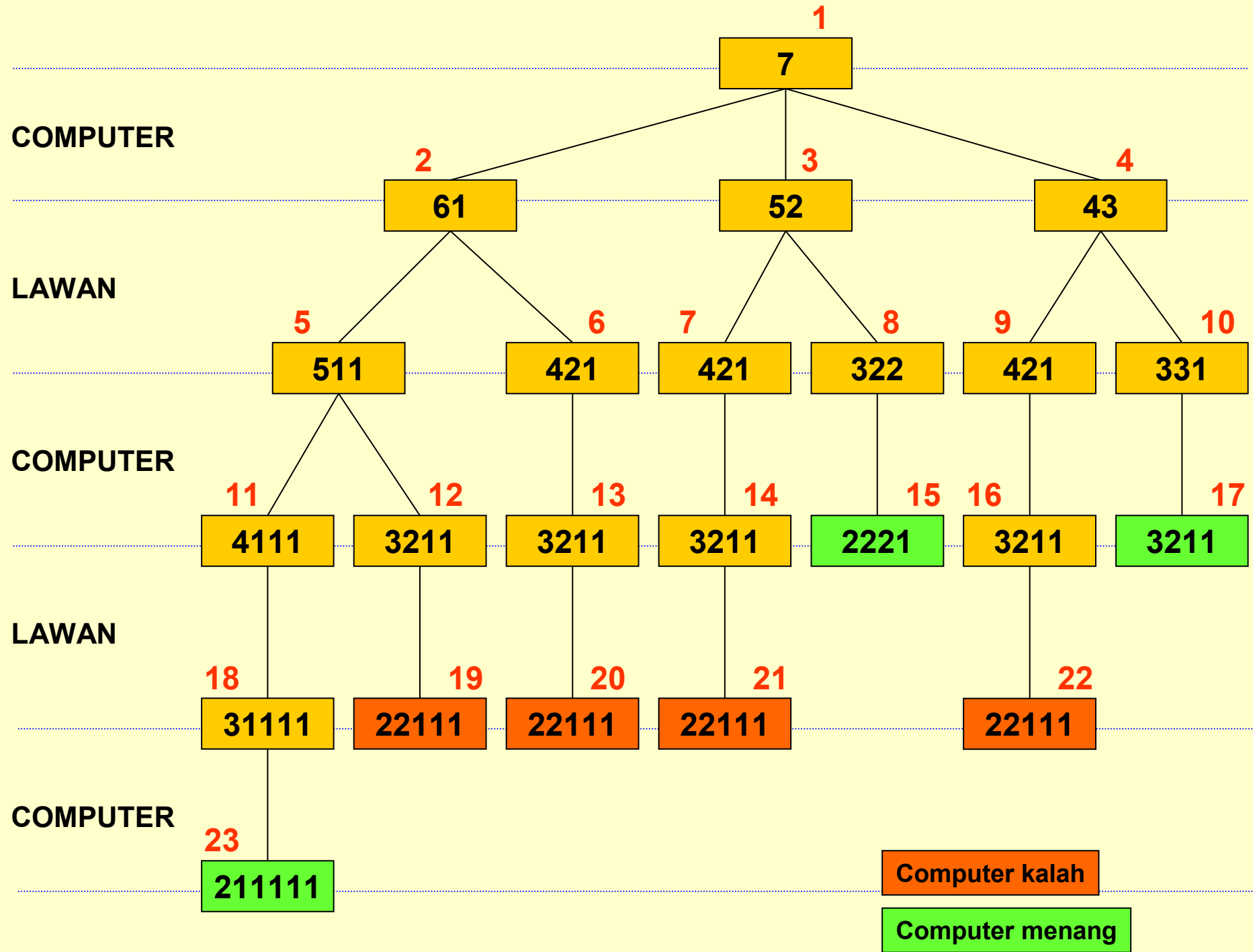


Misalkan jumlah total batang = 7







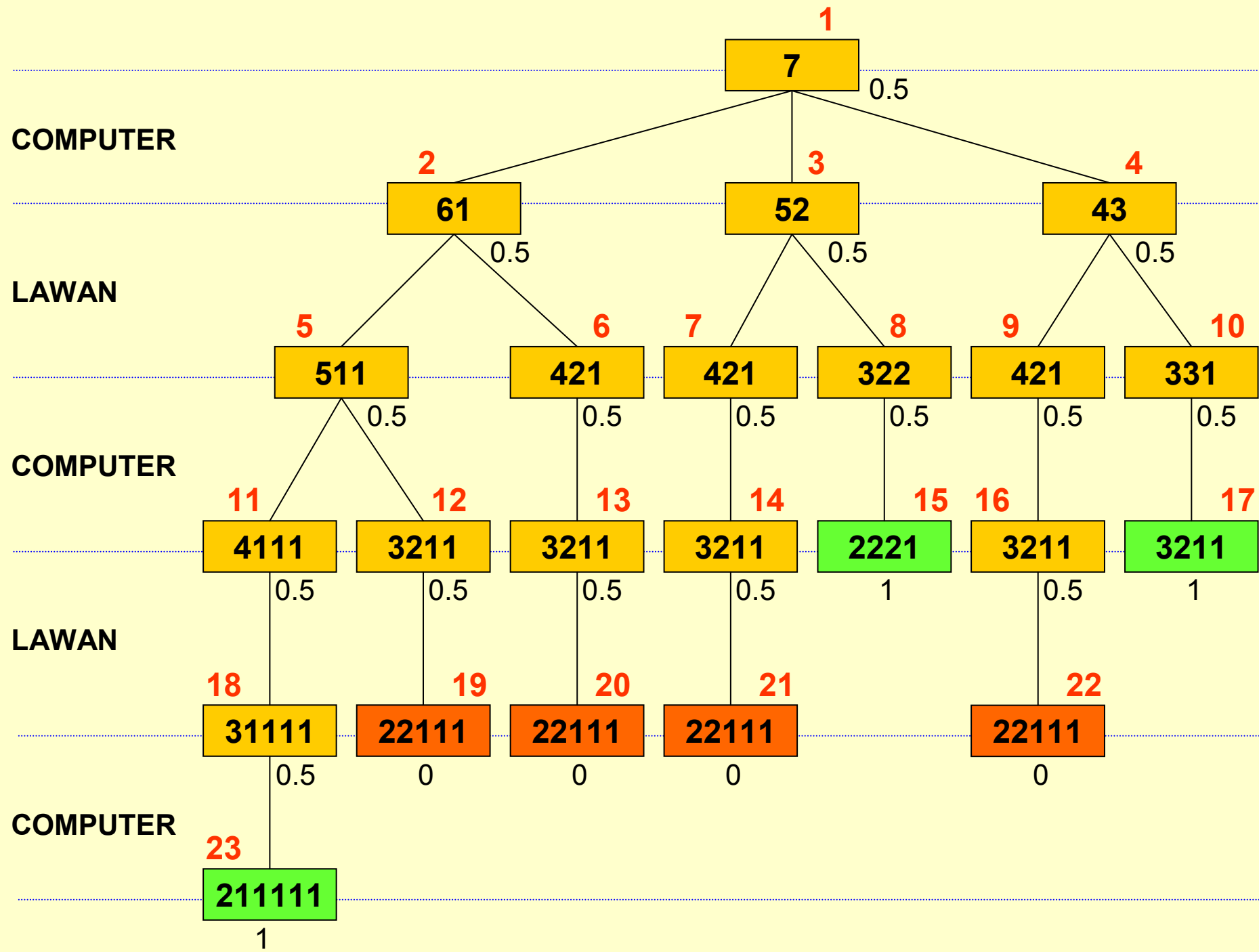


Value function

- Uses Temporal Difference Learning

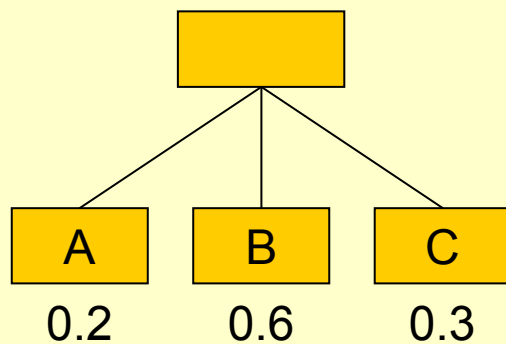
$$V(s) = V(s) + \alpha [V'(s) - V(s)]$$

- Initialization
 - Computer menang : $V(s) \leftarrow 1$
 - Computer kalah : $V(s) \leftarrow 0$
 - State lainnya : $V(s) \leftarrow 0.5$
 - $\alpha = 0.1$



Exploratory move

- Greedy move
 - memilih $V'(s)$ yang terbesar
- ϵ -greedy move
 - Kemungkinan besar memilih $V'(s)$ yang terbesar
 - Ada kemungkinan memilih yang lain secara random

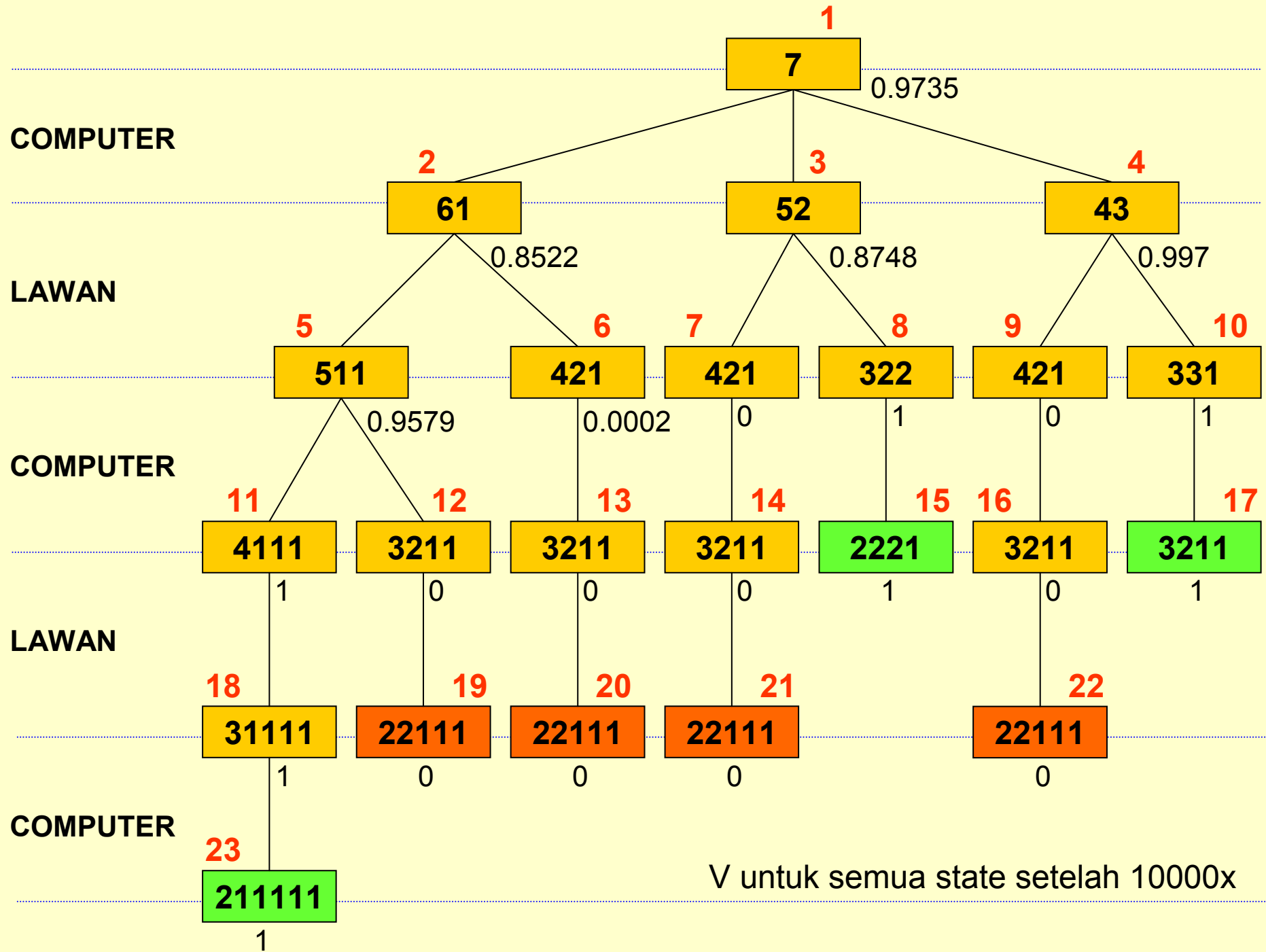


ϵ -greedy = 10%

```
p ← random[0..1]
if (p < 0.1) {
  anakterpilih = pilih anak secara random
} else {
  anakterpilih = pilih anak yang mempunyai  $V'(s)$  terbesar
}
```

Hasil eksperimen

State	10x	100x	1000x	10000x
1	0.5004	0.8906	0.9746	0.9735
2	0.5042	0.9478	0.7871	0.8522
3	0.5	0.5	0.9534	0.8748
4	0.5	0.5262	0.9163	0.997
5	0.5265	0.9948	0.9052	0.9579
6	0.5	0.4066	0.0994	0.0002
7	0.5	0.495	0.2745	0
8	0.5	0.5	1	1
9	0.5	0.495	0.4429	0
10	0.5	0.672	0.9983	1
11	0.6126	0.9991	1	1
12	0.5	0.405	0.0608	0
13	0.5	0.2152	0.0236	0
14	0.45	0.405	0.1029	0
15	1	1	1	1
16	0.5	0.405	0.2657	0
17	1	1	1	1
18	0.8063	0.9999	1	1
19	0	0	0	0
20	0	0	0	0
21	0	0	0	0
22	0	0	0	0
23	1	1	1	1



Selamat mencoba