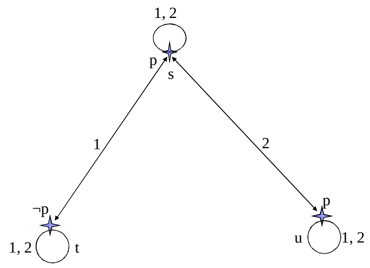
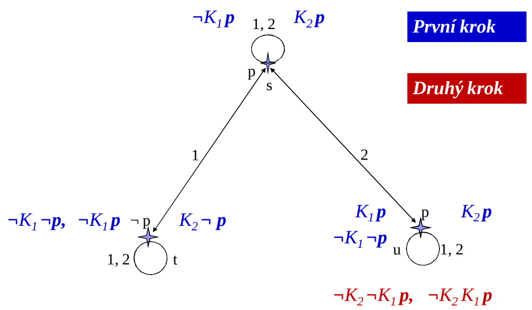
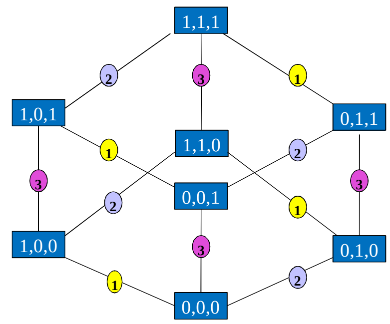
Znalosti v multi-agentních systémech



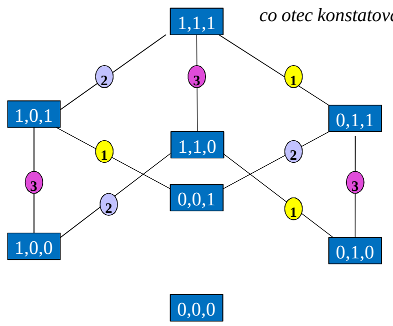


Pozor! Formule $\neg K_1 \neg p$ a $\neg K_1 p$ **nevyjadřují totéž**: srovnej jejich platnost ve stavech **t** a **u.**

Kripkeho struktura pro 3 ušmudlané děti



Kripkeho struktura pro 3 ušmudlané děti poté, co otec konstatoval ***p***



Příklad. Karetní hra 1

Karetní hra pro **3** hráče, jejichž sada karet obsahuje právě **4 ESA** a **4 DESÍTKY**. Každý „vidí“ jen karty svých spoluhráčů a ne ty svoje. Postupně hlásí, zda umí určit své karty - vyhrává první „znalý“!

Máme **4 ESA** a **4 DESÍTKY**, tedy každý z hráčů **1,2,3** může mít buď **DD, DE** nebo **EE.**

Kolo a)

1. Hráč3 vidí, že 1EE a 2DD.
2. Hráč1 i hráč2 ohlásili, že nemohou určit své karty
3. Může hráč3 vyhrát?

Ano, stačí zvážit všechny 3 možnosti a vyloučit ty nemožné!

Kolo b)

1. Jste **hráč1** a vidíte, že 2DD a 3DE.
2. Hráč1, hráč2 i hráč3 v prvním kole ohlásili, že nemohou určit své karty.
3. Může hráč1 v příštím tahu vyhrát?

Kolo c)

1. Jste **hráč2** a vidíte, že 1DE a 3DE.
2. Hráč1, hráč2 i hráč3 v prvním kole ohlásili, že nemohou určit své karty.
3. Hráč1 ani v dalším kole nemůže určit své karty.
4. Víte už, jaké karty držíte?

Příklad – pokračování, karetní hra 1

Máme **4 ESA** a **4 DESÍTKY**, tedy každý z hráčů **1,2,3** může mít buď **DD, DE** nebo **EE.**

Φ = {1**DD,** 1**DE,** 1**EE,** 2**DD,** 2**DE,** 2**EE, …**}

S = { (DD-DD-EE), (DD-DE-DE), (DD-DE-EE), …}

π((DD-DD-EE))(2**DD** & 3**EE**) = *true*

π((DD-DD-EE))(3DD) = *false* ...

$M = (S, \pi, K_1, K_2, K_3)$

Jak se vyjádří, že agent 2 neví jaké on sám má karty?

Např. K_2 (2**DD** ∨ 2**DE** ∨ 2**EE)** & ¬ K_2 **DD** & ¬ K_2 **DE** & ¬ K_2 **EE**

Příklad. Karetní hra 2

G = { 1, 2 } hrají dva hráči 1 a 2

c = { A, B, C} tři karty A, B, C

Φ = { 1A, 1B, 1C, 2A, 2B, 2C}

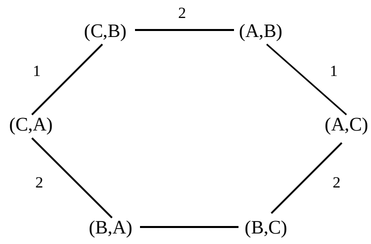
první hráč drží kartu A…

S = { (A,B), (A,C), (B, A), (B, C), (C, A), (C, B) }

množina stavů: první hráč drží A, druhý B, ...

π((A, B))(1A) = *true* π((A, B))(1B) = *false* ...

$M = (S, \pi, K_1, K_2)$



$K_1 = \{[(A,B),(A,C)],[(B,A),(B,C)],[(C,B),(C,A)]\}$

$K_2 = \{[(C,A),(B,A)],[(A,B),(C,B)],[(A,C),(B,C)]\}$

Tento příklad ukazuje, že je potřebné do struktury zařadit i stavy, které agent nepovažuje za možné.

Například ve stavu (*A,B*) *agent1* ví, že stav (*B,C*) není možný. (*Agent1* velmi dobře ví, že drží v ruce kartu *A*.)

Nicméně *agent1* považuje za možné, že *agent2* považuje za možný stav (*B,C*), musíme proto tento stav zařadit do Kripkeho struktury. To je v grafu znázorněno tím, že z uzlu (*A,B*) nevede do uzlu (*B,C*) žádná hrana ohodnocená číslem *1*.

Přitom taková hrana vede z uzlu (*A,B*) do uzlu (*A,C*), ze kterého dále vede hrana do uzlu (*B,C*) ohodnocená číslem *2*.

Zatím jsme se podrobněji nezabývali jazykem, který jsme použili v tomto příkladu. Protože se zajímáme uvažováním o tom, který agent drží kterou kartu, je vhodné za množinu prvotních výroků vzít množinu {*1A, 2A, 1B, 2B,* …}, jejíž prvky interpretujeme výroky „*agent1 drží kartu A*“, „*agent2 drží kartu A*“ atd.

Při této interpretaci prvotních výroků definujeme funkci π zřejmým způsobem.

Je-li *M* Kripkeho struktura popisující tuto karetní hru, potom například platí

$$(M,(A,B)) \models 1A \wedge 2B$$

Snadno se ověří

$$(M,(A,B)) \models K_1(2B \lor 2C)$$

$$(M,(B,C)) \models K_2(2C) \land K_2(1A \lor 1B)$$

$$(M,(A,B)) \models C_G(1A \lor 1B \lor 1C)$$

$$(M,(A,B)) \models C_G(1B \to (2A \lor 2C))$$

$$(M,(A,B)) \models D_G(1A \land 2B)$$

Příklad 3 o 3 mudrcích

Král měl 5 klobouků stejného tvaru a velikosti: 3 z nich byly červené a 2 bílé. Král je ukázal mudrcům a každému z nich nasadil jeden tak, aby mudrc neviděl jeho barvu, ale viděl barvy klobouků těch ostatních.

Pak se začal postupně ptát 1. a 2. mudrce:

- Víte jakou barvu má váš klobouk?
- Oba odvětili NE.
- Třetí mudrc pak řekl „Já už vím!“

Víte také?

Nechť **M** = (S, π,.. K_1, K_2, K_3,..., K_n) je libovolná Kripkeho struktura a **s**∈ S libovolný její stav. Ověřte, že pro libovolné formule ***A***, ***B*** platí

i. **(M, s)** |= (K_i ***A*** & K_i (***A***→***B***)) → K_i ***B***
ii. **(M, s)** |= K_i ***A*** → ***A***
iii. **(M, s)** |= K_i ***A*** → $K_i K_i$ ***A***
iv. **(M, s)** |= ¬ K_i ***A*** → K_i (¬ K_i ***A***)

Lemma.

$(i)\ (M,s) \models E_G^k A \iff (M,t) \models A$ *pro každé t, G − dosažitelné v k krocích*

$(ii)\ (M,s) \models C_G A \iff (M,t) \models A$ *pro každé t, G − dosažitelné z s.*

Důkaz.

(i) se dokáže indukcí podle k, (ii) plyne z (i).

Obě tato tvrzení platí pro libovolné relace přípustnosti K_i (nemusí jít jen o ekvivalence, neboť důkaz nevyužívá žádnou speciální vlastnost relací přípustnosti).