PDV 11 2020/2021

Globální stav a jeho snapshot

Michal Jakob
michal.jakob@fel.cvut.cz
Centrum umělé inteligence, katedra počítačů, FEL ČVUT

Dotaz: NTP – frekvence aktualizace

Standardně se tzv. *polling interval* pohybuje v intervalu **[64s, 1024s]**

Polling interval je průběžně upravován sofistikovaným algoritmem.

Polling interval závisí na zaznamenaném driftu a náhodných fluktuacích (jitter).

- Pokud je chyba stabilní, tak z výchozí hodnoty 64s se postupně polling interval prodlužuje.
- Pokud chyba vzroste, tak se sníží.

Více info:
https://www.eecis.udel.edu/~mills/ntp/html/assoc.html#poll

Globální Stav

Globální stav: množina lokální **stavů procesů** v DS a **stavů** všech **komunikačních kanálů** v jednom okamžiku.

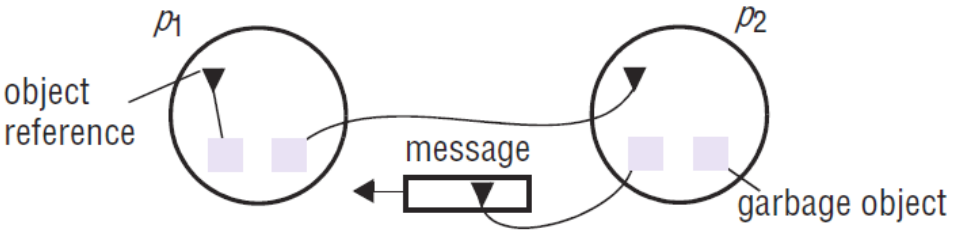
Globální stav se mění, kdykoliv dojde v systému **k akci** (události)

- vykonání instrukce v procesu
- poslání nebo přijetí zprávy

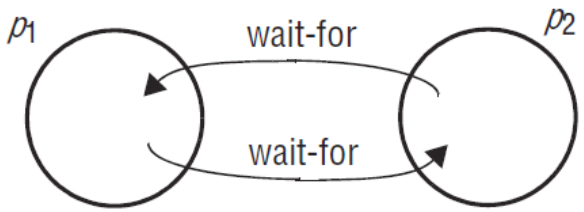
Globální snapshot: záznam globální stavu.

Příklady

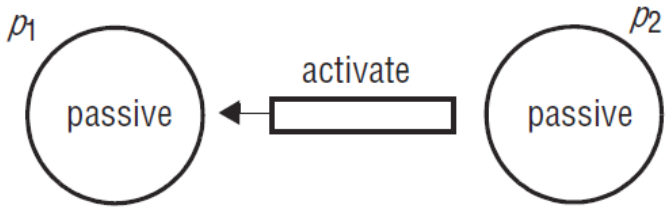
Garbage collection: nutnost identifikovat objekty, na které není *globálně* žádná reference.



Detekce uváznutí (deadlock): nutnost identifikovat cykly *globálním* wait-for grafu.



Detekce ukončení výpočtu: nutnost zajistit, že *všechny* procesy jsou pasivní a v *žádném* kanálu přenosu není žádná potenciálně aktivační zpráva.



Checkpointing za účelem obnovení globálního stavu systému.

…

Globální snapshot

Pokud bychom měli **globální hodiny**, tak zaznamenat snapshot globálního stavu by bylo jednoduché: všechny procesy by zaznamenaly stav v dohodnutý čas, tj. v jednom **fyzickém okamžiku**.

Fyzický okamžik by garantoval **konzistenci** zaznamenaného globálního stavu.

Jak zaznamenat globální snapshotů **bez** globálních hodin?

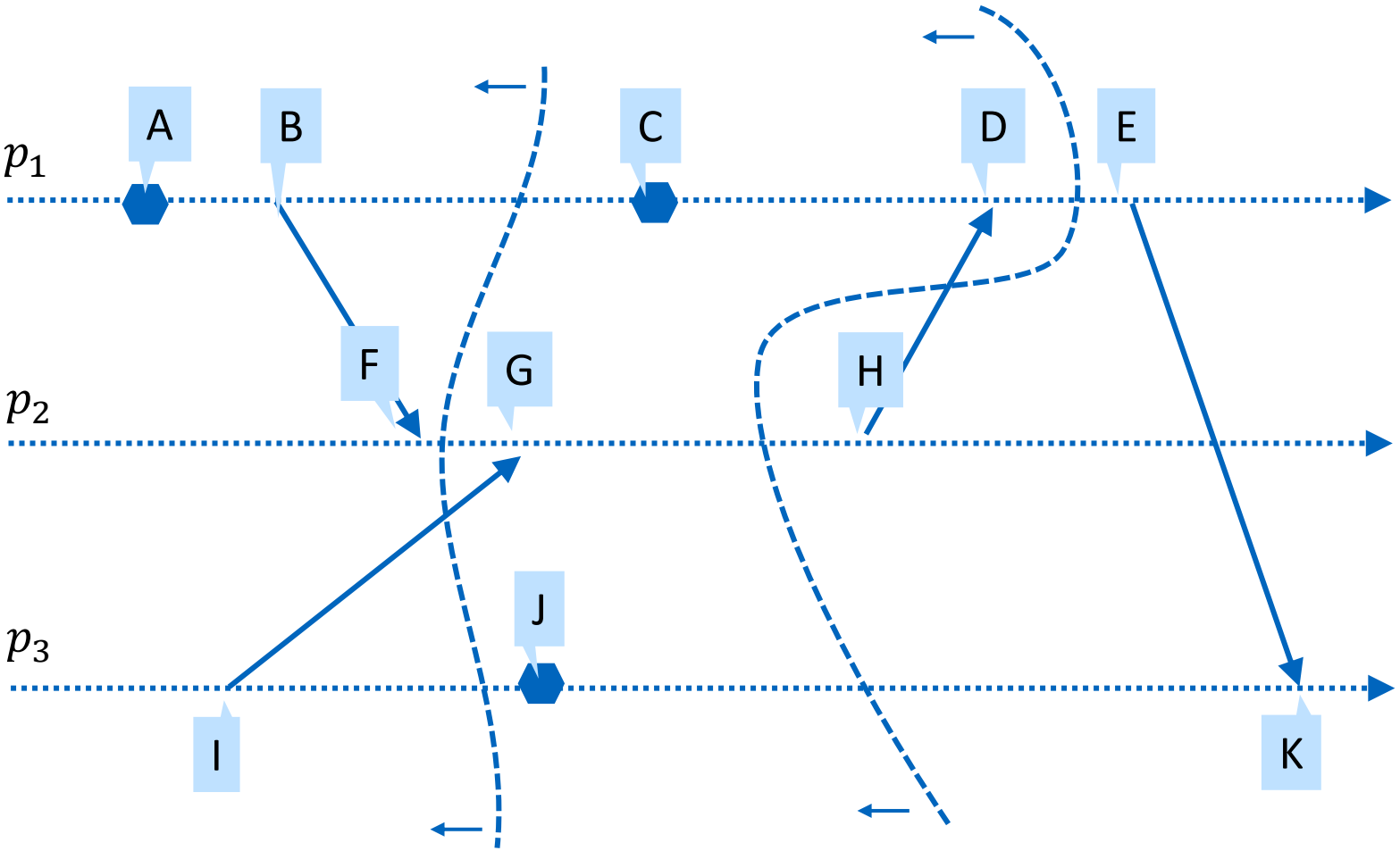
→ **Logický okamžik**

Řez distribuovaného výpočtu

Řez: časová hranice v každém procesu a v každém komunikačním kanále.

- události, které nastanou před řezem, jsou **v řezu**
- události, které nastanou po něm, jsou **mimo řez**.

Řez: Příklad



Konzistentní řez

Definice

Řez R je konzistentní pokud splňuje kauzalitu, tj. pokud pro každý pár událostí e, f v systému platí:

$$f \in R \wedge e \rightarrow f \Rightarrow e \in R$$

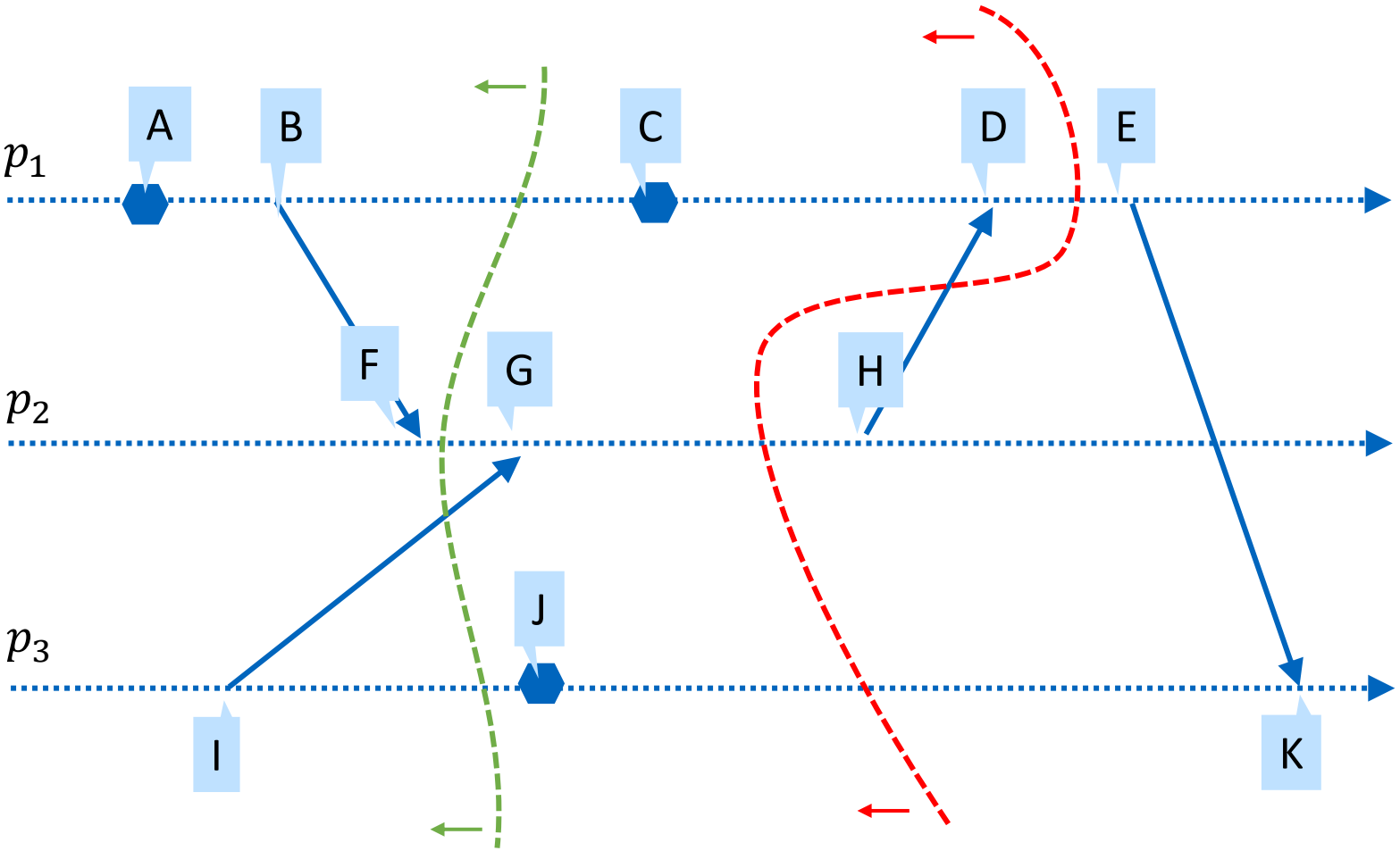
předcházejí dle relace **stalo se před** (tj. nelze, aby v řezu byl „**důsledek**“ a nebyla tam „**příčina**“.)

Konzistentní řez = **logický okamžik**

Konzistentní globální snapshot odpovídá konzistentnímu řezu.

- Globální snapshot je konzistentní, pokud by **mohl** (ale nikoliv musel) být zpozorován externím pozorovatelem daného distribuovaného výpočtu.
- Nekonzistentní řez by nemohl **nikdy** být zpozorován externím pozorovatelem daného distribuovaného výpočtu.

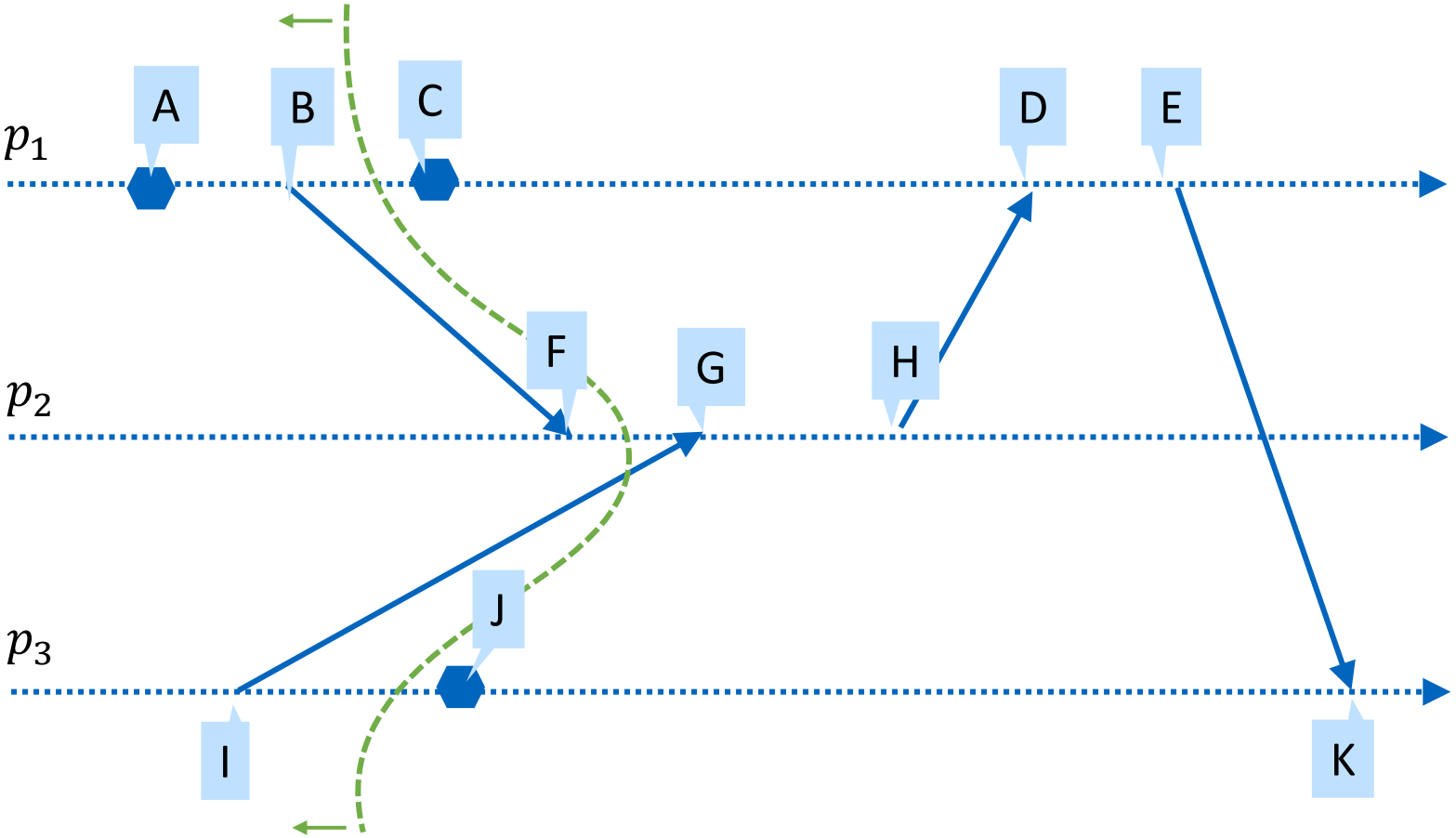
Řez: Příklad



konzistentní řez
(mohl být zpozorován)

nekonzistentní řez
(*nikdy* nemohl být zpozorován)

Řez: Příklad



tentýž **konzistentní** řez

(ale nemohl být v reálu zpozorován –
F nastalo ve fyzickém čase až po C a J)

Výpočet konzistentního globálního snapshotu

Problém výpočtu globálního snapshotu

Cíl: Zaznamenat **globální snapshot**, tj. stav pro každý proces a každý komunikační kanál.

Požadavek: Zaznamenání snapshotu by neměla **interferovat** s během distribuované aplikace a neměla by vyžadovat po aplikaci zastavení posílání aplikačních zpráv.

Model
(Existují i algoritmy se slabšími předpoklady)

- skupina N procesů $p_1, \dots, p_N$
- **FIFO perfektní komunikační kanál** mezi každým párem procesů, tj. zprávy se neduplikují, nevznikají, neztrácejí a jsou doručovány v pořadí odeslání (značení: $C_{i,j}$ kanál přenáší zprávy z procesu p_i do procesu p_j)
- **asynchronní** systém: neznáma, ale **konečná latence**

Předpoklad: Každý proces je schopen zaznamenat svůj vlastní **aplikační stav** (případně low-level systémový stav).

Chandy-Lamport algoritmus pro distribuovaný globální shapshot

(Vytváření snapshotu je distribuované.)

Speciální zpráva: **ZNAČKA ▪**

Jeden (libovolný) z procesů iniciuje vytvoření snapshotů.

Procesy reagují na příjem zprávy **ZNAČKA ▪ .**

Chandy-Lamport algoritmus pro globální shapshot

Zahájení tvorby snapshotu

Iniciující **proces** p_i odešle ZNAČKU ▪ všem ostatním procesům (i sobě)

Příjem ZNAČKY ▪ procesem p_i kanálem $C_{m,i}$

if (p_i dosud nezaznamenal svůj stav) **then**
- p_i **zaznamená** svůj stav;
- p_i **zaznamená** stav kanálu $C_{m,i}$ jako prázdnou množinu;
- p_i *každým* odchozím **kanálem $C_{i,j}$ odešle** jednu ZNAČKU ▪ (předtím než skrze $C_{i,j}$ pošle jakoukoliv jinou zprávu);
- p_i **zapne zaznamenávání** zpráv doručených skrze všechny ostatní příchozí kanály $C_{j,i}$ *kromě* $C_{m,i}$

else
- p_i **zaznamená** stav kanálu $C_{m,i}$ jako množinu všech zpráv, které p_i obdržel skrze $C_{m,i}$ od doby, kdy zahájil záznam $C_{m,i}$;
- p_i **ukončí** záznam kanálu $C_{m,i}$;

end if

Ukončení

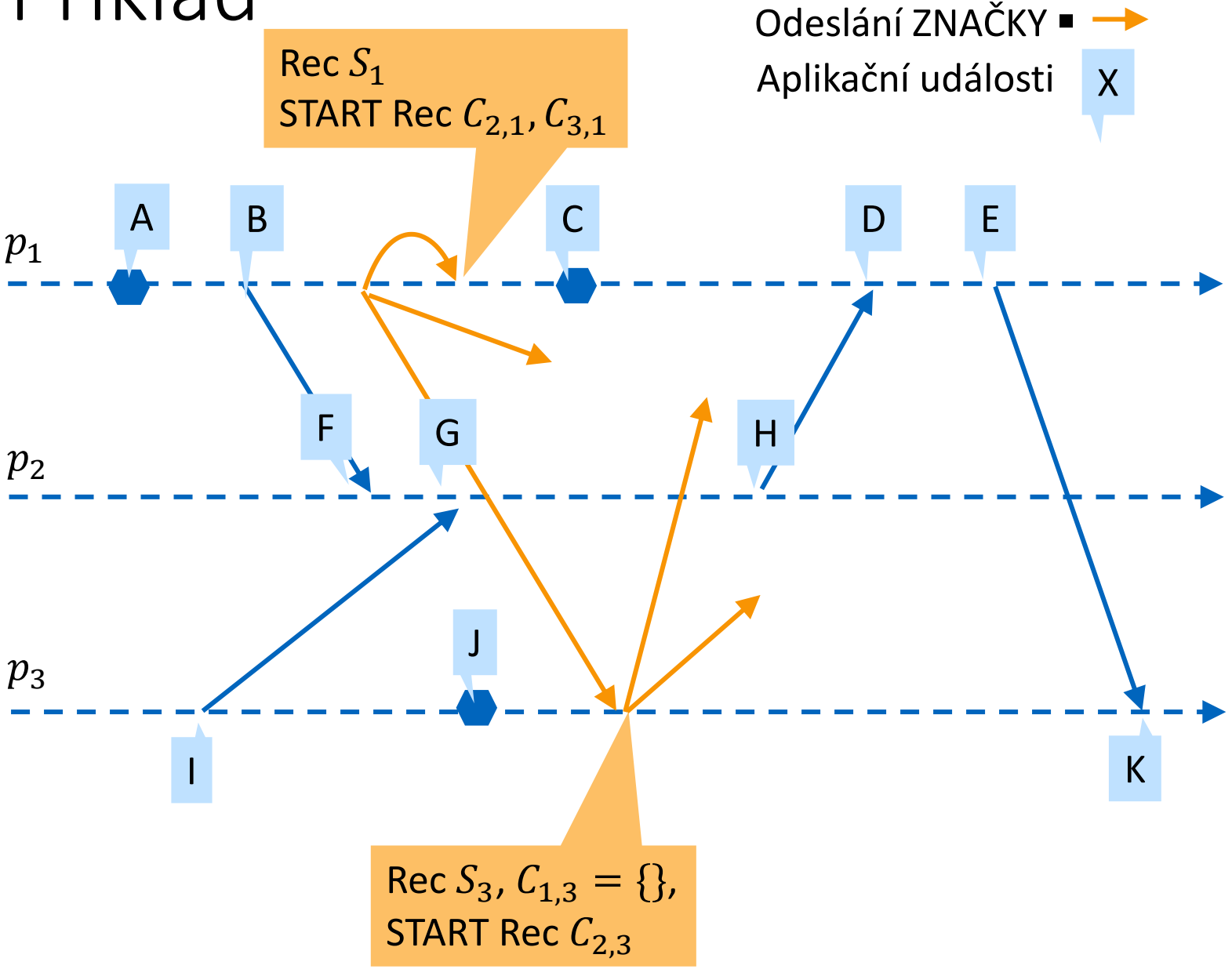
Ukončení algoritmus

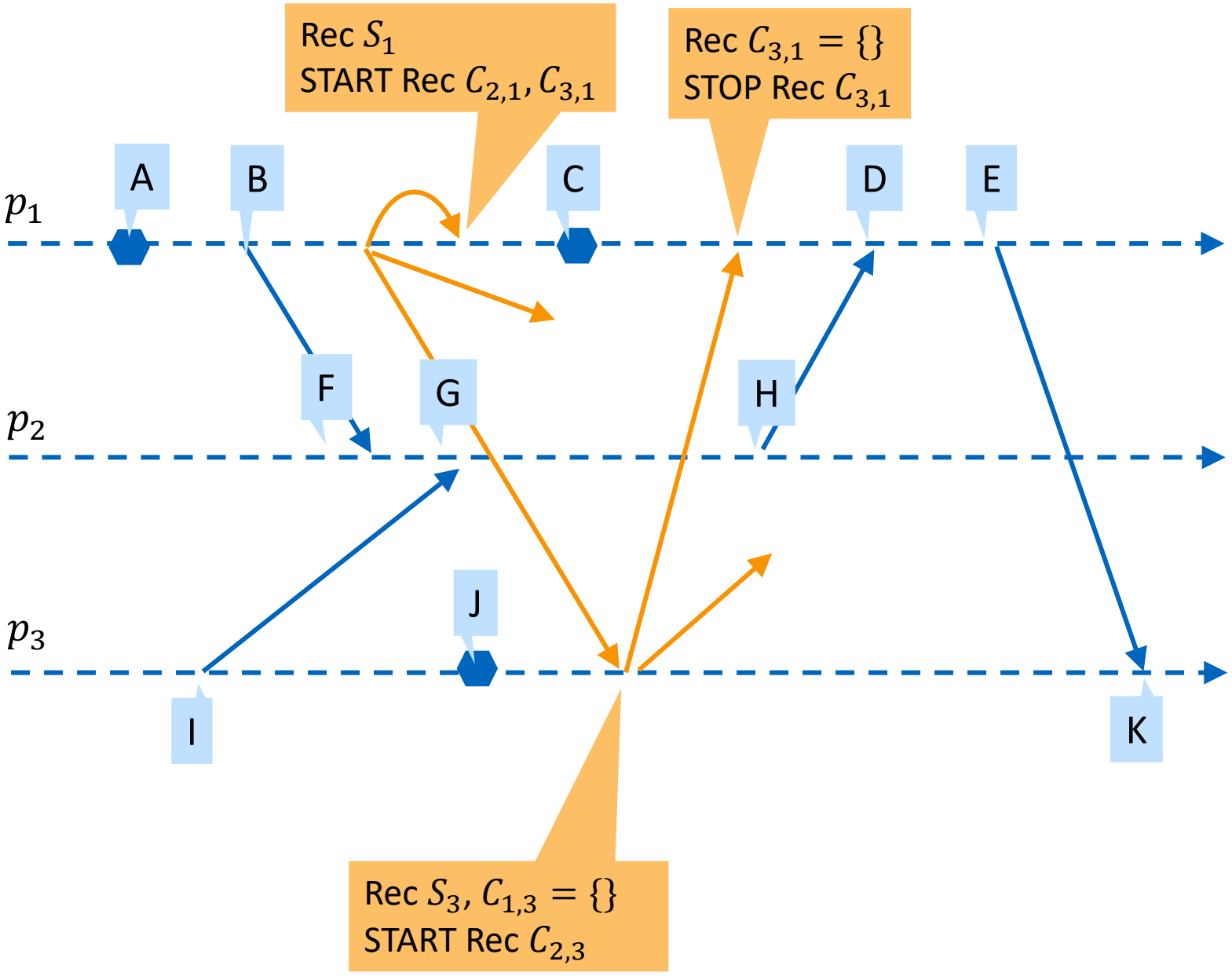
Algoritmus je ukončen jakmile:

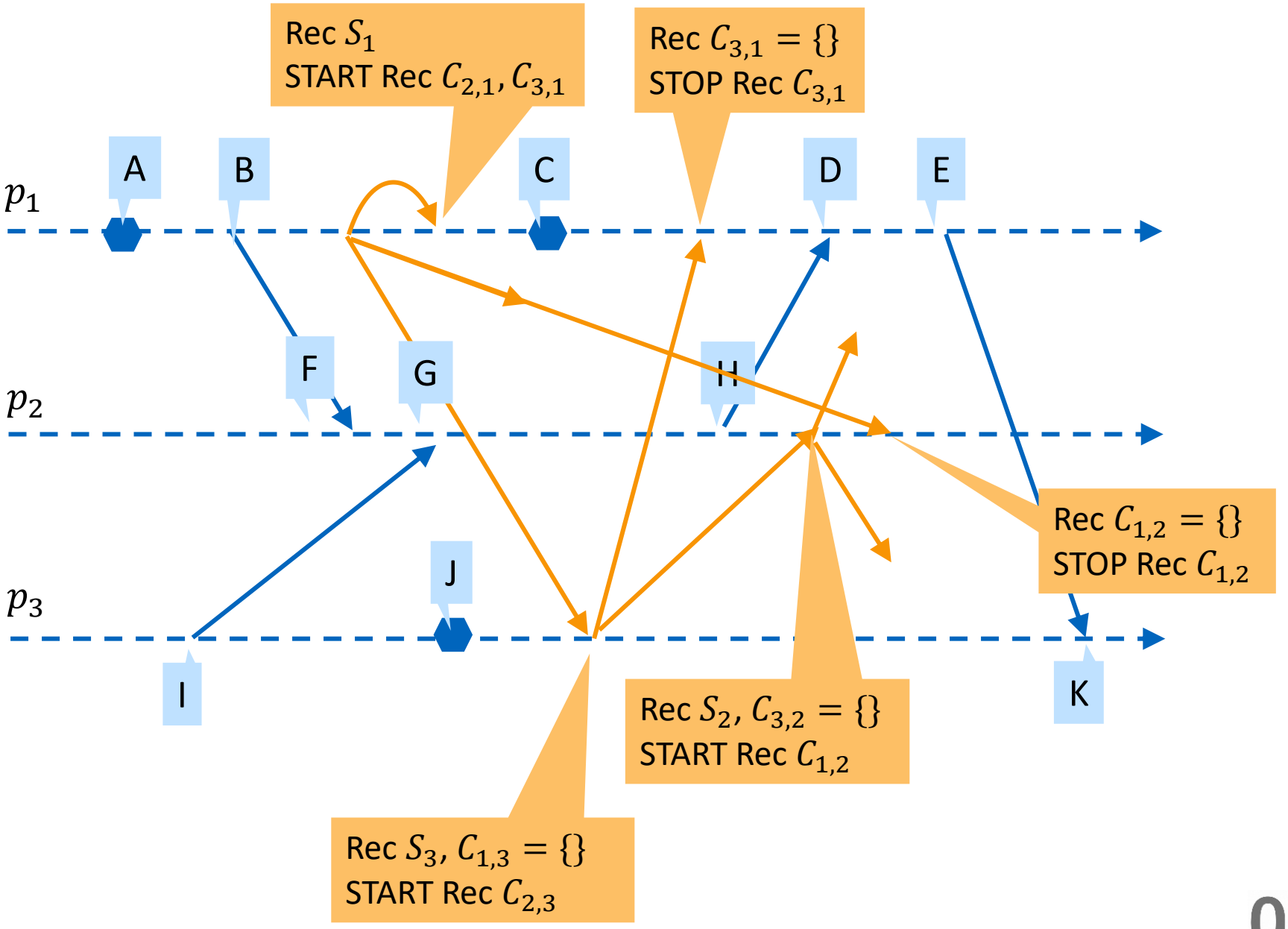
1. všechny procesy zaznamenaly svůj stav
   a
2. všechny procesy přijaly ZNAČKU ▪ na všech $N - 1$ svých příchozích kanálech (aby zaznamenaly stav na všech kanálech)

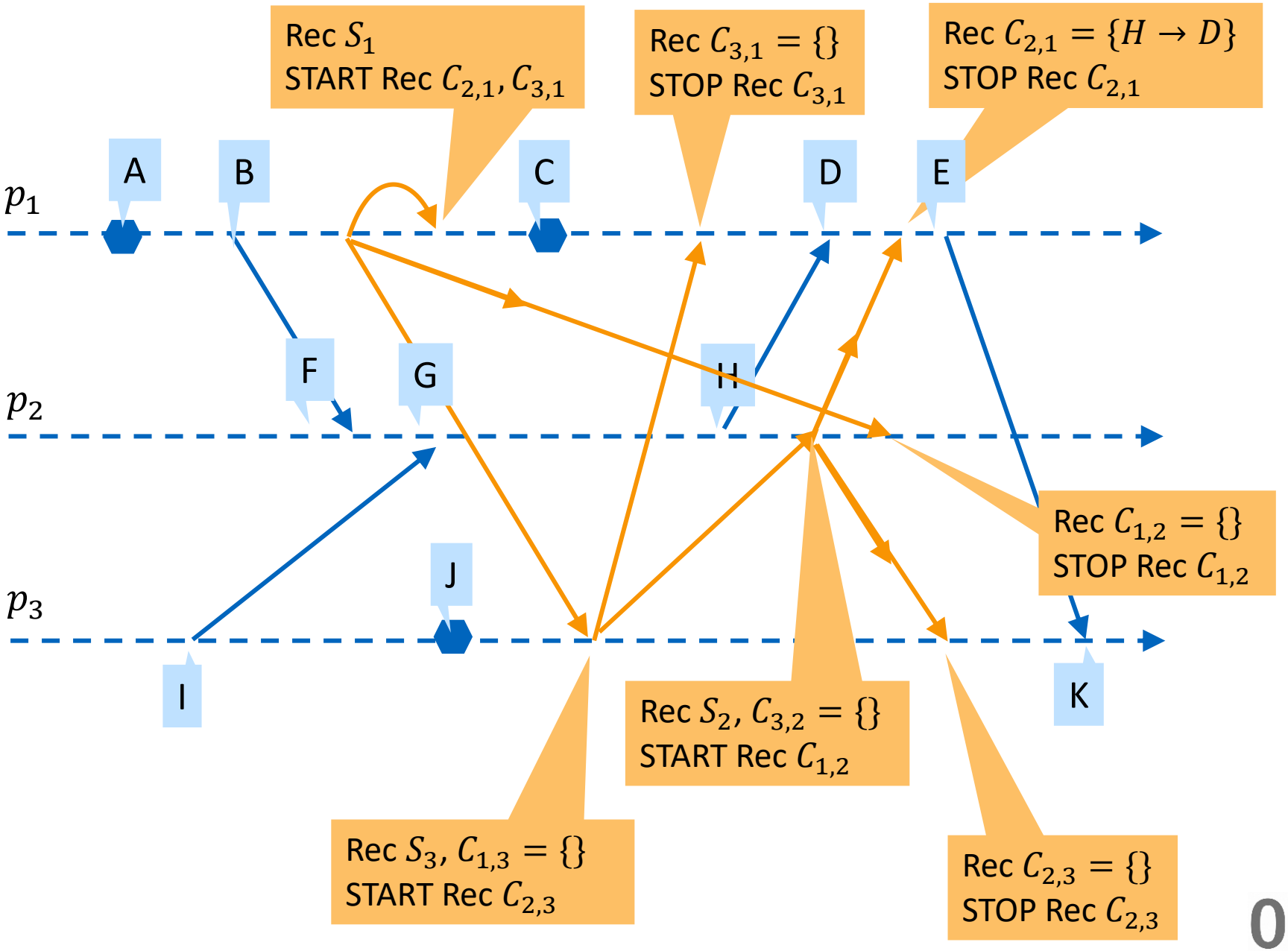
Následně (je-li potřeba) mohou být jednotlivé fragmenty globální stavu posbírány centrálním servery a poskládán **plný globální snaphot**.

Příklad

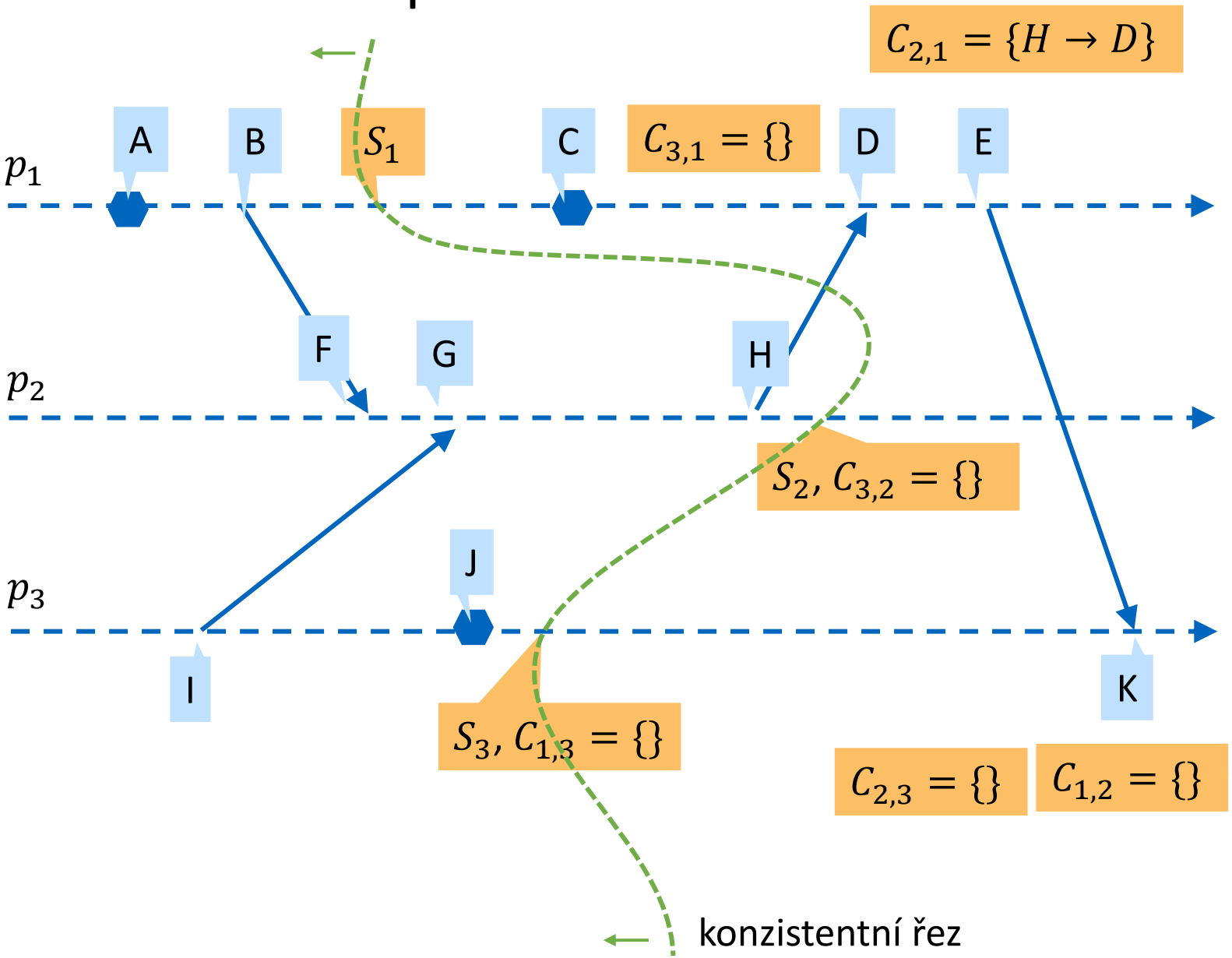

Odeslání ZNAČKY
Aplikační události
X
Rec S_1
START Rec $C_{2,1}, C_{3,1}$
p_1
p_2
p_3
A
B
C
D
E
F
G
H
I
J
K
Rec S_3, $C_{1,3} = \{\}$,
START Rec $C_{2,3}$


Rec S_1
START Rec $C_{2,1}, C_{3,1}$
Rec $C_{3,1} = \{\}$
STOP Rec $C_{3,1}$
A
B
C
D
E
p_1
F
G
H
p_2
J
p_3
I
K
Rec S_3, $C_{1,3} = \{\}$
START Rec $C_{2,3}$


Rec S_1
START Rec $C_{2,1}, C_{3,1}$
Rec $C_{3,1} = \{\}$
STOP Rec $C_{3,1}$
A
B
C
D
E
p_1
F
G
H
p_2
J
p_3
I
K
Rec $C_{1,2} = \{\}$
STOP Rec $C_{1,2}$
Rec $S_2, C_{3,2} = \{\}$
START Rec $C_{1,2}$
Rec $S_3, C_{1,3} = \{\}$
START Rec $C_{2,3}$


Rec S_1
START Rec $C_{2,1}, C_{3,1}$
Rec $C_{3,1} = \{\}$
STOP Rec $C_{3,1}$
Rec $C_{2,1} = \{H \rightarrow D\}$
STOP Rec $C_{2,1}$
p_1
p_2
p_3
A
B
C
D
E
F
G
H
I
J
K
Rec $C_{1,2} = \{\}$
STOP Rec $C_{1,2}$
Rec $S_2, C_{3,2} = \{\}$
START Rec $C_{1,2}$
Rec $S_3, C_{1,3} = \{\}$
START Rec $C_{2,3}$
Rec $C_{2,3} = \{\}$
STOP Rec $C_{2,3}$

Finální Snapshot



Vlastnosti

Vlastnosti Chandy-Lamport

Výsledkem Chandy-Lamport algoritmu pro výpočet globální snapshotu je **konzistentní řez.**

Důkaz:

- nechť R je výsledný řez a e_i a e_j jsou události v procesech P_i a P_j tak, že $e_i \rightarrow e_j$
- pro konzistenci třeba dokázat implikaci: $e_j \in R \Rightarrow e_i \in R$
- tj. pokud $e_j \rightarrow$ "P_j zaznamená svůj stav" pak $e_i \rightarrow$ "P_i zaznamená svůj stav"
- Přepokládejme, že tato implikace neplatí, tj. $e_j \rightarrow$ "P_j zaznamená svůj stav" a "P_i zaznamená svůj stav" $\rightarrow e_i$
- Uvažujme cestu aplikačních zpráv z e_i do e_j (přes další procesy): vzhledem k FIFO kanálu musí ZNAČKY na všech spojích výše uvedené cesty předcházet aplikační zprávy
- Tedy, protože "P_i zaznamená svůj stav" $\rightarrow e_i$, tak musí platit, že P_j obdržel svoji ZNAČKU (a tedy i zaznamenal svůj stav) před e_j
  (tj. nemůže nastat: "P_i zaznamená svůj stav" $\rightarrow e_i \rightarrow e_j \rightarrow$ "P_j zaznamená svůj stav")
- A tedy $e_j \notin R$ → spor

Vyhodnocení stabilních vlastností

Uvažujme **vlastnost distribuovaného výpočtu** jako logickou formuli definovanou nad globálním stavem distribuovaného výpočtu.

Stabilní vlastnost je taková vlastnost, že jakmile je ve výpočtu **jednou** splněna, zůstává splněna **navždy**.

- příklad stabilní vlastnost živosti: výpočet skončil
- příklad stabilní vlastnosti porušující bezpečnost: nastalo uváznutí, objekt je sirotek (neukazuje na něj žádná reference)

Chandy-Lamport algoritmus lze použít pro detekci **stabilních** globálních vlastností.

(Lze ukázat, že pokud nějaká stabilní vlastnost splněna v globálním snapshotu zachyceným snapshot algoritmem, bude splněna i výpočtu splněna i ve fyzickém okamžiku doběhnutí snapshotů algoritmu. Naopak, pokud ve vypočteném globálním snapshotu nějaká stabilní vlastnost splněna není, nemohla být ve výpočtu splněna v okamžiku zahájení snapshotů algoritmu).

Shrnutí

Schopnost **zachytit globální stav** distribuovaného výpočtu je důležitá.

Vytvoření snapshotů by nemělo nijak omezovat probíhající výpočet.

Chandy-Lamportův algoritmus vypočte globální snapshot.

Vypočtený globální snapshot odpovídá **konzistentnímu řezu**.

Globální snapshot může být využit k **detekci stabilních vlastností** výpočtu.