Počítačové siete
Spanning Tree Protocol (STP)

Viacnásobné prepoje switch-ov

- prečo to spraviť
  - odolnosť voči výpadku linky
  - čiastočná odolnosť voči výpadku switch-a
    - výpadok switch-a neznefunkční celú sieť – len časť, ktorá nemá iné prepoje
    - dôležité uzly môžu byť pripojené do viacerých switch-ov, čím zostanú prístupné aj po výpadku switch-a
- prečo to nejde „len tak“
  - vznik slučiek (loop)
    - poslanie rámca na všetky porty spôsobí zacyklenie a zahltenie siete

Viacnásobné prepoje switch-ov

- riešenie
  - deaktivovať nadbytočné linky
    - tak, aby vždy existovala práve jedna cesta medzi ľubovoľnými 2 uzlami siete – vytvorenie **kostry** siete (kostra = spanning tree)
- požiadavky
  - automatická konfigurácia
  - malá časová a komunikačná náročnosť
  - schopnosť automaticky reagovať na zmenu podmienok
    - pridanie/odobratie switch-a, zlyhanie/obnovenie linky, …
  - deterministickosť a ovplyvniteľnosť (pomocou parametrov) výberu použitých switch-ov a liniek v kostre

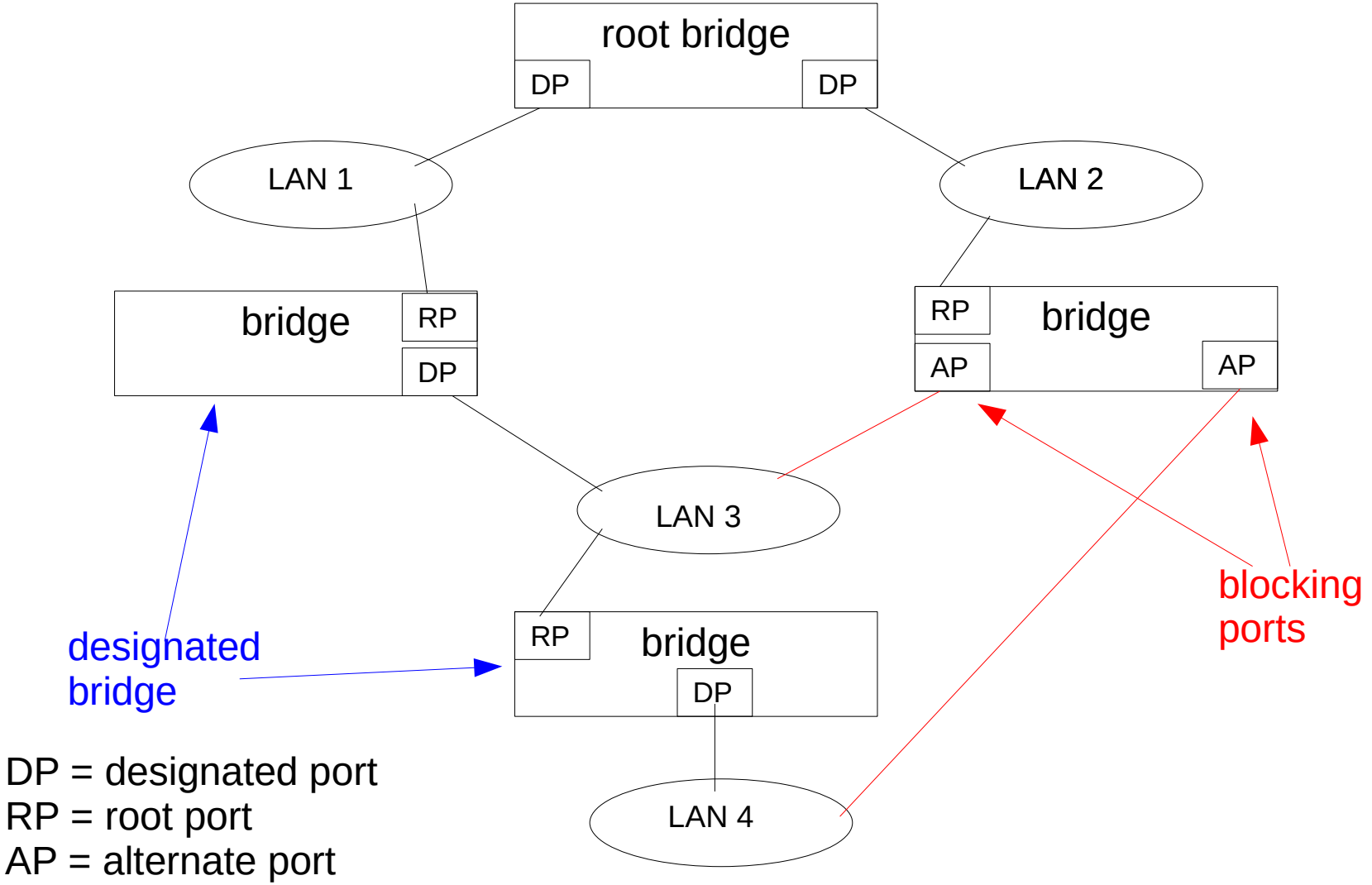
STP – predpoklady

- multicastová adresa pre adresáciu všetkých switch-ov
  - „Bridge Group Address“
  - 01:80:C2:00:00:00
- jedinečný identifikátor switch-a
  - priotita + MAC adresa
- jedinečný identifikátor portu switch-a
  - priorita + číslo portu

STP – základné pojmy

- root bridge
  - switch zvolený za koreň kostry
- designated bridge (pre segment siete)
  - switch, cez ktorý sa prenášajú dáta z/do segmentu
- root port
  - port smerom k root bridge-u
- designated port (pre segment siete)
  - port, cez ktorý sa prenášajú dáta z/do segmentu

STP – základné pojmy



DP = designated port
RP = root port
AP = alternate port

STP – základné pojmy

- path cost (pre každý port)
  - „cena“ linky pripojenej k portu
  - konfigurovateľná
- root path cost (pre každý port)
  - „cena“ cesty z portu k root bridge-u

STP – základné pojmy

- root bridge = bridge s najnižším ID
  - najvýznamnejšia je priorita, potom MAC adresa
- designated port pre segment
  - port s najnižšou root path cost – najlacnejšou cestou ku koreňu
  - ak ich je viac, tak rozdhoduje ID bridge-a
  - ak ich je stále viac, tak ID portu
- designated bridge pre segment
  - bridge obsahujúci designated port pre segment

STP – princíp

- bridge-e si posielajú BPDU (Bridge Protocol Data Unit)
  - ID odosielajúceho bridge-u, ID odosielajúceho portu, ID root bridge a root path cost
- na začiatku si každý myslí, že je root bridge a designated bridge pre všetky pripojené segmenty
- na základe prijatej BPDU bridge koriguje svoju predstavu o root bridge, a designated bridge pre segment
- root bridge pravidelne posiela BPDU
- keď bridge dostane z RP BPDU s lepšou alebo rovnakou informáciou, pošle novú informáciu cez všetky DP
- keď bridge dostane z DP BPDU s horšou informáciou, odpovie cez tento port vlastnou informáciou

STP – princíp

- informácie získané cez STP majú timeout
  - vypršanie timeout-u signalizuje zlyhanie
- ak vyprší informácia prijatá cez niektorý port, bridge sa pokúsi stať designated bridge pre tento segment
- ak vyprší informácia prijatá z RP, bridge si určí nový RP
- ak sa úplne stratí informácia o root bridge, switch sa pokúsi stať root bridge-om

STP – stavy portov

- blocking
  - port neprenáša dáta, ale prijíma BPDU
- listening
  - port ešte neprenáša dáta, čaká na stabilizáciu topológie
  - následne sa buď vráti do blocking alebo postúpi do learning
- learning
  - port ešte neprenáša dáta, ale učí sa MAC adresy
  - po uplynutí stanoveného času prejde do forwarding
  - môže prejsť späť do blocking
- forwarding
  - port plne funkčný
  - môže prejsť späť do blocking

STP – notifikácie o zmenách

- pri zmene topológie
  - môže dôjsť k „presunu“ uzlov siete
  - bridge pošle notifikáciu, že došlo k zmene root bridge-u
    - špeciálny TCN BPDU
  - root bridge pošle notifikáciu všetkým bridge-om
    - nastavením flagu v BPDU
  - bridge-e znížia čas na expiráciu záznamov o MAC adresách

BPDU

- BPDU sa posiela v Ethernet + LLC rámci
  - cieľová adresa (6B) – 01:80:C2:00:00:00
  - zdrojová adresa (6B) – MAC adresa portu
  - dĺžka (2B)
  - LLC header (3B) – 0x42, 0x42, 0x03
  - BPDU

Config BPDU

- protocol ID (2B): 0x0000
- protocol version (1B): 0x00
- BPDU type (1B): 0x00 = config
- flags (1B): bit 0 = TC, bit 7 = TC Ack
- root ID (8B)
- root path cost (4B)
- bridge ID (8B)
- port ID (2B)
- message age (2B)
- max age (2B)
- hello time (2B)
- forward delay (2B)

TCN BPDU

- protocol ID (2B): 0x0000
- protocol version (1B): 0x00
- BPDU type (1B): 0x80 = TCN

STP – konfigurácia

- nastavenie bridge priority
  - 4 (najvyššie) bity
    - 16 bitové číslo, ktoré musí byť násobkom 4096
    - default 0x8000 = 32768
  - 12 bitov – rozšírený ID (používa sa napr. pre VLAN)
  - ďalších 48 bitov je MAC adresa
- nastavenie port priority
  - 4 (najvyššie bity)
    - 8 bitové číslo, ktoré musí byť násobkom 16
    - default 0x80 = 128
  - ďalších 12 bitov je číslo portu

STP – konfigurácia

- hello time
  - ako často posiela root bridge BPDU
  - default 2s
- max age time
  - timeout pre životnosť konfiguračnej správy
  - default 20s
- forward delay
  - čas medzi prechodom z listening do learning a z learning do forwarding stavu
  - default 15s

STP – konfigurácia

- path cost
  - 10Mbps 100
  - 100Mbps 19
  - 1Gbps 4
  - 10Gbps 2

STP – algoritmus

- parametre bridge
  - designated root, root path cost, root port
  - max age, hello time, forward delay
  - bridge ID, bridge max age, bridge hello time, bridge forward delay
    - konfigurovateľné parametre, použité, keď som root bridge
  - topology change detected
    - flag, ktorý nastavím, keď zistím zmenu alebo dostanem TCN
  - topology change
    - flag, ktorý root bridge kopíruje do BPDU

STP – algoritmus

- bridge timers
  - hello timer
    - slúži na pravidelné posielanie BPDU z root bridge
  - topology change notification timer
    - slúži na opakované posielanie TCN, ku ktorým ešte neprišiel Ack
  - topology change timer
    - počas tohto času necháva root bridge nastavený TC flag
    - bridge max age + bridge forward delay

STP – algoritmus

- parametre portu
  - port ID, stav, path cost
  - designated root, designated cost, designated bridge, designated port
    - údaje z prijatej BPDU na porte (ak je to DP, tak vlastné)
  - topology change ack
    - TC Ack pre budúcu poslanú BPDU
  - configuration pending
    - flag, ktorý hovorí, že sa má poslať BPDU (po hold time)
  - change detection enabled

STP – algoritmus

- port timers
  - message age timer
    - sleduje vek prijatej BPDU
  - forward delay timer
    - pre listening a learning stavy
  - hold timer
    - obmedzenie frekvencie BPDU (max 1/s)

Prijatie BPDU

- ak prijatá BPDU obsahuje lepšie alebo rovnaké údaje ako zaznamenané v parametroch portu
  - zapíš údaje
    - zapíš údaje z BPDU do parametrov portu
    - nastav message age timer
  - aktualizuj konfiguráciu
    - vyber root
    - vyber designated porty
  - zmeň stavy portov
  - ak bola BPDU prijatá z (nového) RP, zaznamenaj timeouty do parametrov bridge a vygeneruj BPDU na DP
  - ak bola BPDU prijatá z RP a mala nastavený TC Ack, zastav Topology change notification timer a vynuluj Topology change detected flag

Prijatie BPDU

- ak je na designated porte prijatá BPDU s horšími parametrami
  - t.j. nejaký iný bridge sa snaží stať root bridge-om
  - pošli na tomto porte BPDU
- ak je prijatá TCN BPDU na designated porte
  - spusti Topology change detection
  - pošli potvrdenie
    - nastav Topology change ack
    - pošli BPDU na porte

Výber root bridge a root port

- nájdi port, ktorý nie je designated port a
  - má najmenší designated root
  - má najmenší designated cost + path cost
  - má najmenší designated bridge
  - má najmenší designated port
  - má najmenší port ID
- ak taký neexistuje (si root)
  - nastav designated root (bridge par.) na svoje ID, root path cost = 0, root port = 0
- ak existuje
  - bude to root port
  - nastav designated root (bridge par.) = designated root z root portu
  - root path cost = designated cost + path cost

Výber designated portov

- ako designated port nastav každý port, ak
  - už je designated port
  - designated root portu <> designated root bridge-u
  - root path cost < designated cost portu, alebo = a
  - bridge ID < designated bridge portu, alebo = a
  - port ID < designated port portu
- nastavenie portu ako designated port:
  - designated root portu = designated root bridge-u
  - designated cost = root path cost
  - designated bridge = bridge ID
  - designated port = port ID

Topology change detection

- vykonáva sa pri
  - prijatí TCN BPDU
  - prechode portu do forwarding stavu, ak je change detection enabled a bridge je designated bridge aspoň pre nejaký segment
  - prechode portu z forwarding alebo learning do blocking, ak je change detection enabled
  - keď sa bridge stane root bridge-om
- postup
  - ak si root bridge
    - nastav Topology change flag
    - spusti Topology change timer
  - ak nie si root bridge a ešte nie je nastavený Topology change detected
    - pošli TCN BPDU cez root port
    - spusti Topology change notification timer
  - nastav Topology change detected

Vypršanie Message age timer

- nastav port ako designated port
- aktualizuj konfiguráciu
  - vyber root
  - vyber designated porty
- zmeň stavy portov
- ak sa stávaš root bridge-om
  - max age, hello time, forward delay nastav podľa svojich
  - spusti Topology change detection
  - zastav Topology change notification timer
  - vygeneruj BPDU na všetky designated porty
  - spusti Hello timer

Poslanie BPDU

- ak je aktívny Hold timer, len nastav Config pending, inak
- zostav BPDU
  - root ID, root path cost, bridge ID, port ID z parametrov bridge-u
  - ak si root bridge, message age = 0 inak upravená message age z root portu
  - max age, hello time a forward delay z parametrov bridge-u
  - TC Ack podľa Topology change ack portu
  - TC podľa Topology change bridge-u
- ak je message age < max age
  - vynuluj Topology change ack a Config pending
  - odošli BPDU, spusti Hold timer

Rapid STP (RSTP)

- STP je pomalý
  - max age time (20s) na zistenie straty konektivity
  - forward delay (15s) x 2 na obnovenie konektivity
  - spolu 50s v zlom prípade
  - v dobrom prípade (okamžitá detekcia zlyhania priamo pripojenej linky) 30s
- cieľom RSTP je tento proces urýchliť
- spätná kompatibilia
  - ak je na porte prijatá STP (v. 0) BPDU, na tomto porte bude sa bude používať STP (v. 0)

RSTP – typy portov

- root port
  - aktívny port smerom k root bridge-u
- designated port
  - aktívny port do segmentu siete na designated bridge-i
- backup port
  - neaktívny port do segmentu na designated bridge-i
  - záloha pre designated port
- alternate port
  - neaktívny port do segmentu na inom ako designated bridge-i
  - záloha pre root port

RSTP – typy portov

- edge port
  - port, ktorý je pripojený k LAN segmentu, ku ktorému nie je pripojený žiadny iný bridge
    - pripojený priamo ku koncovému zariadeniu
    - pripojený k časti siete bez podpory STP, ktorá neobsahuje žiadne redundantné spoje
  - nastaviteľná vlastnosť portu
  - umožňuje rýchly prechod do forwarding stavu
    - bez čakania, keďže žiadnu slučku nemôže vytvoriť

RSTP – stavy portov

- discarding
  - zjednotenie blocking a listening
  - neučí sa MAC, nepreposiela dáta
- learning
  - učí sa MAC
  - nepreposiela dáta
- forwarding
  - učí sa MAC
  - preposiela dáta

RSTP BPDU

- protocol ID (2B): 0x0000
- protocol version (1B): 0x02
- BPDU type (1B): 0x02 = RSTP BPDU
- flags (1B): b.0 = TC, b.1 = Prop, b.2-3: Port role, b.4 = Lrn, b.5 = Fwd, b.6 = Agr, b.7 = 0
  - Port role: 0=Unknown, 1=Alternate/Backup, 2=Root, 3=Designated
- root ID (8B)
- root path cost (4B)
- bridge ID (8B)
- port ID (2B)
- message age (2B)
- max age (2B)
- hello time (2B)
- forward delay (2B)
- version 1 length (1B): 0

RSTP – zrýchlenie

- konfiguračná BPDU sa posiela z každého designated portu pravidelne
  - v STP len ako reakcia na prijaté BPDU z root portu
- rýchly prechod do forwarding stavu
  - keď je isté, že aktiváciou portu nevznikne slučka
  - mechanizmus požiadaviek a potvrdení namiesto dlhého čakania
- krátky čas (3x Hello time) výpadku prichádzajúcich BPDU, kým začne konať
  - v STP default 10x Hello time (20s)
- krátky forwarding delay (Hello time)
- čas rekonfigurácie môže byť rádovo v ms až pár sekúnd
  - STP 30 – 50s

RSTP – priority vector

- (root, root path cost, desig. bridge, desig. port, port)
- message PV
  - hodnoty prijaté z BPDU
- port PV
  - uložené hodnoty z message PV na porte
- root path PV
  - vypočítané pre každý port, ktorý má port PV získaný z prijatej správy – root path cost sa zvýši o port path cost
- bridge PV
  - (B, 0, B, 0, 0) – ak by bol B root, z tohto by odvodzoval vysielané PV
- root PV
  - buď bridge PV alebo najmenší root path PV
- designated PV
  - pre každý port odvodený od root PV tak, že za designated port dosadím príslušný port ID
  - toto je PV, ktorý budem posielať na porte, ak bude desig. portom pre segment

RSTP – voľba roly portov

- ak je port PV neaktuálny, port je **desig. port**
- ak je port PV prijatý a aktuálny a je z neho odvodený root PV, tak port je **root port**
- ak je port PV prijatý a aktuálny, nie je z neho odvodený root PV a desig. PV portu nie je lepší a port PV.desig. bridge je iný bridge, port je **alternate port**
- ak je port PV prijatý a aktuálny, nie je z neho odvodený root PV a desig. PV portu nie je lepší a port PV.desig. bridge je tento bridge, port je **backup port**
- ak je port PV prijatý a aktuálny, nie je z neho odvodený root PV a desig. PV portu je lepší, port je **desig. port**
- ak port PV je vlastný, port je **desig. port**

RSTP – voľba roly portov

- Keď sa port má stať desig. portom
  - do port PV sa prenesú údaje z desig. PV

RSTP – zmena stavu portu

- alternate a backup port
  - prejde do discarding stavu
- root port
  - ak bol forwarding, zostáva
  - ak nebol, prepne do discarding všetky nedávno bývalé root porty a prejde do learning a následne forwarding
    - ak bol nedávno backup, tak počká aj na expiráciu 2xHelloTime

RSTP – zmena stavu portu

- designated port
  - oznámi svoj úmysel poslaním správy s nastaveným Proposal flagom susedovi
    - predpokladom je, že má na porte len 1 suseda
  - keď mu sused potvrdí, že je to OK (t.j. príjme zo susedovho root portu správu s Agreement flagom), tak prejde do learning a následne do forwarding
    - resp. po vypršaní času pre prípad STP
  - v prípade edge portu môže prejsť hneď

RSTP – Proposal & Agreement

- keď dostanem Proposal
  - t.j. sused chce aktivovať designated port
  - spustím synchronizáciu ostatných portov a následne pošlem Agreement naspäť susedovi
    - port je synchronizovaný, ak je discarding alebo edge alebo som na ňom prijal Agreement od suseda
    - ak port nie je synchronizovaný a nie je discarding, tak prejde do discarding (čím sa stane synchronizovaný) a následne cez neho pošlem Proposal susedovi
    - synchronizovaný port = istota, že port (a porty ďalších bridge-ov za ním) sú v súlade s rolami portov, teda že nemôže vzniknúť cyklus

RSTP – zmena topológie

- deaktivácia portu
  - vymaže naučené MAC adresy na tomto porte
- aktivácia portu
  - vymaže naučené MAC adresy na ostatných portoch a pošle TCN na všetky aktívne porty (desig. a root)
- prijatie TCN
  - vymaže naučené MAC adresy na ostatných portoch a pošle cez ne TCN
- ignoruje zmenu stavu edge portu, nevymazáva MAC adresy naučené na edge porte

RSTP – prijatie BPDU

- ak príjmem BPDU z desig. portu
  - s lepším message PV ako port PV alebo z rovnakého desig. bridge a portu ako má port PV, tak aktualizujem port PV a zbehnem nové stanovenie rol
  - s rovnakým message PV, tak len aktualizujem timer
  - s horším message PV a s Learning z iného desig. portu na mojom desig. porte, tak začnem nový pokus stať sa desig. portom – zjavne tu niečo nesedí a iný bridge sa pokúša nastaviť svoj port ako designated pre tento segment
- ak príjmem BPDU z root / alternate / backup portu
  - tak spracujem prípadný Agreement (čo mi umožní prejsť s Designated portom do Learning / Forwarding stavu bez ďašieho čakania)

STP a VLAN

- klasický STP a RSTP nerieši VLAN
- Per-VLAN-Spanning-Tree (PVST+)
  - CISCO špecialita
  - pre každú VLAN vytvára samostatnú kostru
- Multiple Spanning Tree Protocol (MSTP)
  - ďalšia štandardizovaná verzia STP
  - vychádza z RSTP
  - umožňuje existenciu viacerých kostier
    - a teda efektívnejšie využitie liniek viacerými VLAN

MSTP

- jedna globálna kostra – Common and Internal Spanning Tree (CIST)
  - skladá sa z Common Spanning Tree (CST)
    - prepája STP a RSTP bridge a MST regióny
  - a z Internal Spanning Tree (IST) v každom MST regióne
- MST región
  - MST bridge-e (a príslušné segmenty siete) so spoločnou konfiguráciou – mapovaním VLAN na inštancie MSTI
- MSTI (MST Instance) – samostatná kostra vnútri MST regiónu pre nejakú podmnožinu VLAN

MSTP

- CIST root
  - globálny root bridge
- CIST regional root
  - root IST v regióne
  - pripája región k CST
- MSTI regional root
  - root pre MST inštanciu v regióne

MSTP roly portov

- CIST roly
  - root, designated, alternate, backup – ako v RSTP
- MSTI roly
  - root, designated, alternate, backup
  - master port
    - CIST root port CIST regional root bridge-u
    - zabezpečuje prepojenie MSTI a siete mimo MST regiónu

MSTP

- komunikácia vnútri regiónu
  - použijú sa linky v rámci príslušnej MSTI alebo IST vnútri regiónu
- komunikácia z regiónu von (alebo naopak)
  - použijú sa linky v rámci MSTI alebo IST smerom
    - k CIST regional root bridge-u a cez master port von z regiónu, alebo
    - k listom a odtiaľ do iného regiónu

MSTP

- CIST priority vector
  - CIST root ID, external root path cost,
  - CIST regional root ID, internal root path cost,
  - Designated bridge ID, Designated port ID, Receiving port ID
- MSTI priority vector
  - MSTI regional root ID, internal root path cost,
  - Designated bridge ID, Designated port ID, Receiving port ID

MSTP

- priorita bridge-u, priorita portu
  - nastavovateľné separátne pre CIST a každú MSTI
- port path cost
  - zvlášť external port path cost
    - používané pre kalkuláciu ceny v rámci CST
  - zvlášť internal port path cost pre IST
  - zvlášť internal port path cost pre každú MSTI
- cena externej cesty je v rámci regiónu rovnaká

MSTP – priradenie rol portom

- určia sa CIST roly
  - identifikuje sa CIST root bridge
  - určí sa CIST root port
  - určia sa CIST designated porty
  - určia sa CIST alternate a backup porty
- určia sa MSTI roly
  - určí sa master port
  - identifikuje sa MSTI root bridge
  - určí sa MSTI root port
  - určia sa MSTI designated porty
  - určia sa MSTI alternate a backup porty

MSTP BPDU

- protocol ID (2B): 0x0000
- protocol version (1B): 0x03
- BPDU type (1B): 0x02 = RSTP/MSTP BPDU
- flags (1B): b.0 = TC, b.1 = Prop, b.2-3: Port role, b.4 = Lrn, b.5 = Fwd, b.6 = Agr, b.7 = 0
  - Port role: 0=Unknown, 1=Alternate/Backup, 2=Root, 3=Designated
- CIST root ID (8B)
- external root path cost (4B)
- CIST regional root ID (8B)
- port ID (2B)
- message age (2B)
- max age (2B)
- hello time (2B)
- forward delay (2B)
- version 1 length (1B): 0
- ...

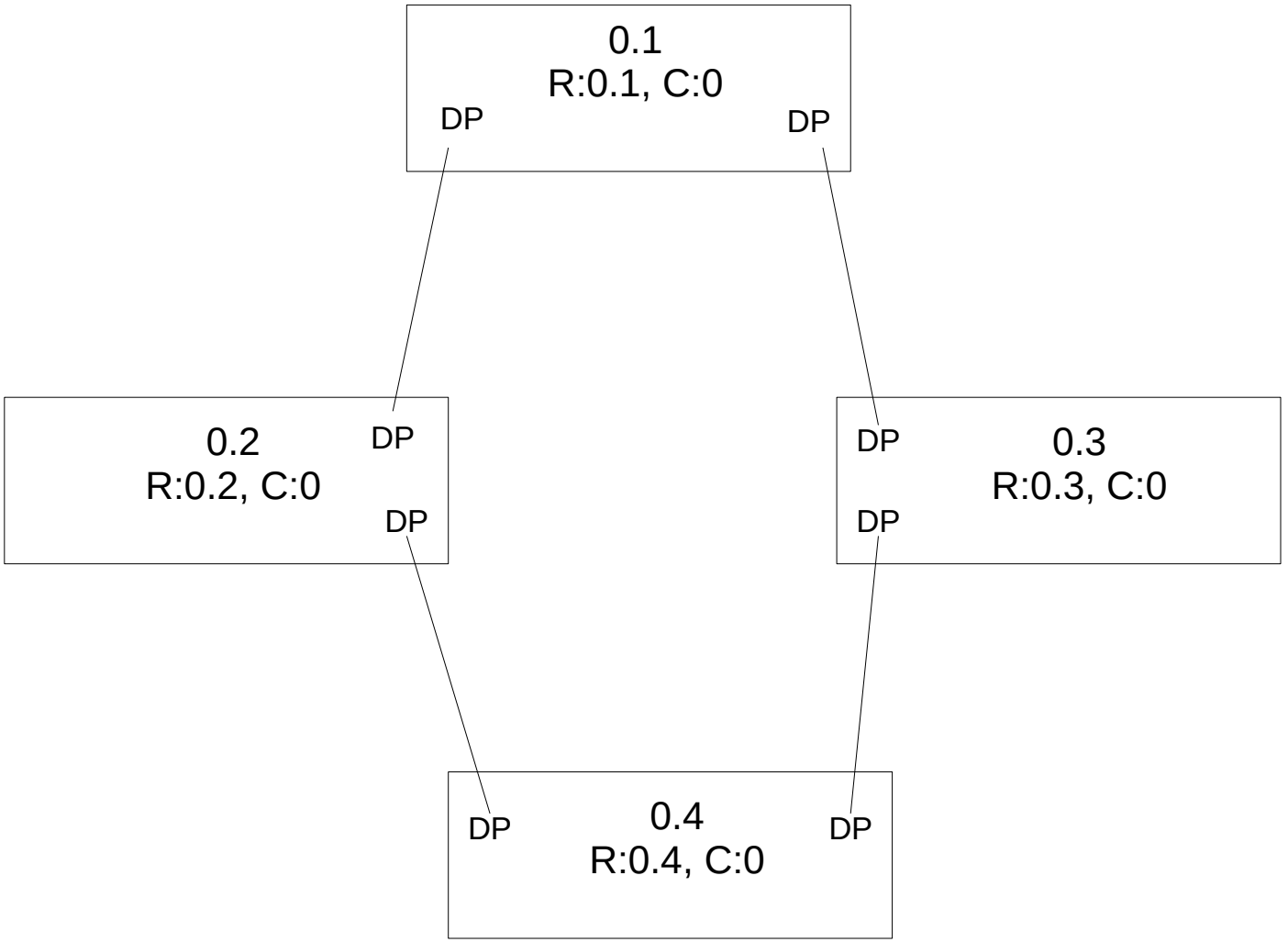
MSTP BPDU

- Version 3 length (2B)
- MST Configuration ID (51B)
- CIST internal root path cost (4B)
- CIST bridge ID (8B)
- CIST remaining hops (1B)
- MSTI configuration messages (0 – 64 x)
  - MSTI flags (1B) – (pribúda port role 0 = master)
  - MSTI regional root ID (8B) – obsahuje MSTID v 12 bitoch rozšíreného system ID
  - MSTI internal root path cost (4B)
  - MSTI bridge priority (1B)
  - MSTI port priority (1B)
  - MSTI remaining hops (1B)

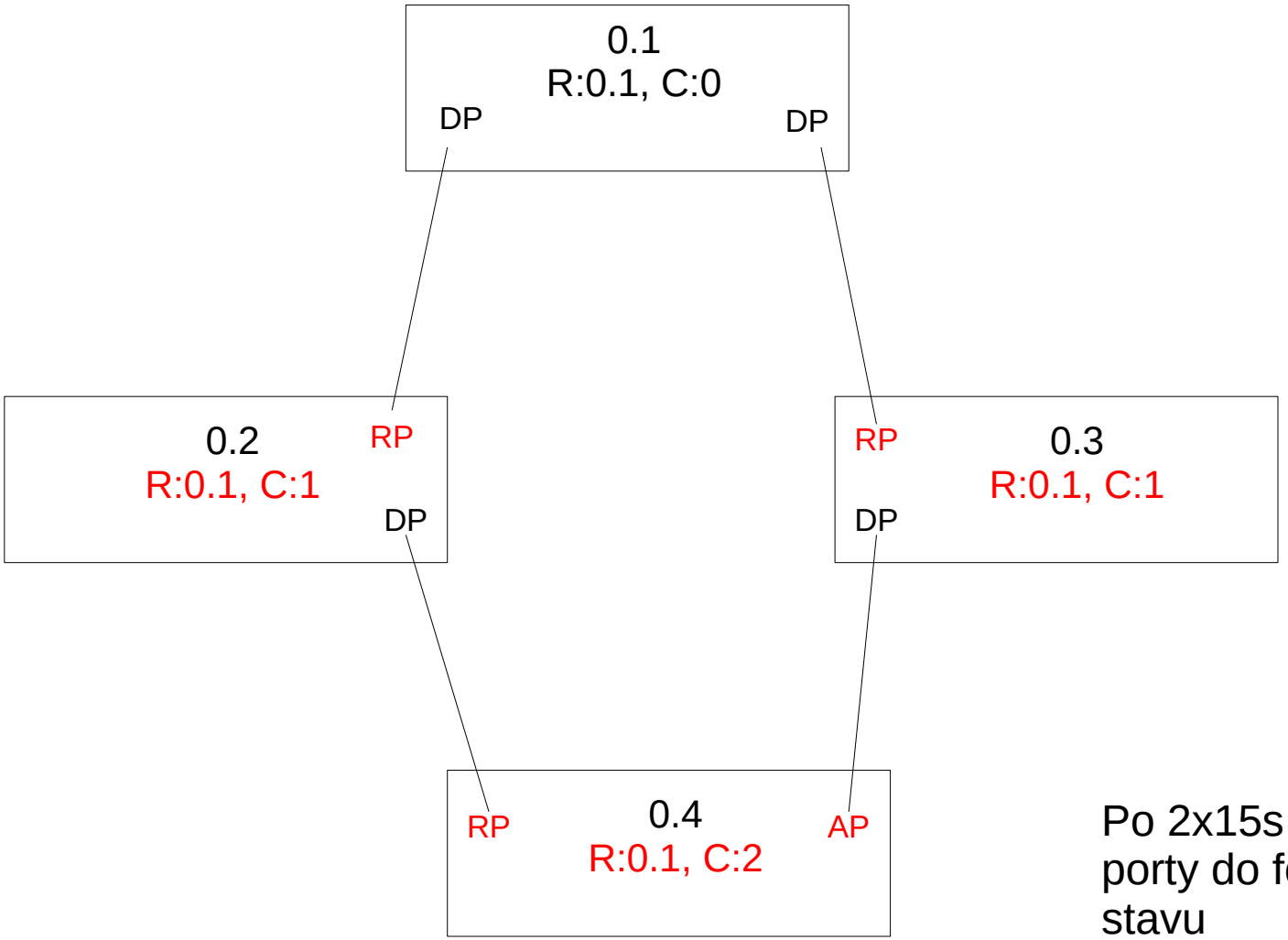
MSTP Configuration ID

- slúži na identifikáciu bridge-ov so spoločným priradením VLAN k MSTID
  - Format selector (1B) = 0
  - Config name (32B) – default MAC v hex
  - Revision level (2B) – default 0
  - Config digest (16B)
    - HMAC-MD5 z tabuľky 4096 x 2B
      - MSTID pre každú VLAN (0 = IST, prvá a posledná položka = 0)
      - kľúč: 0x13AC06A62E47FD51F95D2BA243CD0346

STP – príklad

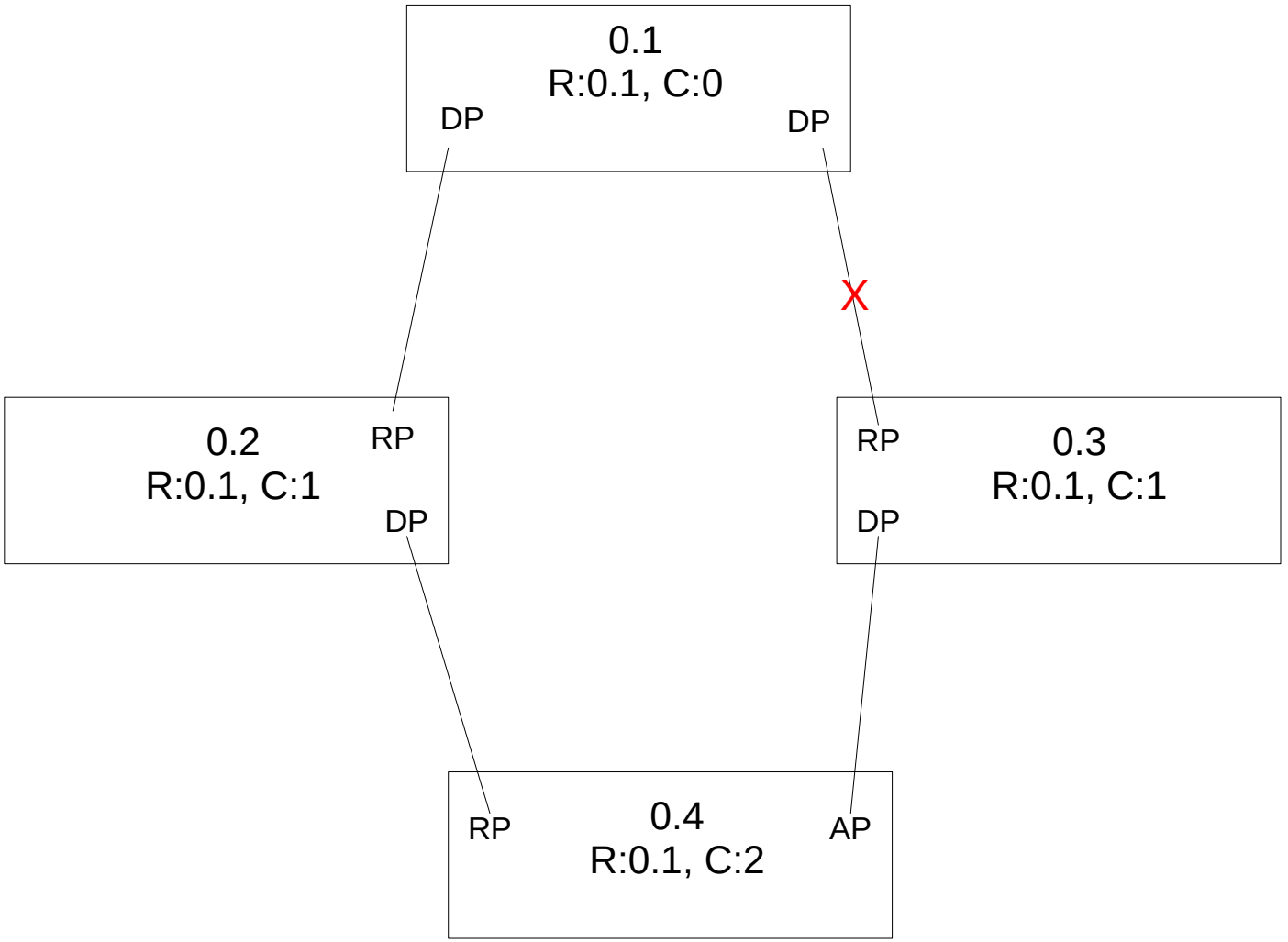


STP – príklad

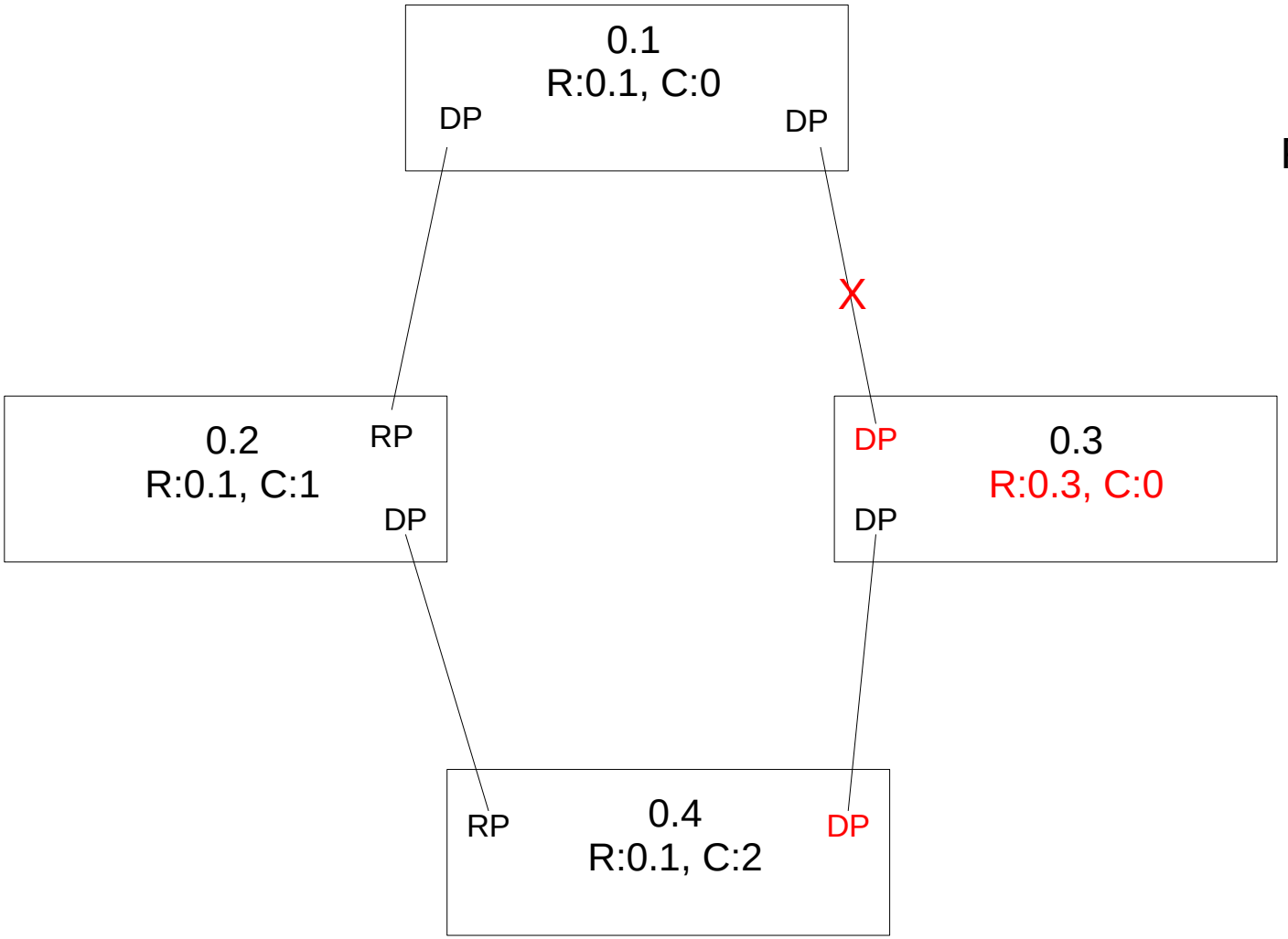


Po 2x15s prejdú porty do forwarding stavu

STP – príklad

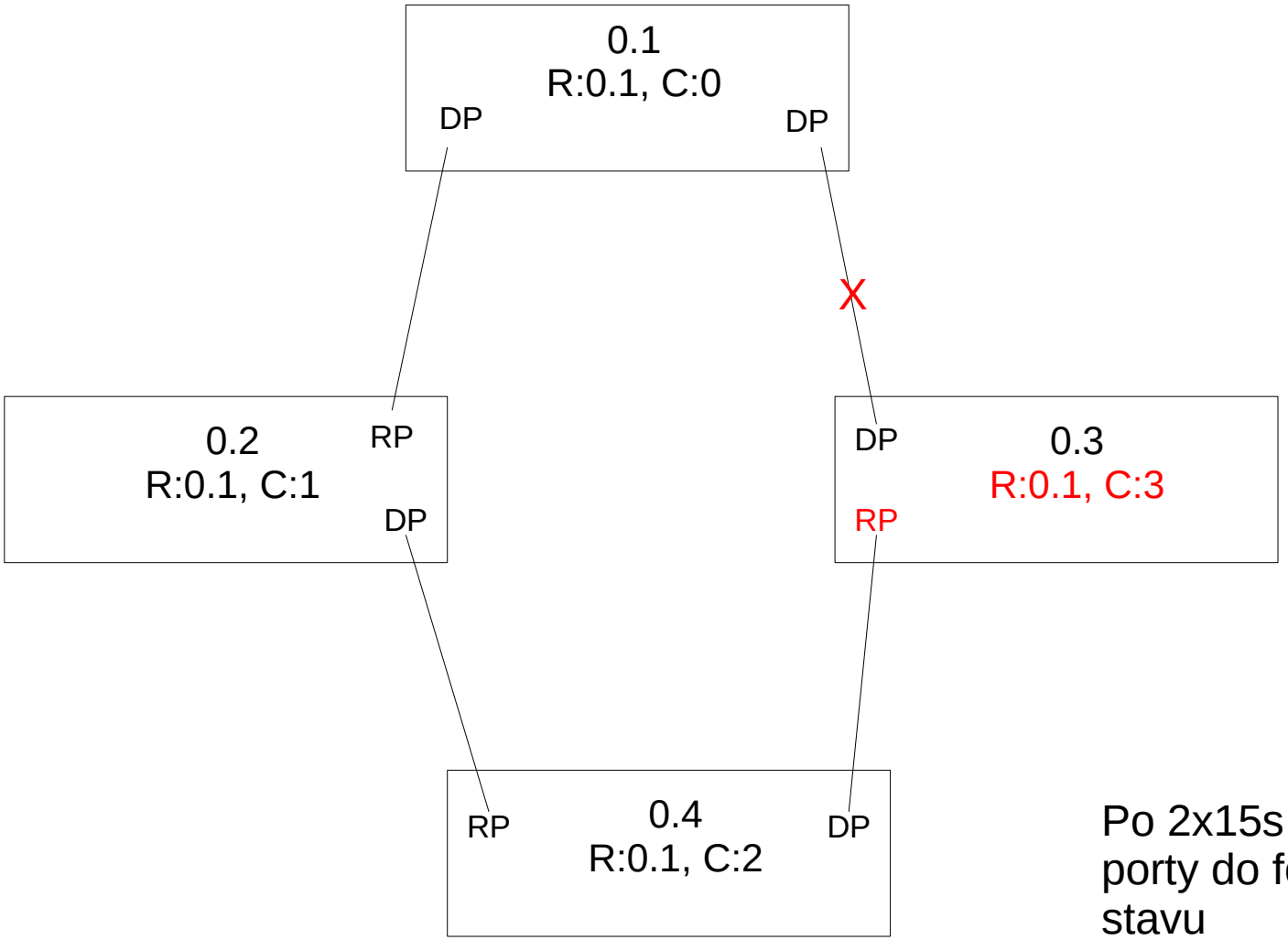


STP – príklad



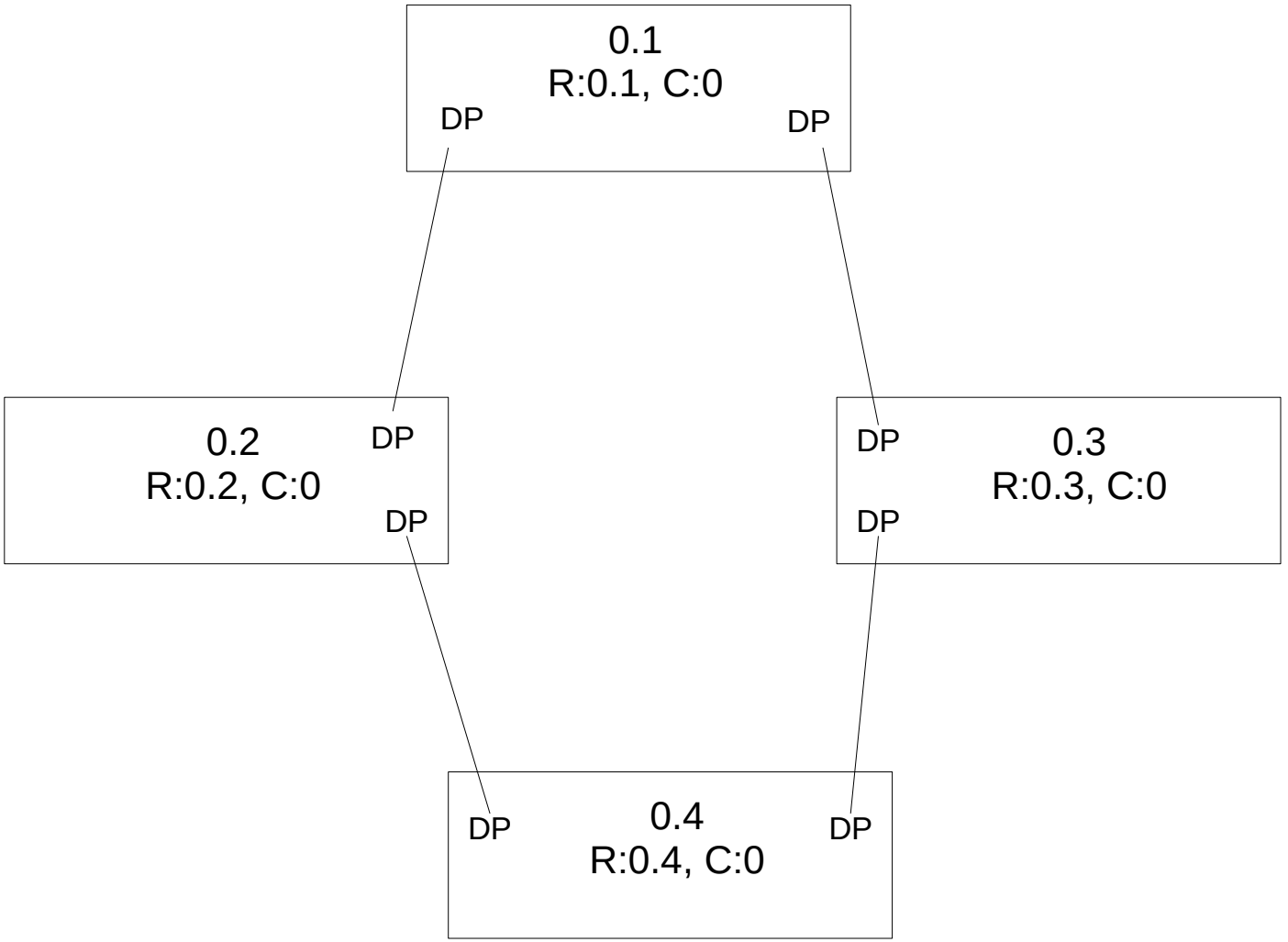
Po 20s

STP – príklad

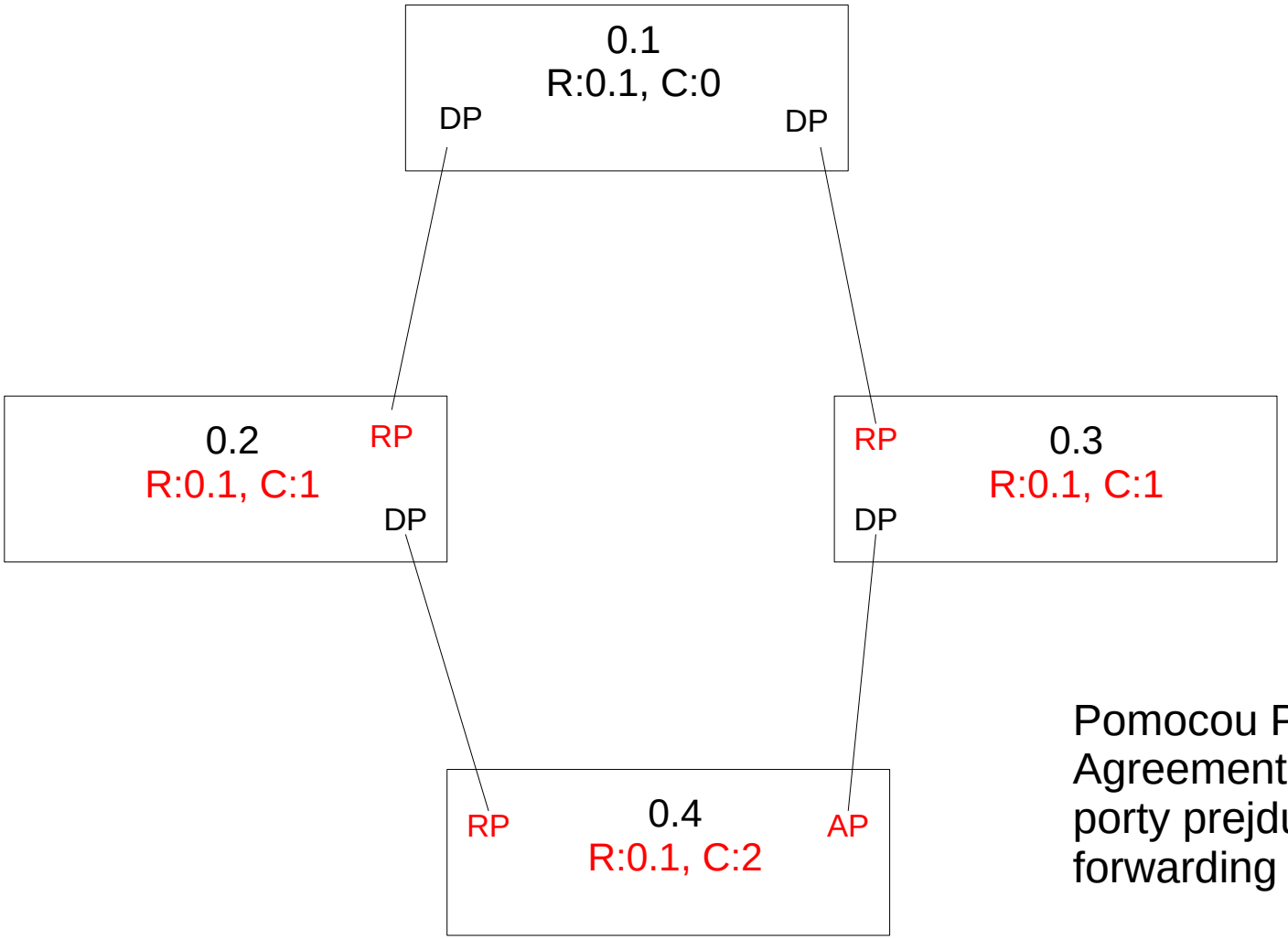


Po 2x15s prejdú porty do forwarding stavu

RSTP – príklad

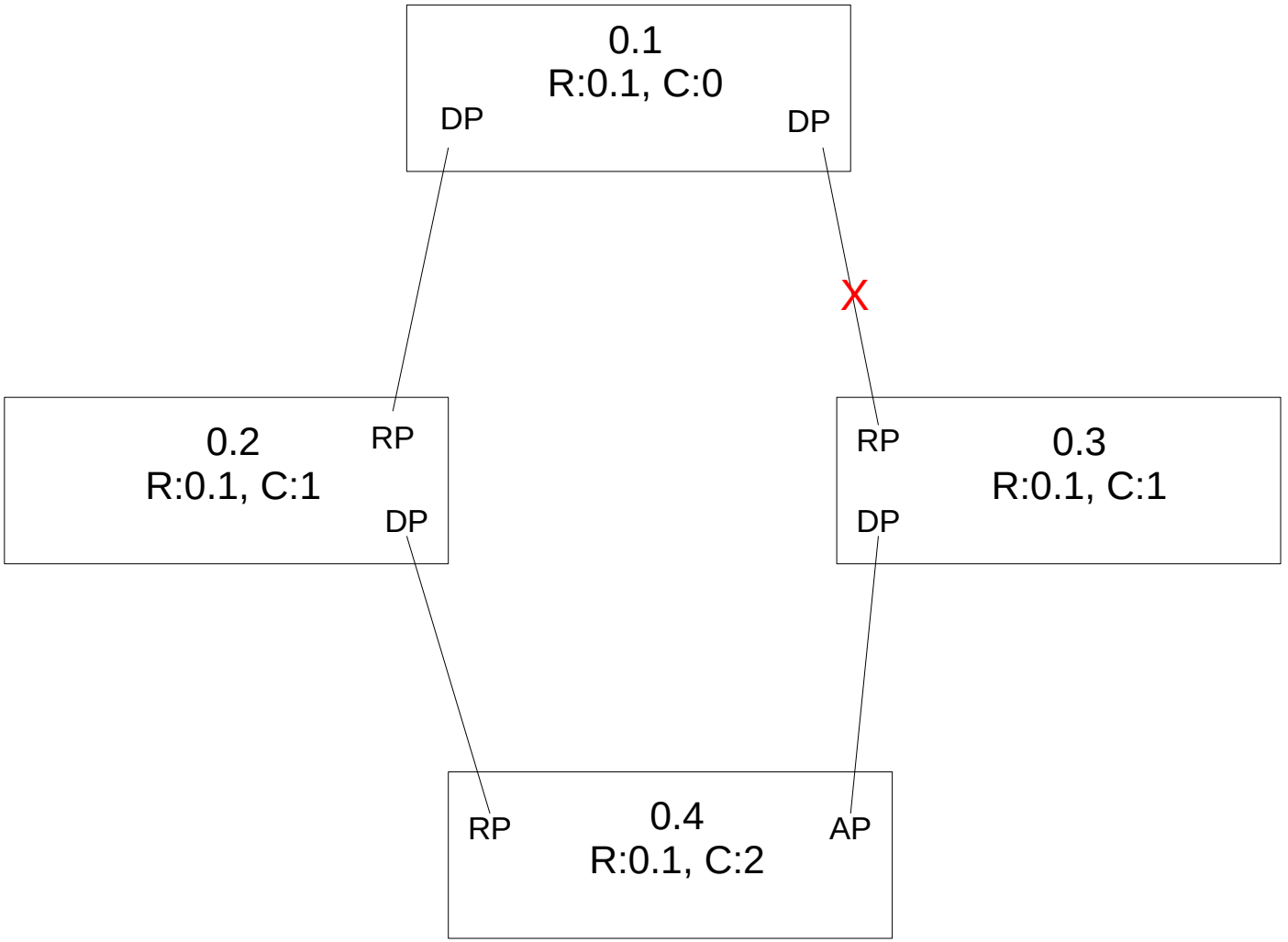


RSTP – príklad

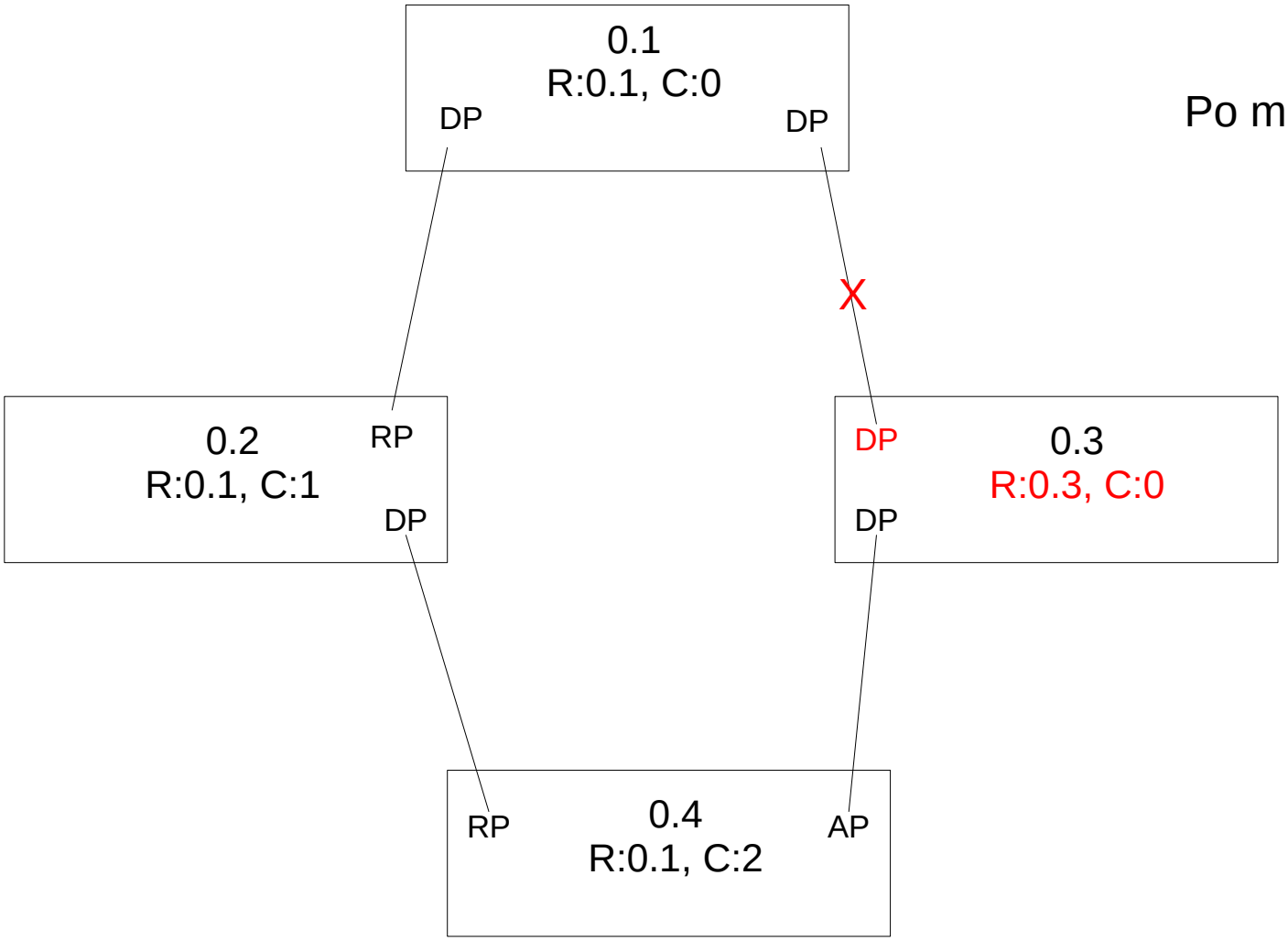


Pomocou Proposal – Agreement mechanizmu porty prejdú rýchlo do forwarding stavu

RSTP – príklad

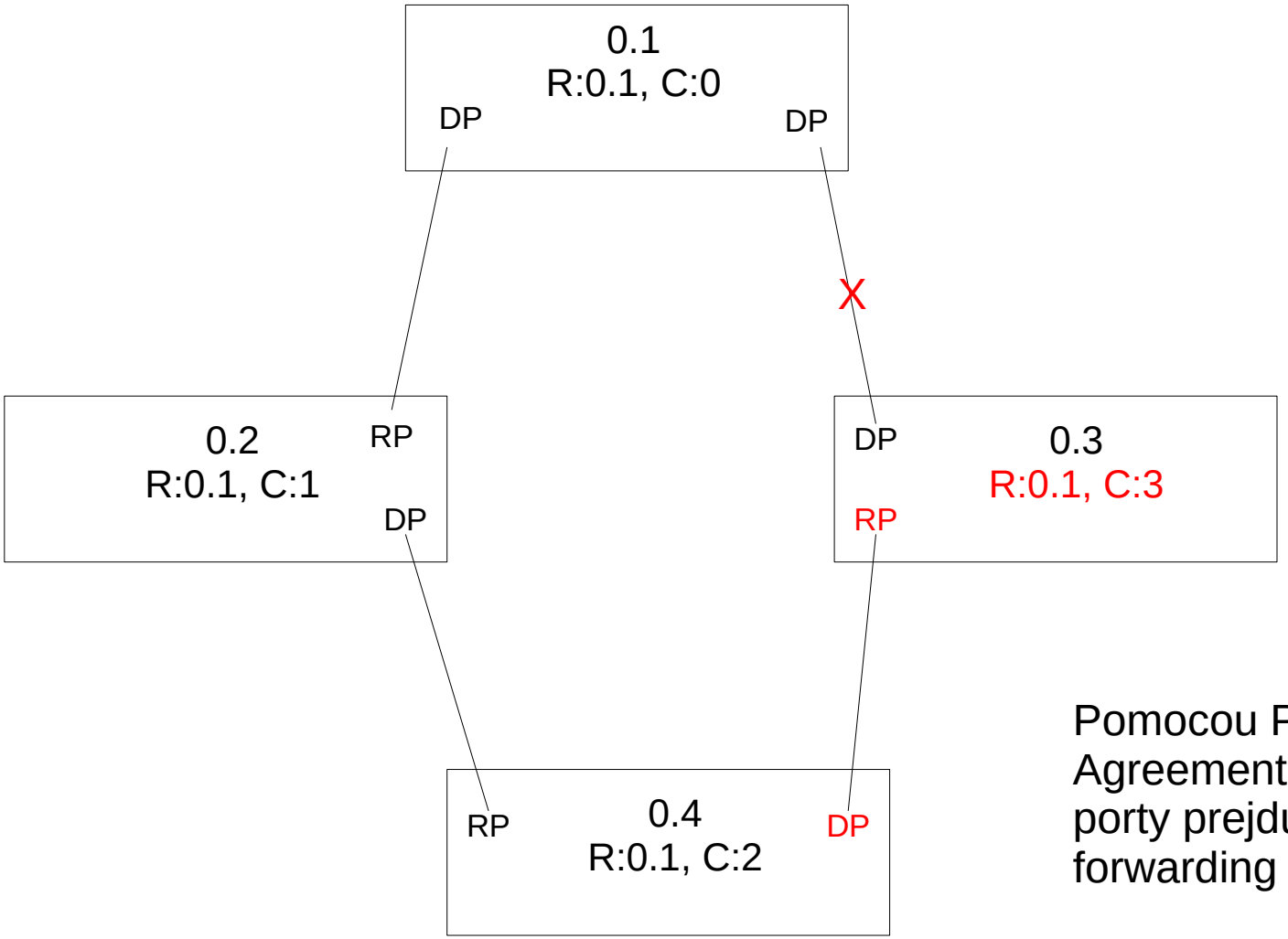


RSTP – príklad



Po max. 6s

RSTP – príklad



Pomocou Proposal – Agreement mechanizmu porty prejdú rýchlo do forwarding stavu