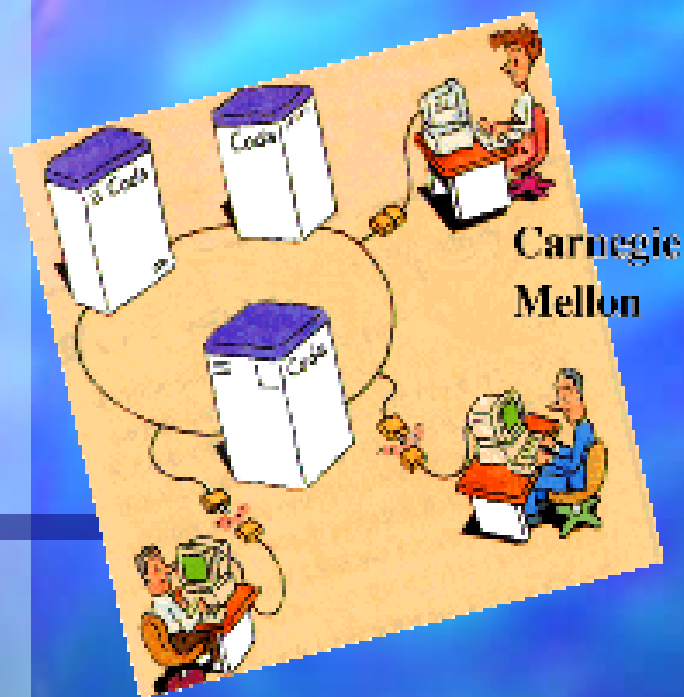




Rozproszony system plików Coda

Krzysztof Lichota

AFS

System plików dla środowiska akademickiego:

- częstsze odczyty niż zapisy
- rzadkie współdzielenie pliku do zapisu
- powtarzalność dostępu do pliku
- sekwencyjny odczyt

Cele AFS

- Skalowalność
- Semantyka zbliżona do one-copy
- Sensowna wydajność

Charakterystyka AFS

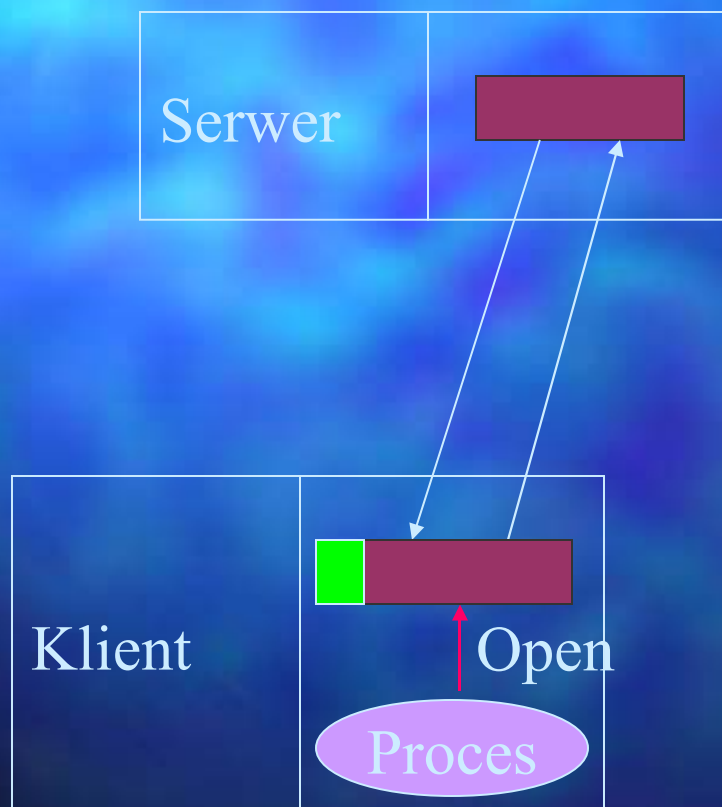
- Trwały cache u klienta
- whole-file caching
- whole-file serving

Replikacja read-only

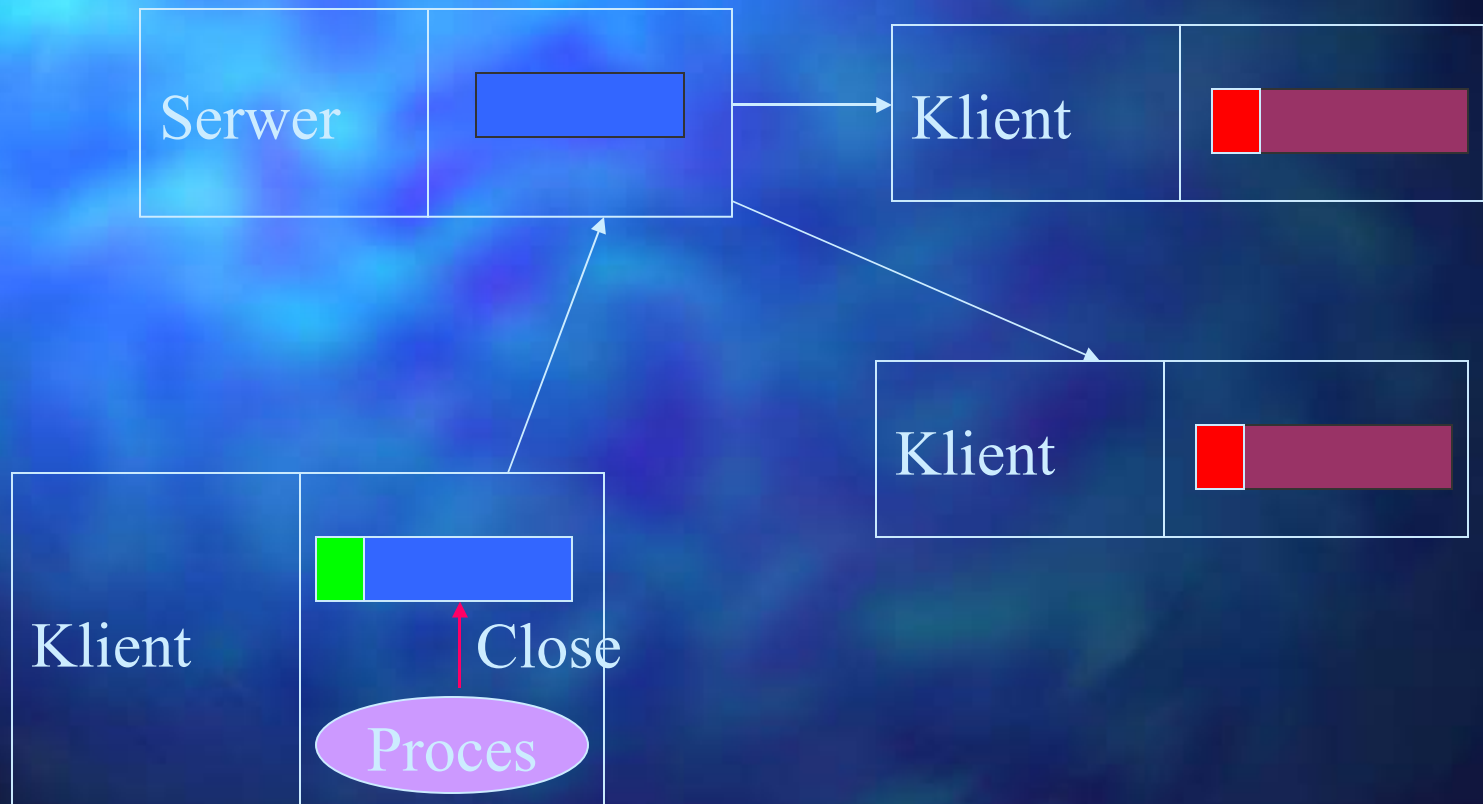
Semantyka

- Sprawdzenie aktualności tylko przy otwarciu pliku
- Dalsze operacje lokalne
- Zamknięcie pliku zamazuje poprzednią zawartość na serwerze
- Spójność cache utrzymywana mechanizmem „callback”

Sprowadzenie do cache



Unieważnienie kopii



Cechy callbacków

- Nie musimy kontaktować się z serwerem przy każdym otwarciu pliku
- Nie zawsze mamy najbardziej aktualną kopię (zgubione callbacki)

Zgubione callbacki wykrywamy
przepytując okresowo serwery

Wady AFS

Replikacja read-only

- Wrażliwość na awarie serwerów
- Wrażliwość na awarie sieci

Niemożliwe działanie bez połączenia do serwera

- Wrażliwość na krótkie awarie
- Brak wsparcia dla komputerów przenośnych

Założenia Coda

Skalowalność

Semantyka podobna do one-copy

Odporność na awarie:

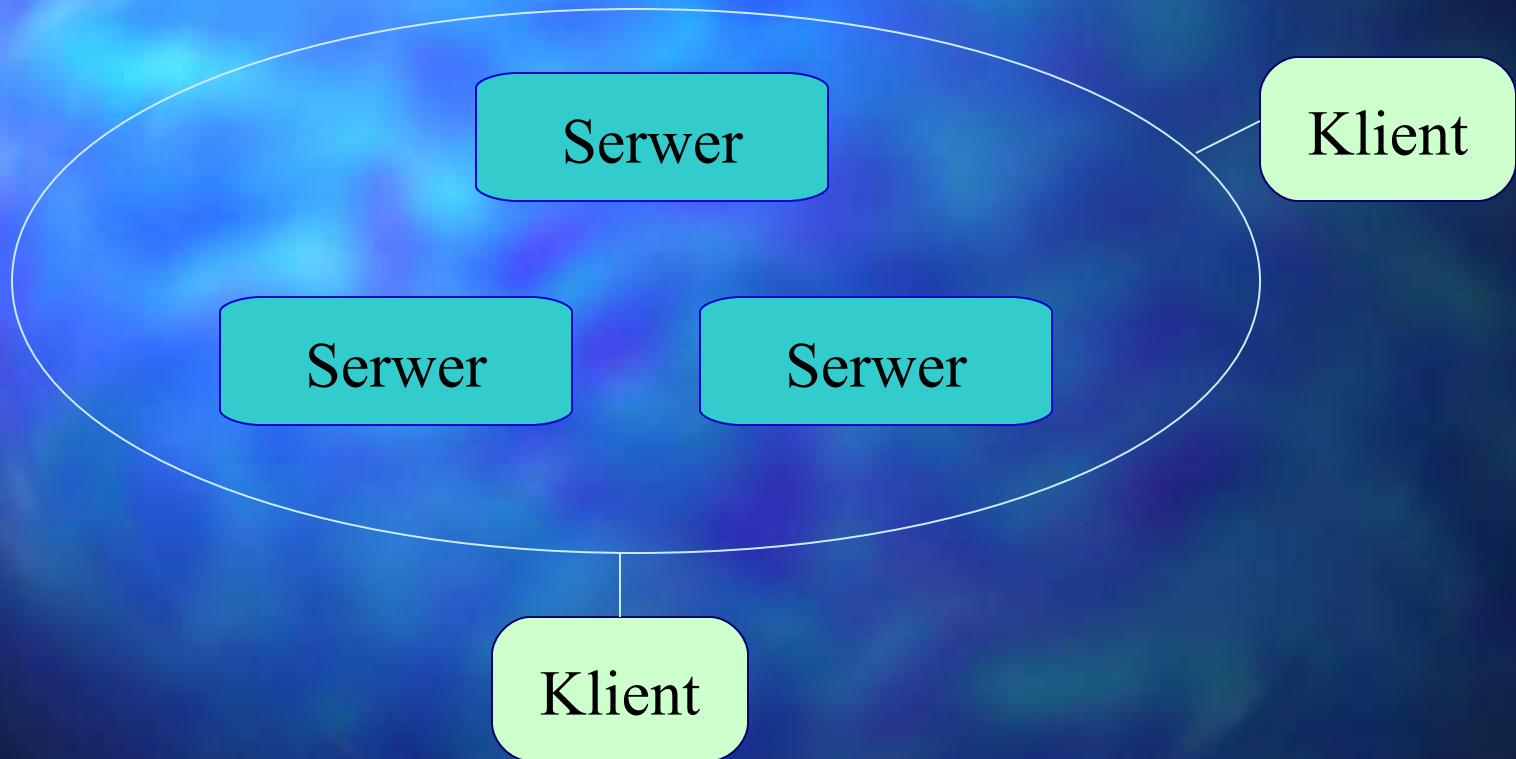
- replikacja read-write
- korzystanie z cache gdy nie ma połączenia

} AFS

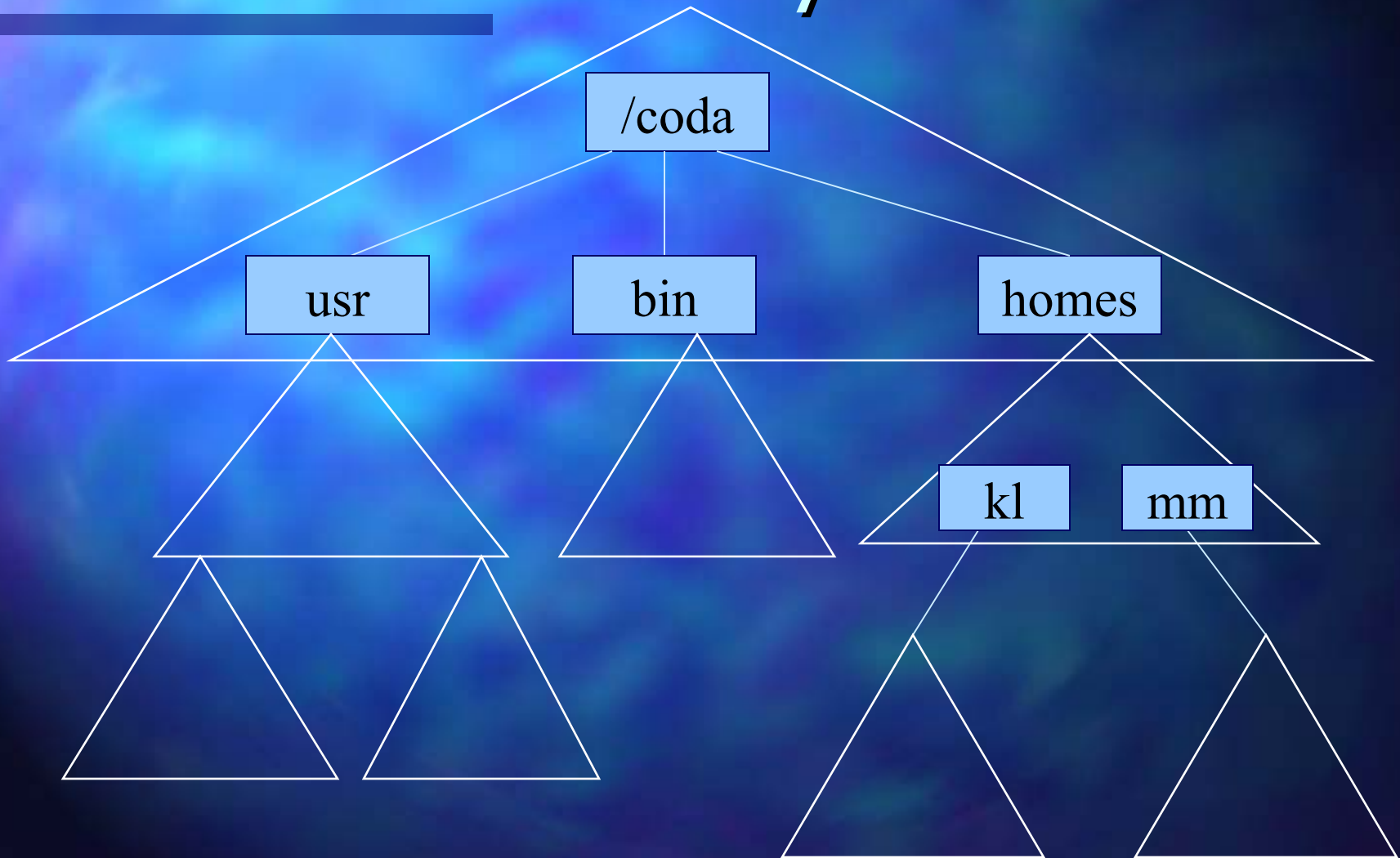
Optymistyczna replikacja

- Jedyna możliwość dla komputerów przenośnych
- Brak ograniczeń przy rozdzielaniu serwerów
- Konflikty występują, ale są rzadkie

Coda cell



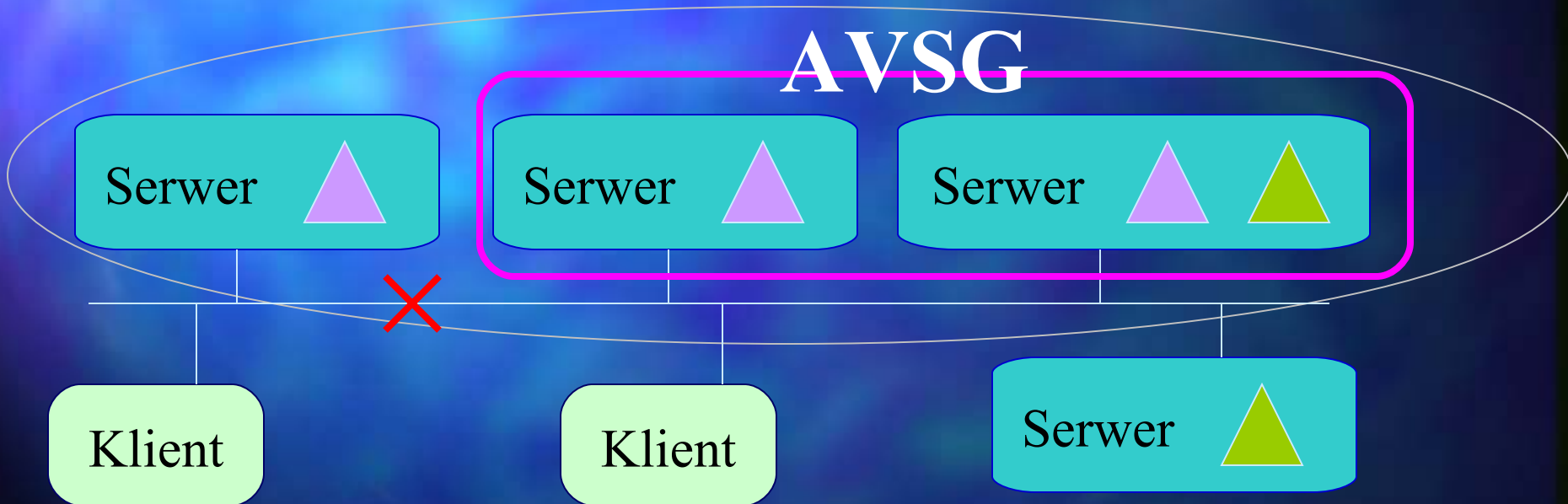
Woluminy



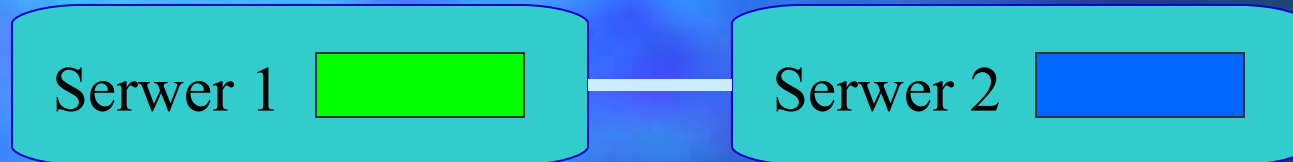
AVSG i VSG

VSG

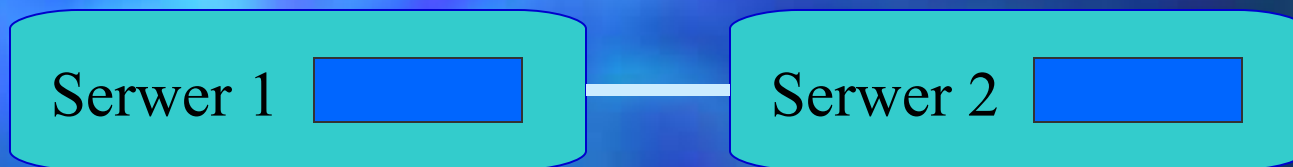
AVSG



Konflikt replik

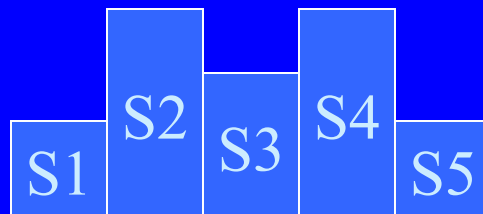


Dominacja repliki



CVV

LSID



CVV

VSG

Równość replik

Mocna równość

- LSID takie same
- CVV takie same

Słaba równość

- LSID się zgadzają
- CVV różne

Wykrywanie konfliktu



Dominacja

S1

S1	S2	S3
----	----	----

S2

S1	S2	S3
----	----	----

S3

S1	S2	S3
----	----	----

S1	S2	S3

S1	S2	S3

S1	S2	S3

S1	S2	S3

Rozsyłanie informacji

- Read-one, write-all
- „Wybrany serwer”
- MultiRPC

Semantyka operacji open

Sukces:

Mamy kopię nie starszą niż T sekund i spójną w danym AVSG

- sprowadziliśmy aktualną kopię
 - nie mieliśmy jej w cache
 - nasza kopia została unieważniona
- skorzystaliśmy z kopii w cache
 - kopia jest aktualna lub
 - zgubiliśmy callback

Semantyka operacji open

Porażka:

- kopie w AVSG są w konflikcie

Semantyka operacji close

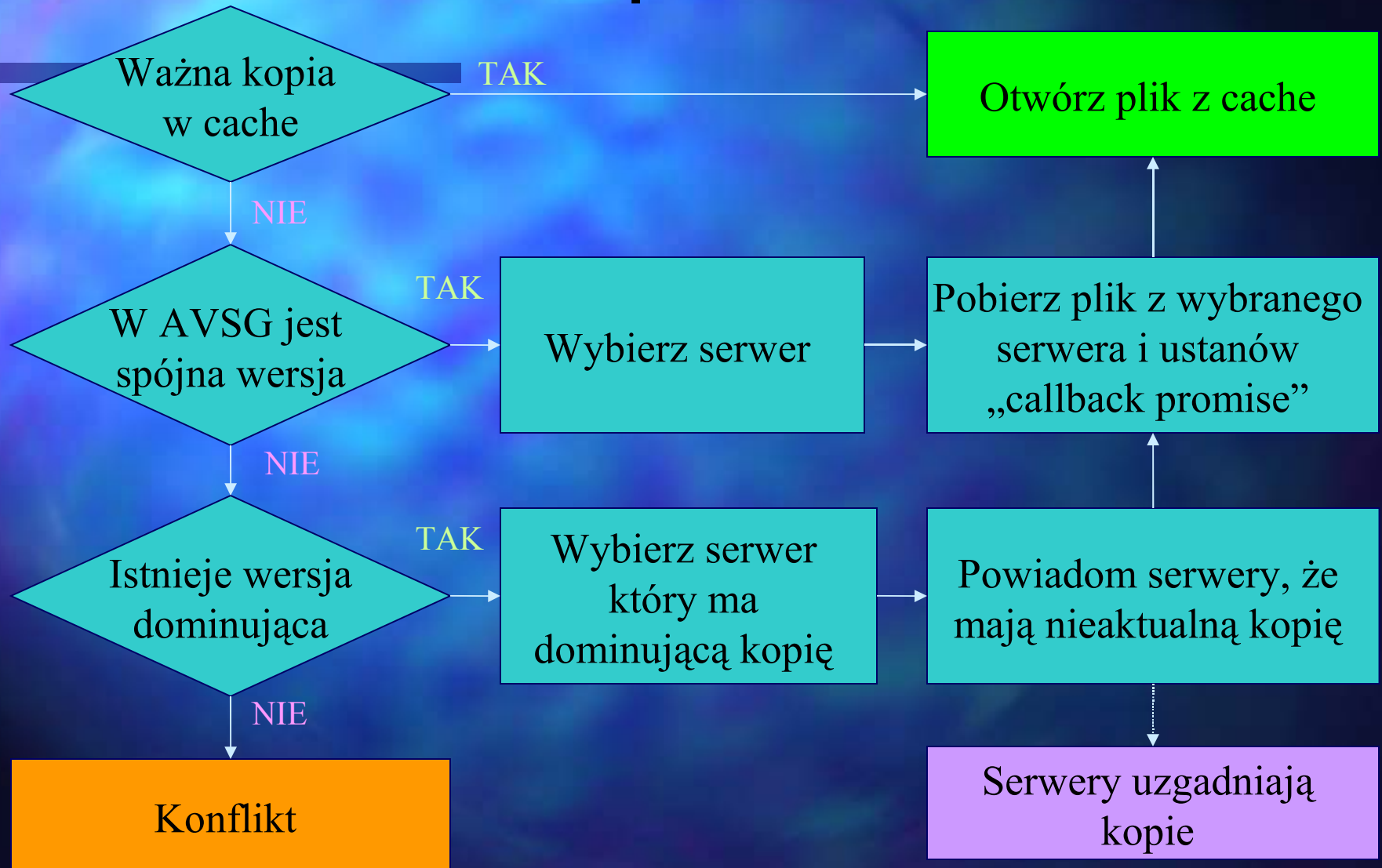
Sukces:

- wszystkie serwery z AVSG uaktualniły swoją kopię

Porażka:

- kopie w AVSG są w konflikcie

Open



Close

- Sprawdzenie przez serwery, czy kopia dominuje nad ich kopiami lub jest równa
- Zatwierdzenie nowego LSID i tymczasowego CVV na serwerze
- Rozesłanie ostatecznego CVV do serwerów

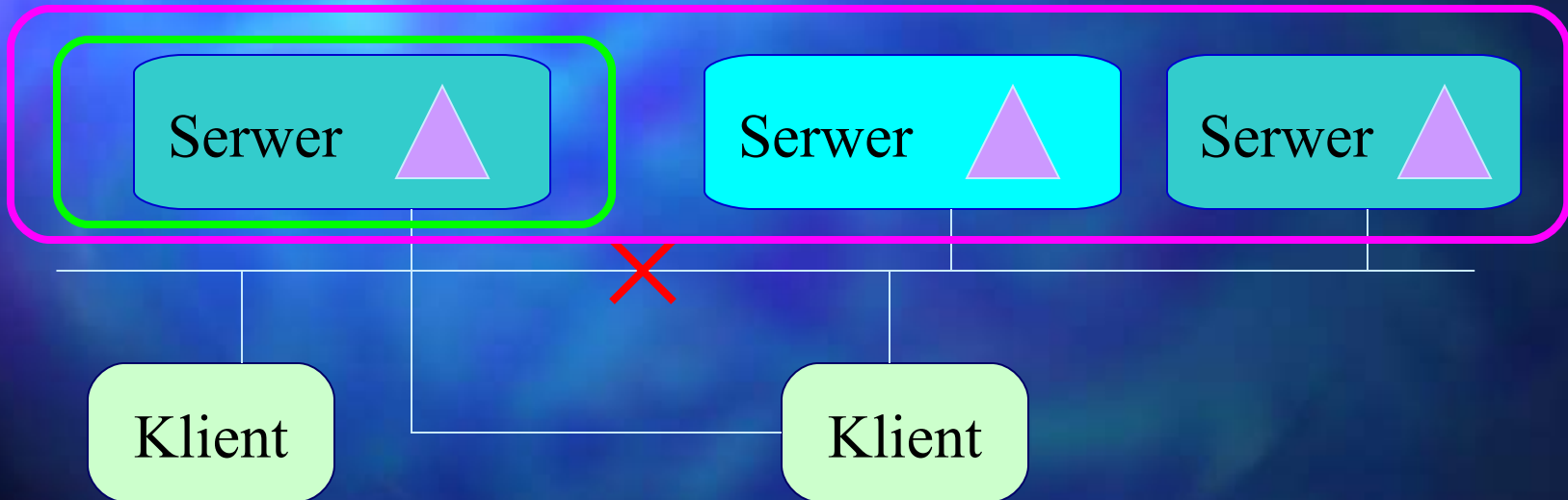
Probing

Co T sekund próbujemy skontaktować się z każdym serwerem z VSG

- odpowiada nowy serwer
- nie zgłasza się poprzednio dostępny serwer

Callback z wybranego serwera

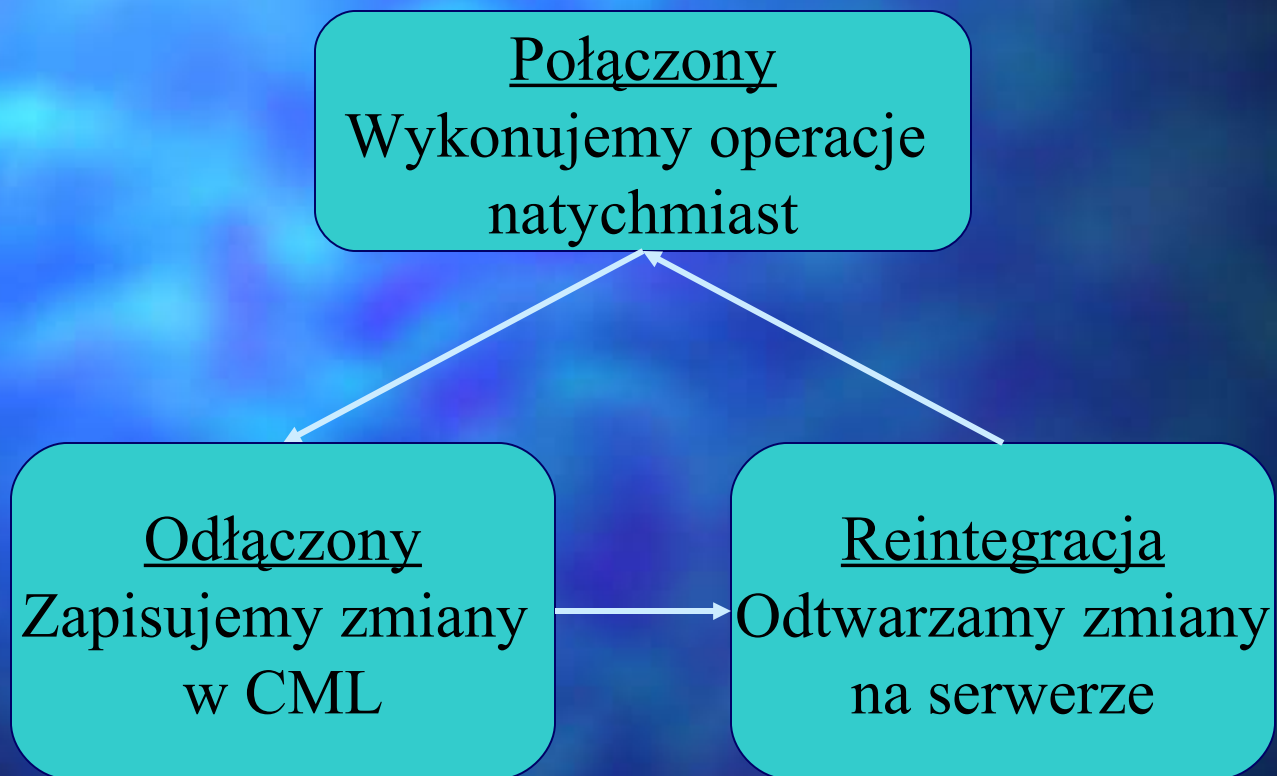
Callback promise ustanawiany jest tylko na serwerze, z którego pobieramy plik



Volume CVW

- Zawiera sumę operacji na danym woluminie
- Niezgodność powoduje unieważnienie wszystkich plików z danego woluminu

Disconnected operation



Hoarding

Gromadzimy potrzebne pliki przed odłączeniem

- priorytetowa lista plików
- pliki „sticky”
- spy

Reintegracja

- Odtwarzamy zmiany na serwerze
- Rozstrzygamy konflikty jeśli się da
- Jeśli nie przechowujemy pliki w covolume

Słabe połączenie

- Duże opóźnienia przy zamykaniu pliku

Write-disconnect

- Odczytujemy z serwerów
- Zapisujemy zmiany jak w disconnected operation

Podumowanie

- Dobra wydajność
- Sensowna semantyka
- Dobra skalowalność
- Wygoda przy użyciu komputerów przenośnych
- Możliwość korzystania z niepewnych sieci (IR, packet radio)