

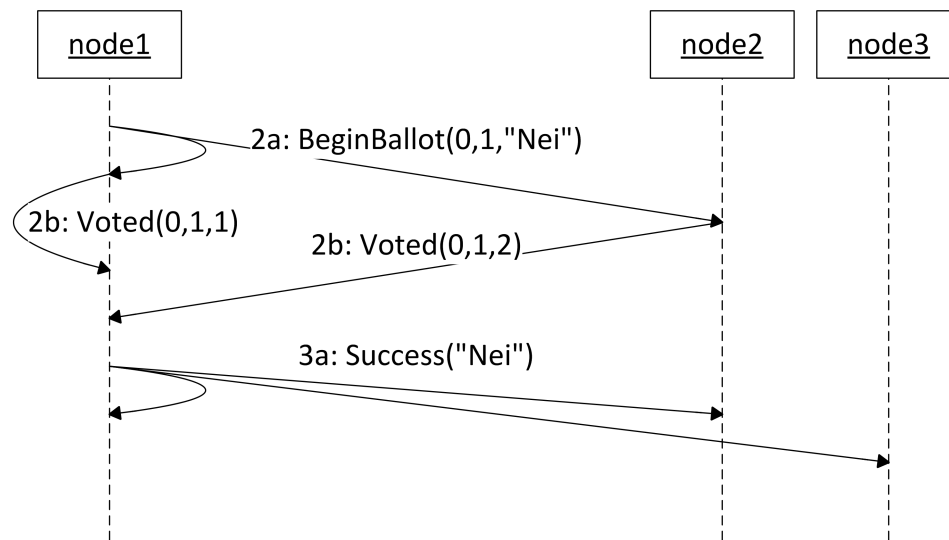
Paxos consensus

Anta at det er 3 noder og at node 1 vil fremme dekretet "NEI" og de andre nodene dekretet "JA" hvis de får muligheten til det.

1. Forklar forløpet i Paxos consensus hvis node 1 er utpekt til å starte protokollen og det bare trengs én avstemningsrunde for å avslutte protokollen.
2. Forklar forløpet hvis node 1 og node 2 starter protokollen samtidig og en av de to lykkes med sin avstemningsrunde.
3. Anta at node 1 og node 2 starter protokollen samtidig. Gi et eksempel på et forløp der Paxos consensus ikke terminerer.
4. Anta at det ikke er noe i veien med nettet eller nodene slik at alle meldinger kommer frem i rimelig tid. Hvis alle nodene ønsker å fremme samme forslag til dekret (f.eks. "JA"), vil Paxos consensus da alltid terminere? Forklar.
5. Betrakt følgende situasjon: Node 1 er utpekt til å starte protokollen, og "overbooker" ved å sende fase 2a-meldinger til både node 2 og 3. Den planlegger å benytte enten $\{1,2\}$ eller $\{1,3\}$ som det egentlige kворumet i fase 2, avhengig av hvilken som svarer først; node 1 trenger bare svar fra en av dem for å kunne fortsette med fase 3. Anta at det bare trengs én avstemningsrunde fra node 1 for å avslutte protokollen fordi node 2 besvarer alle meldingene den får fra node 1. Imidlertid er nettverket til node 3 veldig tregt, så avstemningsrunden til node 1 avsluttes uten at node 3 mottar noen meldinger. Først etterpå oppfører nettet seg normalt igjen, og node 3 mottar en svært forsinket fase 2a-melding fra node 1, men ingen fase 3-melding. Forklar hvordan node 3 kan benytte protokollen til å få resultatet av avstemningen.

Løsningsforslag

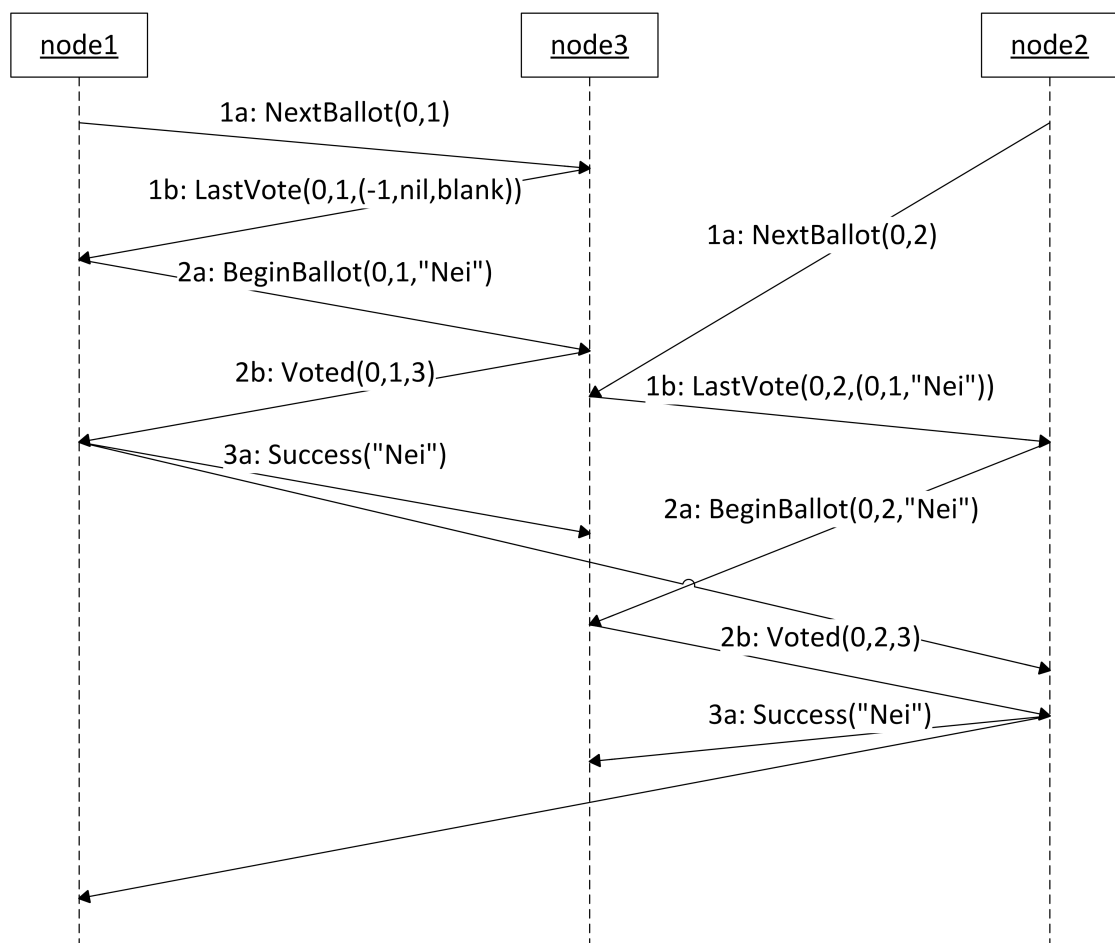
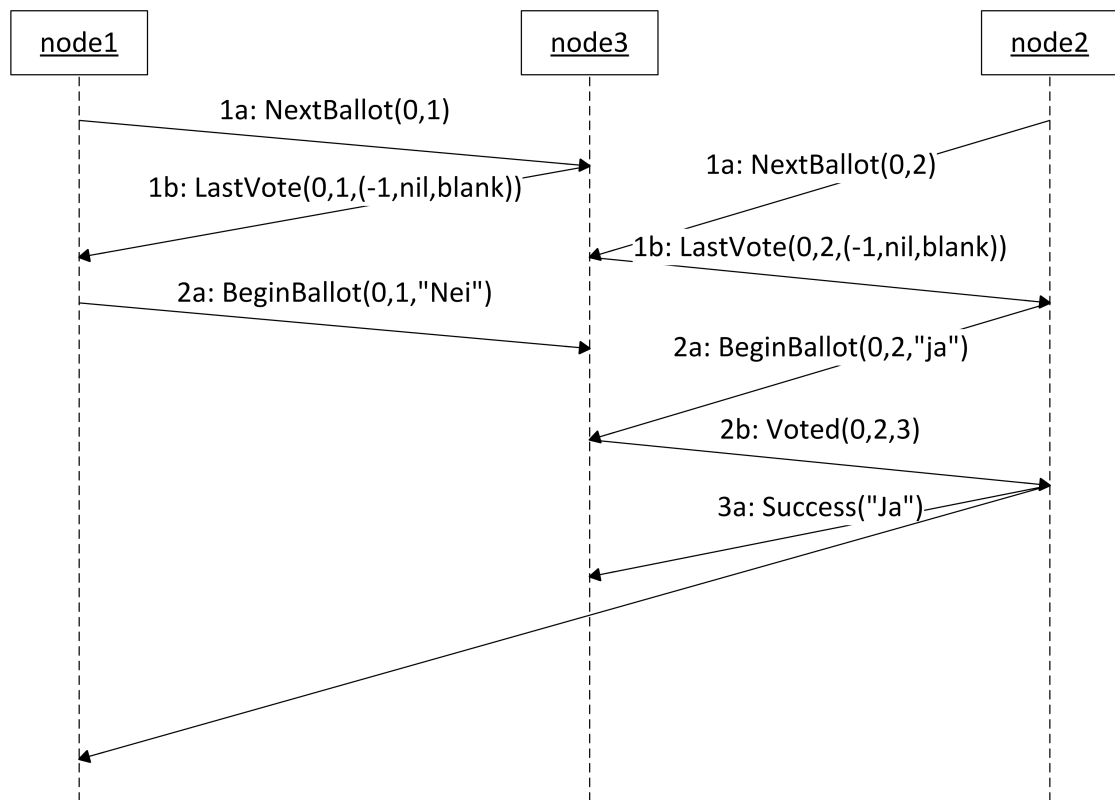
Oppgave 1.



Kommentarer:

- Når en bestemt node er utpekt til å starte protokollen, så starter den protokollen med rundenummer 0 i fase 2a.
- I figuren over har node 1 valgt $Q = \{1,2\}$ som den majoriteten den velger å benytte i fase 2. Vi har tegnet inn at node 1 sender meldinger også til seg selv, men i praksis vil det jo ikke gå noen meldinger internt i en node - node 1 vil simpelthen justere innholdet av sine lokale variable til det innholdet de skal ha etter hver fase, som om det ble sendt meldinger internt.
- Protokollen sier at det skal sendes fase 2a-meldinger til en majoritet av nodene. Hvis antall noder er $2F+1$, betyr dette at det skal sendes melding til minst $F+1$ noder. Det er ikke noe i veien for å sende meldingen til langt flere enn $F+1$ noder. Da kan fase 3 påbegynnes straks $F+1$ svar er kommet inn. Så i det gitte tilfellet med 3 noder kunne det vært like lurt å sende fase 2a-meldinger til både node 2 og 3. Da kan node 1 fortsette med fase 3 i det øyeblikket den får en 2b-melding fra en av de andre.

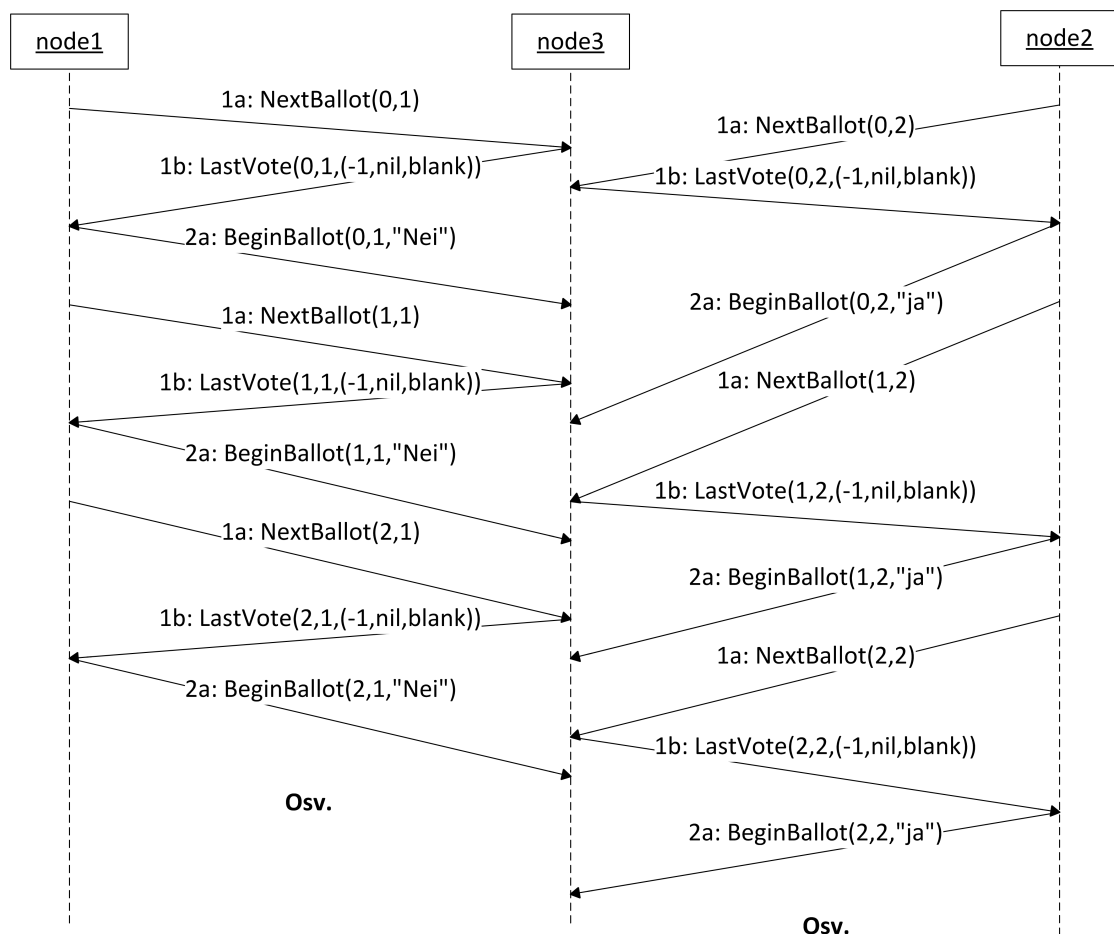
Oppgave 2.



Kommentarer:

- Det er flere mulige forløp; over vises to av dem. I begge benytter node 1 mengden $\{1,3\}$ og node 2 mengden $\{2,3\}$ i fase 1 og 2. I begge mottar dessuten node 3 en fase 1a-melding fra node 1 før den mottar en fase 1a-melding fra node 2. Node 1 og 2 starter begge med rundenummer 0. Da er det den noden som har høyest id som har høyest rundenummer, dvs. node 2. (Vi har i figurene utelatt meldinger som går fra en node til noden selv.)
- Tilfelle 1: Node 3 får fase 1a-meldingen fra node 2 før den får fase 2a-meldingen fra node 1. Da vil node 3 følge opp avstemningsrunden til node 2, så ytterligere meldinger fra node 1 vil bli ignorert. Node 3 besvarer 2a-meldingen fra node 2. Når node 2 får 2b-svaret fra node 3, har den et flertall (node 3 og seg selv) med dekretet "Ja", og avrunder med fase 3.
- Tilfelle 2: Node 3 får fase 1a-meldingen fra node 2 etter at den har fått (og besvart) fase 2a-meldingen fra node 1. Da vil node 3 følge opp avstemningsrunden til node 2, men node 1 har alt fått det flertallet den trenger (node 3 og seg selv) med dekretet "Nei" og avrunder med fase 3. Node 2 får via 1b-svaret fra node 3 vite at dekretet "Nei" allerede er blitt fremmet, og fortsetter derfor med å fremme dette dekretet i fase 2. Når node 2 får 2b-svaret fra node 3, avrunder den med dekretet "Nei" i fase 3.

Oppgave 3.



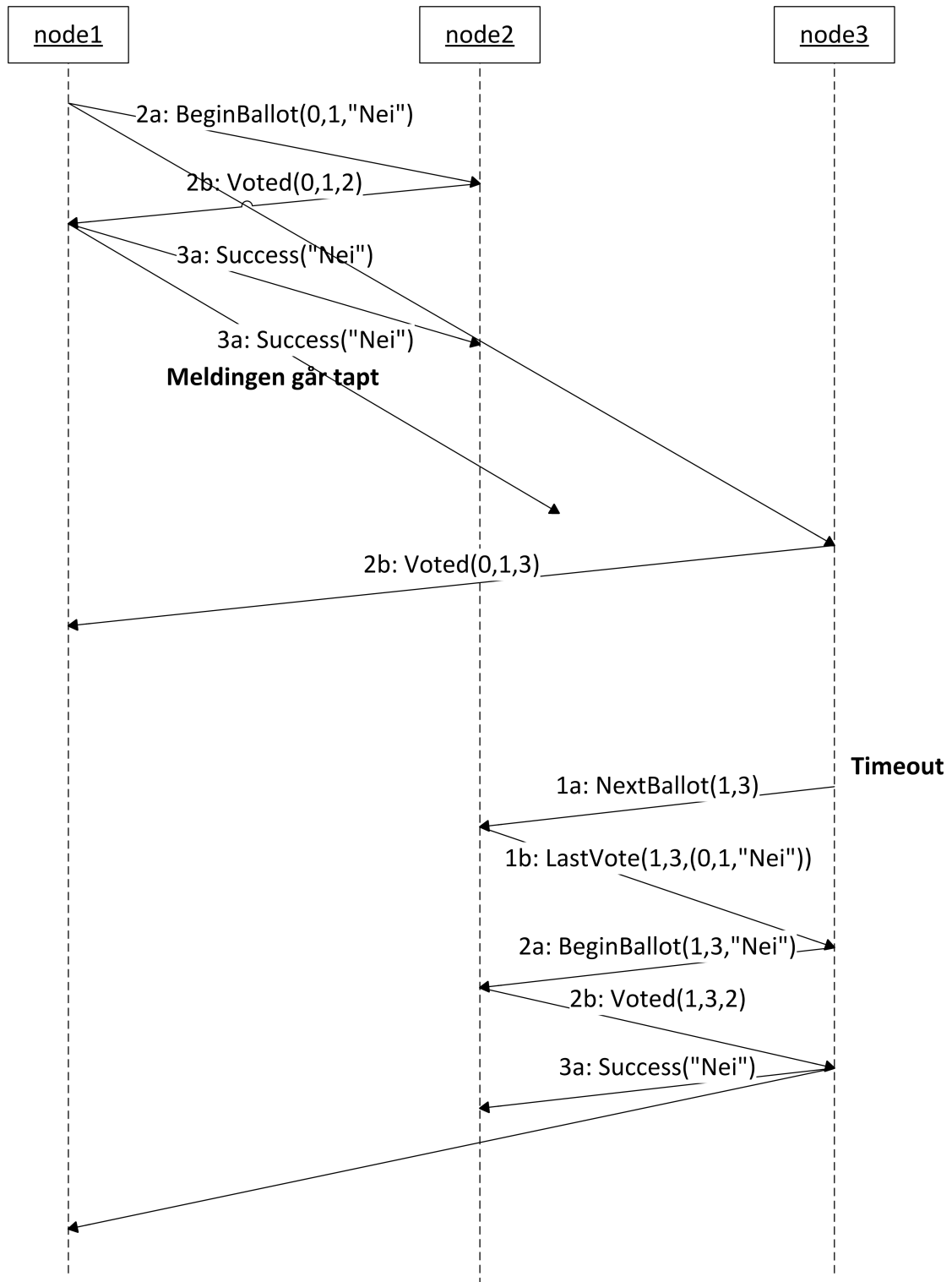
Kommentarer:

- Node 1 og node 2 starter på nye avstemningsrunder når de ikke får svar på sine 2a-meldinger innen en viss tidsfrist. Men denne tidsfristen er satt så kort at det stadig påbegynnes nye runder. Hver gang node 3 ser en ny fase 1a-melding, har den et høyere rundenummer enn det noden har sett til nå, og noden skal da følge opp denne avstemningsrunden og droppe meldinger for andre (lavererangerte) runder. Dermed besvarer ikke node 3 noen av de fase 2a-meldingene den mottar. (Figuren viser ikke meldinger fra en node til noden selv.)
- En node kan også ødelegge for sine egne avstemningsrunder hvis tidsfristene for å få svar er for korte. F.eks. kan node 1 stadig starte på nye avstemningsrunder umiddelbart etter at den har sendt fase 2a-meldingene sine og før den får svar på disse.

Oppgave 4.

Nei, terminering har ikke noe å gjøre med hva dekretet inneholder. Forløpet i eksempelet i oppgave 3 er uavhengig av dekretenes.

Oppgave 5.



Kommentarer:

- Noder kan starte ytterligere avstemningsrunder for å få vite resultatet av protokollen.
- I oppgaven var en av ideene bak formuleringene at node 3 skjønner at protokollen er påstartet ved at den mottar fase 2a-meldingen fra node 1. Denne meldingen er svært forsinket, men protokollen sier ikke noe om at node 3 må være klar over dette; node 3 besvarer i utgangspunktet alle meldinger på vanlig vis. Imidlertid er det viktig at den på ett eller annet tidspunkt tar initiativ til egne avstemningsrunder for å få fremdrift i protokollen. Til dette kan det benyttes en timeout. I forløpet over velger node 3 ved timeout å benytte {2,3} som majoritetsmengde i fase 1 og 2. Node 2 vil besvare ytterligere avstemningsrunder (med høyere rundenummere) selv om den har fått resultatet av avstemningen. Via svaret på fase 1 får node 3 så vite at dekretet "Nei" har vært oppe til avstemning, og benytter derfor dette dekretet i fase 2. Node 2 følger lojalt opp også denne fasen, og node 3 kan konkludere med at dekretet "Nei" er vedtatt (fordi den har fått svar fra alle i sitt kворum) og sende fase 3-melding til alle. Det er ikke noe problem for en node å motta flere fase 3-meldinger; de vil alltid inneholde samme dekret. (Figuren viser ikke meldinger fra en node til noden selv.)
- Generelt bør/må protokollen implementeres slik at alle noder som er involvert i protokollen, vet om dette slik at de kan ta et selvstendig ansvar for protokollen ved behov. Nodene kan benytte timeouts for å avgjøre om de skal gjøre noe for å drive protokollen fremover. Hvis f.eks. node 3 ikke hører noe som helst fra noen av de andre nodene innen en viss tidsfrist, bør den selv starte på en avstemningsrunde. Den starter da en avstemningsrunde med rundenummer 1 i fase 1a. (Siden node 1 var forhåndsutpekt til å starte protokollen i dette tilfellet, bør rundenummer 0 reserveres for node 1; de øvrige nodene starter sine avstemningsrunder med 1 eller høyere.) Hvis rundenummer 1 er for lavt i forhold til de rundene andre noder har iverksatt og sagt seg villig til å delta i, vil det ikke komme noen svar. Da vil noden etter en ny timeout forsøke med rundenummer 2, osv.