Progetto di Sistemi Distribuiti

Dott. Diego Borsoi
Dott. Filippo Callegari
DMIF, University of Udine, Italy

Version 0.1, 13 aprile 2021

Capitolo 1

Introduzione

Il progetto in questione riguarda la creazione di un sistema distribuito per la comunicazione fra dispositivi all'interno di una rete sparsa, tramite l'utilizzo di eventi.

1.1 Descrizione del problema

Ogni dispositivo corrisponde ad un nodo della rete ed è caratterizzato da:

- **Id** : numero univoco del nodo.
- **Stato** : tupla di variabili che rappresentano delle specifiche caratteristiche del nodo (es. temperatura di un sensore, stato di accensione di una termocoppia, ecc).
- **Regole** : insieme di regole del tipo ECA (Event Condition Action) che possono attivarsi a seguito di un evento inviato al nodo. Queste regole possono essere di due tipi: locali, l'azione modifica solamente lo stato del nodo in cui si attiva, oppure globale, l'azione viene inviata a tutti i nodi della rete perché venga letta ed, in caso la valutazione della guardia associata sia positiva, eseguita.

$$\{event;\ condition;\ action \mid \text{if } guard \text{ then } action\}$$

Lo stato della rete si evolverà ogni qual volta un evento verrà attivato, andando a sua volta ad innescare eventuali nuovi eventi e creando quindi una sequenza di azioni a cascata.

1.2 Struttura dell'implementazione

La rete in questione ha una struttura a mesh sparsa (cioè ogni nodo sarà al più connesso a un numero di nodi molto basso, rispetto alla totalità). I nodi sono idempotenti in modo tale da avere un sistema fortemente decentralizzato. Le

varie comunicazioni fra i nodi sono eseguite al di sopra di connessioni TCP, in tal modo possiamo garantire la consegna di ogni messaggio nell'ordine prestabilito. Per quanto concerne invece le comunicazioni riguardanti il sistema di *heartbeat* (3.2), queste vengono eseguite utilizzando connessioni UDP.

1.3 Trasparenze

Di seguito sono descritte le trasparenze che concergono e sono implementate dal sistema:

Trasparenza ai fallimenti: nel momento in cui un nodo fallisce/si disconnette, il resto della rete continua a funzionare normalmente.

Trasparenza alla scalabilità: la rete può espandersi in dimensione senza che il funzionamento dei nodi vari.

Trasparenza alla mobilità: un nodo può spostarsi all'interno della rete senza che il funzionameto suo e degli altri nodi vada a modificarsi.

1.4 Algoritmi

Il sistema implementa solamente due algoritmi:

- **Flooding Algorithm:** viene usato per la comunicazione di un'azione a tutta la rete nel momento in cui in un nodo una regola globale viene attivata.
- **Lamport clock (modificato):** viene utilizzata una versione modificata del Lamport clock per identificare i vari $flood$ eseguiti; questo clock viene incrementato solamente dall'invio (o ricezione) di un messasggio, e non dalle azioni interne ad un nodo.

Il sistema non implementa particolari algoritmi essendo che si vuole realizzare una rete distribuita dove ogni nodo conosce esclusivamente i vicini ed evolve il suo stato solamente a causa di eventi ricevuti tramite dei messaggi.

1.5 Testing

Per testare il sistema verrà utilizzata un'entità *Ambiente*, la quale simulerà:

- la creazione iniziale della rete, caricando da dei file appositi la struttura degli stati, la lista di regole e la conformazione della rete
- la scoperta di nuove connessioni
- variazioni di variabili legate all'ambiente (es. temperatura registrata da un sensore)

- fallimenti di nodi
- ritardi nell'invio di messaggi fra nodi

Ogni modulo verrà testato singolarmente ed infine verranno eseguiti dei test completi del sistema.

1.6 Piano di sviluppo

Le future fasi di sviluppo seguiranno il seguente ordine:

1. Riunione con il committente per convalidare la risoluzione del problema
2. Implementazione ambiente virtuale per la gestione dei nodi
3. Implementazione della struttura del nodo
4. Implementazione sistema *heartbeat*
5. Implementazione del sistema algoritmico
6. Test totale
7. Validazione

Capitolo 2

Analisi

In questo capitolo vengono descritti nel dettaglio requisiti funzionali e non funzionali della soluzione proposta.

2.1 Requisiti Funzionali

I requisiti funzionali individuati sono:

- **Categorizzazione di un nodo:** ogni nodo ha un tipo il quale ne identifica lo stato e le sue regole;
- **Modifica delle regole di un nodo:** ogni tipo di nodo può avere le sue regole, codificabili attraverso la programmazione dello stesso;
- **Modifica dello stato di un nodo:** ogni evento permette di avere o degli *effetti locali* o degli *effetti globali*:
  - ***effetti locali***: la regola va a modificare lo stato interno;
  - ***effetti globali***: la regola può modificare lo stato delle variabili interne, e può generare un evento sugli altri nodi;
- **Aggiunta dinamica di un nodo:** un nodo può essere aggiunto alla rete in qualsiasi momento senza perturbarne la dinamicità, limitando l'aggiornamento ai vicini a cui si collega.
- **Esecuzione di un'azione ricevuta dai vicini:** nel momento in cui un nodo riceve un messaggio dai propri vicini esso va a verificare la soddisfacibilità della guardia (se presente) e nel caso di una valutazione positiva viene eseguita l'azione associata, andando quindi a modificare il proprio stato.
- **Attivazione di una regola:** ogni qual volta avviene un cambiamento nello stato di un nodo, viene eseguito un controllo delle regole, per vedere se gli eventi generati possano attivare una o più regole del nodo; nel caso in cui una regola venga attivata, in base al tipo (locale o globale) viene portata a termine l'azione corrispondente.

2.2 Requisiti Non Funzionali

I requisiti non funzionali individuati sono:

- **Decentralizzazione:** nessun nodo ha il controllo dell'ordine degli eventi, grazie al fatto che ogni nodo è idempotente;
- **Tolleranza ai guasti:** poiché tutti i nodi sono idempotenti, nel momento in cui un nodo si scollega dalla rete, la rete rimanente continua ad operare normalmente;
- **Etereogenità:** fintanto che i nodi aggiunti utilizzano il protocollo descritto, qualunque nodo di qualunque tipo (hardware o categoria) potrà essere aggiunto alla rete;
- **Scalabilità:** l'aggiunta dinamica dei nodi alla rete permette di scalare orizzontalmente con estrema facilità;
- **Trasparenze:** le trasparenze implementate sono quelle descritte al capitolo precedente (paragrafo 1.3).

Capitolo 3

Progetto

In questo capitolo vengono descritti in modo più approfondito l'achitettura del progetto, i moduli, i protocolli e gli algoritmi utilizzati.

3.1 Architettura logica

Essendo tutti i nodi costruiti al medesimo modo, di seguito presentiamo la struttura di uno singolo di essi.

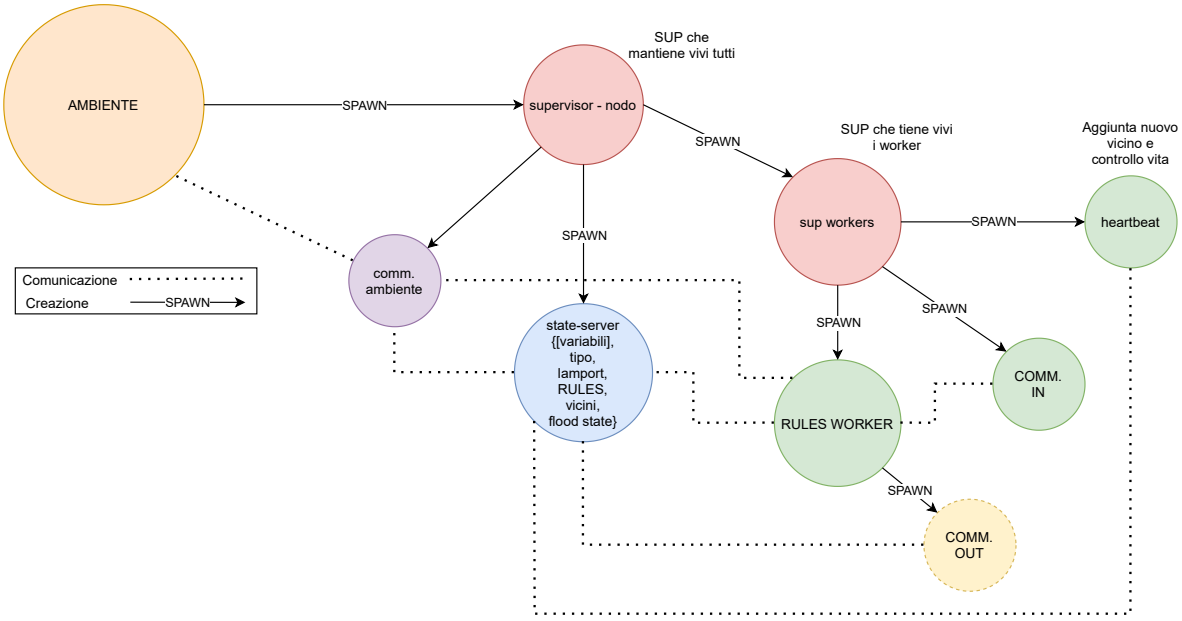

Figura 3.1: Struttura gerarchica dei moduli di un nodo e visualizzazione delle connessioni fra di essi.

Come si può vedere dalla figura 3.1, ogni nodo è formato dai seguenti moduli:

- **Supervisor nodo:** modulo che cerca di manterene sempre operativi gli altri moduli interni. Se questo componente si riavvia equivarrebbe ad un riavvio del nodo, e quindi la conseguente perdita della modifica agli stati interni. I moduli da lui controllati sono:
  - **State server:** questo modulo manterrà tutte le variabili locali al nodo, che verranno modificate durante l'operatività dello stesso. Queste variabili sono:
    - stato del nodo;
    - lista delle regole associate al nodo;
    - tipo del nodo;
    - stato delle waves;
    - informazioni sui nodi vicini;
    - id univoco del nodo;
  - **Supervisor dei workers:** modulo che si occupa di gestire i moduli a lui dipendenti, riavviandoli in caso di "crash". Questi sono:
    - **heartbeat sense:** si occupa di controllare la "vitalità" dei vicini e di istanziare le connessioni con essi;
    - **communication in:** si occupa di gestire tutti i messaggi in ingresso relativi al nodo in questione;
    - **rules worker:** si occupa di applicare le azioni ricevute dai nodi vicini ed eventualmente eseguire una delle regole a lui locali al momento dell'attivazione; lui si occuperà anche di propagare le regole che generano degli eventi verso gli altri nodi, interpellando il modulo "*communication out*" (modulo apposito per l'invio dei messaggi ai nodi vicini).

Come è stato accennato in precedenza verrà sviluppato un ulteriore modulo chiamato "***Ambiente***": questo modulo permette di simulare le interazioni che avverrebbero nel mondo reale. Nel dettaglio, le funzionalità del modulo sono:

- il "discovery" dei vicini;
- cambiamento delle variabili non dipendenti dalle regole (temperatura dell'ambiente/GPS/...);
- simulazioni di disservizi di rete;
- simulazione di guasti (temporanei o non) di un nodo;
- topologia della rete.

Dal punto di vista del nodo ci troviamo quindi costretti ad aggiungere un ulteriore modulo fittizio (***comunicazione ambiente***) per permettere la comunicazione con l'ambiente.

3.2 Protocolli ed algoritmi

Di seguito verranno descriti nel dettaglio i vari protocolli ed algoritmi utilizzati.

Controllo e attivazione delle regole

Nel momento in cui il sistema riceve un'azione da eseguire (dopo aver controllato la validità e che appartenga ad una wave non ancora ricevuta) si innesca la seguente serie di azioni:

1. il modulo *communication IN* invia l'azione da eseguire al modulo *rules worker*;
2. quest'ultimo utilizza una funzione dello *state server* per modificare lo stato del nodo in accordo all'azione ricevuta;
3. il *rules worker* esegue quindi un controllo sulle regole andando ad identificare quali possono essere attivate dalla modifica appena eseguita;
4. per ciascuna regola che viene attivata viene testata la condizione e in caso di risultato positivo:
   - se la regola è del tipo *locale*, viene eseguita l'azione associata
   - se invece la regola è del tipo *globale*, viene eseguita l'azione associata (in caso di guardia con valutazione positiva) e viene passata ad un processo di *communication OUT*, istanziato appositamente, che genera una nuova wave di messaggi inviando ai vicini la nuova azione.

Heartbeat

L'algoritmo di "heartbeat" serve per mantenere consistente lo stato dei vicini di un nodo. Questo infatti controlla la loro vitalità e sarà componente chiave per l'aggiunta di un nuovo nodo.

L'algoritmo si suddivide quindi in due componenti:

- *ECHO*: similmente al protocollo ICMP, si fa una richiesta di echo al vicino. Se questa "*ECHO_RQS*" andrà a buon fine, il primo nodo che istanzia una richiesta di "*ECHO*" richeverà un pacchetto di "*ECHO_RPL*". Definiamo come τ il tempo che intercorre tra un messaggio di *ECHO* ed un'altro. Se un nodo non rispondesse entro 2τ ad una *ECHO_RQS*, questo verrà considerato come non più collegato. Ogni *ECHO_RPL* conterrà il "Lamport" attuale del vicino.
- *ADD_NEW_ND*: similmente al protocollo DHCP, nella fase di aggiunta di un nodo alla rete, il nuovo nodo si annuncia al suo vicino "fisico", chiedendo le informazioni essenziali per poter partecipare attivamente alla rete. Questo sarà spiegato più in dettaglio nella prossima sezione.

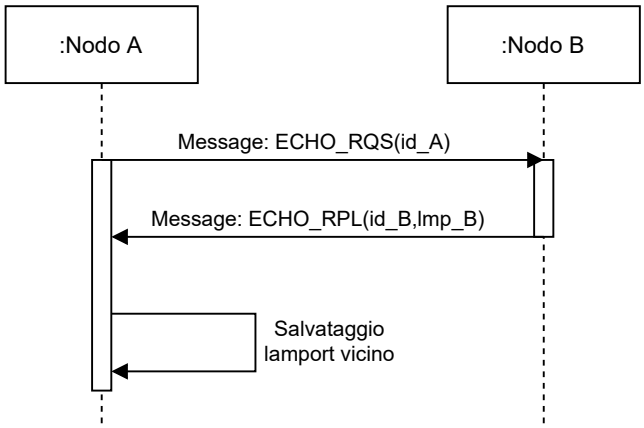

Figura 3.2: Sequence diagram dei messagi usati per il sistema di *heartbeat*.

Aggiunta di un nuovo nodo

L'aggiunta di un nodo è una parte complicata del sistema: bisogna tener conto della possibilità che la rete si partizioni. Questo si può verificare nel caso in cui un nodo si riavvii. Il partizionamento della rete è visto come caso particolare di aggiunta di un nodo alla rete.

L'aggiunta di un nuovo nodo si compone dei seguenti passi:

1. *ADD_NEW_ND*: il nuovo nodo manda la richiesta di aggiunta alla rete a tutti i suoi vicini inviando il proprio id;

2. *ADD_NEW_NB*: il nodo che deve aggiungere il nuovo nodo risponde con un messaggio contenente $(lamport, id)$.

A questo punto, una volta che per ogni vicino ho i suoi parametri di lamport, i casi possibili sono 2:

1. **tutti i nodi hanno medesimo lamport:** imposto il mio lamport al lamport comune;

2. **esiste un lamport massimo:** imposto il mio lamport al valore massimo, e rispondo con *UPD_LMP* a tutti i miei vicini che non hanno il lamport al massimo. Questi a loro volta manderanno a tutti i loro vicini, con lamport diverso da quello scelto, il nuovo valore.

Algoritmo: Flooding

Ogni qual volta si crea una regola che genera un evento "globale", ogni nodo spedirà ai suoi vicini un messaggio contenenta un'azione da eseguire. Questi, una volta ricevuto, aggiorneranno il loro "wave count" (tenuto dal Lamport

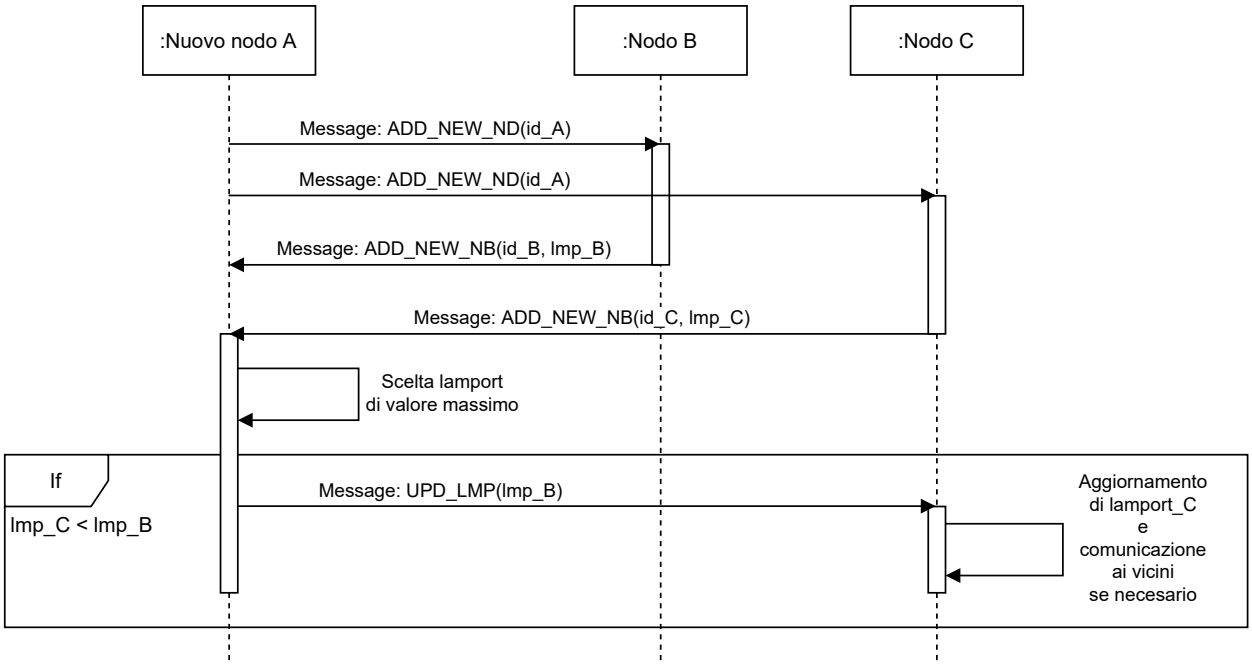

Figura 3.3: Sequence diagram dei messagi usati per laggiunta di un nuovo nodo alla rete.

modificato, spiegato successivamente), e passerà all'esecuzione dell'istruzione condizionata contenuto nel messaggio. Al fine di evitar la presenza di messaggi vecchi nella rete, ogni messaggio conterrà la coppia $(lamport, id\_gen)$: questo verrà salvato localmente nel nodo ricevente, e verrà mantenuto in memoria al fine di verificare se un messaggio ricevuto non è già stato eseguito. Ogni nodo quindi spedirà una copia del messaggio a tutti i suoi vicini (a patto non l'abbia già ricevuto in passato), meno a quello da cui l'ha ricevuto. Queste strategie applicate faranno si che i messaggi circolanti nella rete siano il minor numero possibile, e nell'eventuale creazioni di cicli nella rete, dovuta alla topologia della stessa, siano soppressi alla prima occasione utile.

Algoritmo: Lamport modificato

Il "Clock di Lamport" ci permette di dare una certa conseguenza temporale alle azioni. Questo sarà un numero intero crescente, ed identificherà ogni wave generata. Alla generazione di una wave ogni nodo userà il suo lamport interno, incrementato, per identificarla. Questo ci permette, come già abbiamo spiegato, di controllare la propagazione dei messaggi. Sia quindi Lmp_i il lamport della nuova wave ricevuta e Lmp_n il lamport del nodo corrente. Le azioni intraprendibili sono:

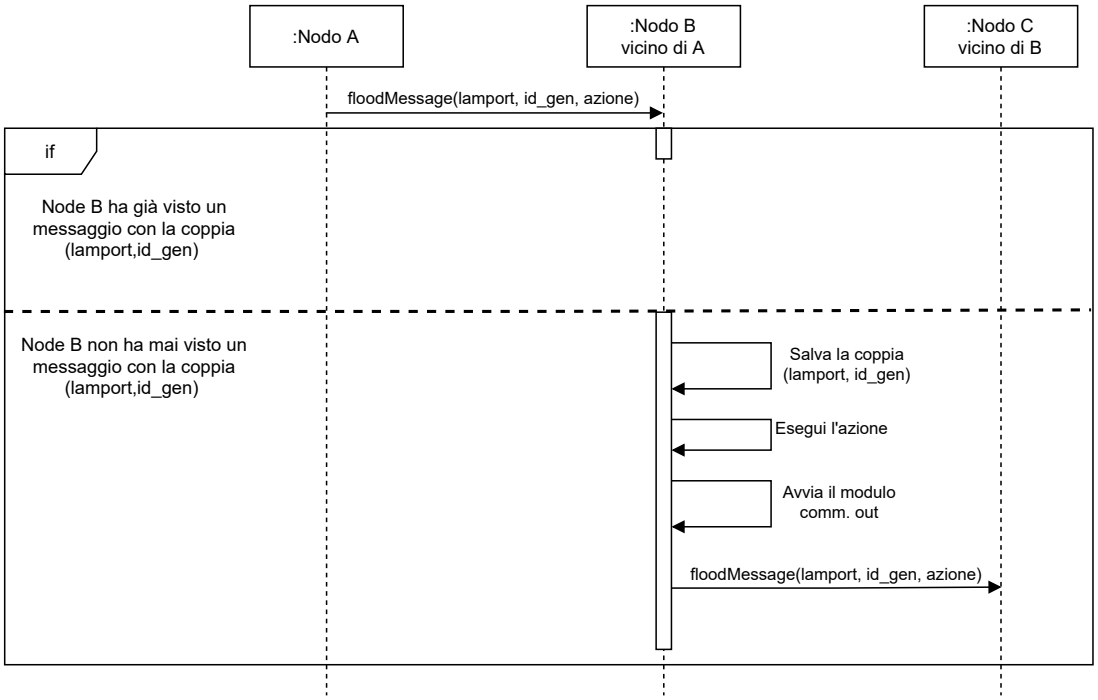

Figura 3.4: Sequence diagram dei messaggi usati per l'algoritmo di flooding.

- $Lmp_i = Lmp_n + 1$: eseguo immediatamente l'azione in essa contenuta ed aggiorno il valore di Lmp_n;
- $Lmp_i > Lmp_n + 1$: aspetto un preimpostato timeout (es: 5τ) prima di eseguire l'azione di Lmp_i, in modo tale da poter ricevere le wave mancanti, e quindi mantenere l'ordine causale delle azioni;
- $Lmp_i < Lmp_n$: vado ad eseguire l'azione solo nel caso in cui tutte le variabili coinvolte non siano già state modificate da una wave ad essa successiva, cioè con $lamport > Lmp_i$.

3.3 Architettura fisica e deployment

Per quanto riguarda l'architettura fisica è necessario l'utilizzo di microcalcolatori, come dei "Raspberry" o degli "Arduino". Ogni nodo corrisponderebbe fisicamente ad una di queste schede, avendo quindi la possibilità d'utilizzare più nodi per il medesimo "apparato".

Non è strettamente necessario l'ausilio di microcalcolatori: potrei concentrare all'interno di un singolo calcolatore più nodi, a patto che siano gestiti in maniera consona.

Come descritto in precedenza, questi comunicherebbero con protocollo TCP ed UDP, non interessandoci quindi di tutta la rete sottostante.

Visto l'esempio a cui abbiamo pensato, si ritiene ideale l'utilizzo di connessioni wireless.

3.4 Piano di sviluppo

Le **funzionalità di base** che verranno implementate sono:

- comunicazione tra nodi;
- sistema di flooding;
- sistema di heartbeat;
- gestione dell'aggiunta di un nodo alla rete;
- gestione del riavvio di un nodo nella rete;
- impostazione iniziale dei nodi (ambiente);
- programmazione dei nodi;
- sistema basico d'esecuzione delle regole.

Sono state inoltre individuate le seguenti **funzionalità avanzate**:

- sistema avanzato d'esecuzione delle regole;
- salvataggio dello stato del nodo su files interni al controllore;
- riprogrammazione dinamica del nodo;
- implementazione di un *calculus* locale.