

Università degli Studi della Calabria

Facoltà di Ingegneria

Istituto per il Calcolo e le Reti ad Alte Prestazioni – ICAR/CNR

Dipartimento di Elettronica, Informatica e Sistemistica – DEIS

Corso di Laurea in Ingegneria Informatica

TESI DI LAUREA

**Definizione di un protocollo di distribuzione
e scoperta di risorse su Griglia basato su un
sistema ad agenti**

Relatore

Ing. Carlo Mastroianni

Candidato

Raffaele Giordanelli

Matr. 76162

Anno accademico 2005/2006

Indice

| | |
|---|---------|
| Introduzione | pag. 3 |
| 1. Tecnologie di Rete e Paradigmi Correlati | pag. 7 |
| 1.1. Grid computing | pag. 8 |
| 1.1.1. Definizione di Griglia Computazionale | pag. 9 |
| 1.1.2. Le motivazioni dell'uso delle Griglie | pag. 11 |
| 1.1.3. Il Cluster | pag. 13 |
| 1.1.4. Le applicazioni delle Griglie | pag. 16 |
| 1.1.5. La comunità delle Griglie | pag. 18 |
| 1.2. Peer-to-Peer e Griglie Computazionali | pag. 19 |
| 1.3. Web Services | pag. 24 |
| 1.4. Swarm Intelligence e Sistemi ad Agenti | pag. 28 |
| 1.5. Clustering e Sorting | pag. 35 |
| 2. Riorganizzazione e scoperta delle risorse in una Griglia | pag. 40 |
| 2.1. Riorganizzazione delle risorse: il protocollo ARMAP | pag. 42 |
| 2.1.1. Entropia | pag. 45 |
| 2.1.2. Gestire una Griglia dinamica con ARMAP | pag. 47 |
| 2.2. Scoperta delle risorse: il protocollo ARDIP | pag. 48 |
| 2.2.1. Identificazione dei peers rappresentativi | pag. 49 |
| 2.2.2. Ricerca parzialmente informata | pag. 50 |
| 2.2.3. Meccanismo di stigmergia | pag. 52 |
| 3. Simulazioni e risultati | pag. 53 |
| 3.1. Risultati sul protocollo ARMAP | pag. 55 |
| 3.2. Scalabilità del protocollo ARMAP | pag. 59 |
| 3.3. Risultati sul protocollo ARDIP | pag. 61 |
| 3.4. Scalabilità del protocollo ARDIP | pag. 72 |
| 4. Conclusioni | pag. 74 |
| Bibliografia | pag. 76 |

Introduzione

Una Griglia Computazionale è un ambiente sviluppato per il calcolo distribuito. Gli utenti non condividono solo files, come generalmente avviene con i peer to peer, bensì risorse software e hardware, più generalmente capacità d'elaborazione.

Un ruolo fondamentale nel funzionamento di una Griglia è assunto dall'*Information System*, sistema informativo. Gli utenti si servono di questo modulo per scoprire le risorse offerte dai nodi di una Griglia, esplorare le caratteristiche di tali risorse e controllare la loro disponibilità. Il servizio chiave fornito dai Grid Information Systems è il servizio di resource discovery, che è invocato da un utente, quando deve scoprire ed usare le risorse software o hardware.

Nella maggior parte dei framework di Griglia sviluppati finora, l'information system è strutturato generalmente secondo un approccio centralizzato o gerarchico. In entrambi i casi, il difetto di tali soluzioni è la limitata scalabilità, dovuta al fatto che con tali modelli si va a creare un collo di bottiglia su quei nodi che rivestono un ruolo di coordinazione nell'information system.

Un information system può invece essere reso scalabile, e più efficiente, capovolgendo il concetto su cui è costruito: non più approcci centralizzati o gerarchici, ispirati quindi al modello client-server, bensì basare il sistema informativo su un modello distribuito, tramite il paradigma peer to peer, rendendo ogni nodo della Griglia indipendente dagli altri.

Il peer to peer è un paradigma il cui successo è dovuto alle applicazioni di file sharing, e che è esploso tra la fine degli anni '90 e l'inizio del nuovo millennio.

Senza scendere nei particolari, è utile ricordare che questo sistema, contrapponendosi al modello client-server, e se

opportunamente calibrato, porta a distribuire il carico e le risorse tra gli utenti (o peers, cioè utenze di pari grado).

In tal modo il sistema informativo può funzionare indipendentemente dalla quantità di nodi e di risorse nella Griglia.

Un sistema informativo di Griglia siffatto deve basarsi su due caratteristiche principali: la replica e la distribuzione di informazioni circa risorse e servizi, e, successivamente, un'appropriata distribuzione di tali informazioni tra gli hosts della Griglia stessa.

Questa soluzione si basa sulle caratteristiche di meccanismi epidemici, studiati per la disseminazione di informazioni nei sistemi distribuiti, al fine di aiutare la scoperta delle risorse[1, 2], e su sistemi auto adattativi, in cui una sorta di *intelligenza di massa* emerge dal lavoro di una moltitudine di agenti semplici che interagiscono con l'ambiente in cui vivono[3].

Con questo sistema, le informazioni riguardo le risorse sono replicate e disseminate attraverso la Griglia da degli agenti, con l'intento di creare ristrette regioni nella rete a bassa entropia, le quali siano specializzate in una particolare classe di risorse.

Successivamente, un protocollo di ricerca parzialmente informato guida i messaggi di query verso una regione i cui nodi possiedono un vasto insieme di risorse della classe richiesta.

Tanto gli agenti quanto i messaggi di query si muovono attraverso la rete indipendentemente dallo stato globale della Griglia, garantendo, così, scalabilità e robustezza.

Lavorando a questo sistema, si è assunto che le risorse offerte da una Griglia orientata ai servizi siano appunto *Servizi di Griglia*, i quali vengono gestiti e pubblicati da una *Organizzazione Virtuale di Griglia*.

Si è assunto inoltre che questi servizi possano essere classificati semanticamente in accordo alle loro caratteristiche: una classe di risorse è definita come un insieme di servizi che rispecchiano specifiche proprietà[4, 5].

Generalmente, infatti, una query non è inviata per ricercare una singola risorsa, ma, per raccogliere informazioni circa una risorsa appartenente ad una data classe. Dopo aver ricevuto un certo numero di responsi, l'utente o il cliente può scegliere le risorse che risultano più appropriate ai suoi fini.

Questa tesi definisce e propone l'uso di due protocolli: ARMAP e ARDIP.

ARMAP(Ant-based Replication and MApping Protocol) serve a disseminare le informazioni sulle risorse della Griglia, copiandole secondo una loro precedente classificazione in regioni specializzate in una particolare classe di risorse. Ciò si raggiunge tramite un approccio multi agente ispirato al comportamento delle formiche, che raccolgono e raggruppano oggetti dal loro ambiente[6].

ARDIP(Ant-based Resource DIsccovery Protocol) è un protocollo di ricerca parzialmente informato, che sfrutta la riorganizzazione delle risorse raggiunta tramite ARMAP. Il concetto principale è il seguente: se un grande numero di risorse della classe desiderata è immagazzinato in un'area ristretta della Griglia, diventa conveniente guidare le richieste verso tale regione, per massimizzare il numero di risorse trovate e minimizzare il tempo di risposta ed il traffico della rete.

Per la sua natura di protocollo parzialmente informato, ARDIP cerca di combinare i benefici di entrambe le tecniche di ricerca di risorse, cieca e informata, attualmente utilizzate nelle reti P2P. Infatti un approccio puramente cieco, usando quindi flooding o cammino casuale [7, 8], è particolarmente semplice ma causa un eccessivo carico della rete. D'altronde, un approccio puramente informato, basato quindi su tabelle di routing o ricerche probabilistiche adattative, generalmente richiede un'organizzazione delle risorse molto strutturata, poco pratica in Griglie ampie, eterogenee e dinamiche.

Un'operazione di ricerca ARDIP consta di due fasi. Nella prima è usato un cammino casuale[7, 8]: messaggi di query vengono inviati dal peer richiedente e viaggiano attraverso la Griglia con un meccanismo *cieco*, cioè non informato.

Nella seconda fase, nel momento in cui una query arriva sufficientemente vicino ad una regione specializzata, la ricerca diviene informata: la query sarà guidata verso un peer rappresentativo eletto nella regione stessa, trovando così un gran numero di risultati utili.

Uno dei maggiori pregi di questa soluzione è il fatto che gli agenti ARMAP e le query ARDIP si muovono senza conoscere lo stato globale della Griglia, ma si comportano secondo un meccanismo decentralizzato basato sul feromone, anche questo preso in prestito dal comportamento delle formiche.

Grazie a queste scelte, l'utilizzo combinato dei protocolli ARMAP e ARDIP si rivelerà pienamente scalabile, ed utilizzabile in una Griglia indipendentemente dal numero di risorse e dall'estensione della rete.

La tesi è strutturata come segue.

Il primo capitolo tratta le tecnologie che stanno dietro la definizione dei due protocolli: peer to peer, Griglie computazionali, sistemi ad agenti, clustering e sorting.

Il secondo capitolo entra nel dettaglio dei due protocolli, mentre il terzo illustra i risultati delle simulazioni effettuate per verificarne la bontà.

Nel quarto ed ultimo capitolo, infine, si traggono le dovute conclusioni.

1. Tecnologie di rete e paradigmi correlati

La progettazione e lo sviluppo dei protocolli ARMAP e ARDIP, per la definizione di un nuovo schema per la realizzazione dei Sistemi Informativi nelle Griglie Computazionali, ha richiesto una conoscenza più o meno profonda, secondo i casi, di alcune tecnologie, che in questo capitolo saranno illustrate, al fine di semplificare la comprensione del lavoro realizzato.

Il primo paragrafo descrive cosa sono le *Griglie Computazionali*, nonché gli scopi per cui sono nate. Questo paragrafo è necessario per comprendere l'ambito in cui si va ad inserire il lavoro descritto in questa tesi.

Il secondo paragrafo illustra l'odierna integrazione tra le Griglie e i sistemi *peer to peer*, mentre il terzo dà una panoramica dei metodi di ricerca di risorse utilizzati in una rete peer to peer.

Una risorsa è, in una Griglia Computazionale, un *Servizio di Griglia*, che, per motivi di interoperabilità tra sistemi e di adeguamento agli standard della rete, non è altro che un *Web Service*. Tale tecnologia è trattata nel quarto paragrafo.

Per quanto riguarda invece le specifiche tecniche del protocollo, nel quinto paragrafo è discussa la *Swarm Intelligence* nei *Sistemi ad Agenti* che permette, tramite un approccio distribuito (l'utilizzo di uno "sciame" di agenti), di ottenere risultati prossimi agli approcci centralizzati, con tutti i vantaggi di distribuzione del carico di lavoro che ciò comporta, nonché le tecniche di *Clustering* e *Sorting*, nel sesto paragrafo, che trovano ampia applicazione nel lavoro effettuato dagli agenti nella riorganizzazione delle risorse.

1.1. Grid Computing

Il termine *Grid Computing* (letteralmente, "calcolo a Griglia") indica un'infrastruttura distribuita per consentire l'utilizzo di risorse di calcolo e di storage provenienti da un numero indistinto di calcolatori (anche e soprattutto di potenza non particolarmente elevata) interconnessi da una rete (solitamente, ma non necessariamente, Internet). Il termine "Griglia" deriva dalla similitudine fatta dai primi ideatori del Grid Computing, secondo i quali in un prossimo futuro si sarebbe arrivati ad utilizzare le risorse di calcolo alla stessa stregua dell'energia elettrica, ovvero semplicemente attaccando una spina all'infrastruttura energetica (Power Grid).

L'idea del Grid computing, di cui recentemente si sente spesso parlare come la prossima rivoluzione dell'informatica (come a suo tempo fu il World Wide Web), risale a circa metà degli anni Novanta, ma l'idea di accesso remoto alle risorse informatiche di calcolo e di storage nasce da lontano. Nel 1969, Len Kleinrock, uno dei padri di Arpanet, scrisse quasi profeticamente:

"We will probably see the spread of 'computer utilities', which, like present electric and telephone utilities, will service individual homes and offices across the country."

Una delle prime definizioni proposte da Ian Foster e Carl Kesselman, nel 1998, relativamente alle Griglie computazionali, recita:

"A computational Grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities".

1. Tecnologie di rete e paradigmi correlati

In seguito, tale concetto si è arricchito a fronte di nuove tematiche politico-sociali; le Griglie Computazionali divengono quindi la piattaforma di riferimento per la condivisione coordinata delle risorse e per la soluzione di problemi in organizzazioni virtuali, dinamiche e multi-istituzionali:

"The sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources, as is required by a range of collaborative problem-solving and resource-brokering strategies emerging in industry, science, and engineering. This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules form what we call a virtual organization".

1.1.1 Definizione di Griglia Computazionale

Oggi, si possono riassumere e coordinare le varie successive definizioni in tre punti chiave, assumendo quindi che Grid, o Griglia Computazionale, è un sistema che [9]:

- Coordina risorse che non sono soggette ad un controllo centralizzato. (integra e coordina risorse ed utenti che non sono definite in un unico dominio di gestione, per esempio differenti unità amministrative di stesse o differenti compagnie, dove le varie politiche di sicurezza, appartenenza, autenticazione ed esecuzione divengono trasparenti);
- Utilizza interfacce e protocolli standard, aperti e general-purpose. (un sistema Grid si basa su tali protocolli ed interfacce per la gestione delle autorizzazioni, dell'autenticazione e per l'accesso alle risorse disponibili);
- Produce un'accettabile qualità di servizio. (un sistema Grid permette di gestire le proprie risorse per ottenere diverse qualità di servizio come tempi di risposta, disponibilità, sicurezza e performance, per soddisfare completamente la richiesta utente; in questo contesto le Griglie computazionali così ottenute rappresentano un sistema combinato più efficiente rispetto alla somma delle sue singole parti).

Particolare attenzione è rivolta al fatto che i diversi protocolli di gestione debbano essere standard: risulta infatti quasi scontato che per ottenere la condivisione di risorse fra ambienti diversi ed eterogenei sia necessaria la presenza di interfacce di comunicazione comuni.

1.1.2 Le motivazioni dell'uso delle Griglie

I computer sono utilizzati per modellare e simulare problemi scientifici ed ingegneristici complessi, diagnosticare analisi mediche, controllare equipaggiamenti industriali, previsioni del tempo, operazioni bancarie, biotecnologie e molto altro. L'osservazione principale è che la media degli ambienti computazionali esistenti risulta inadeguata ad assolvere tali funzioni al fine di ottenere obiettivi così sofisticati e complessi. Il termine "Computational Grid" si avvicina fortemente al concetto di "electric power Grid", modello delle reti di distribuzione dell'energia elettrica, al quale idealmente si ispira. Una rete elettrica rende oggi ampiamente disponibili e collega fonti di energia eterogenee (vi sono centrali elettriche di innumerevoli tipi diversi, che sfruttano le più svariate tecnologie), distanti, gestite da soggetti ed enti diversi: basti pensare al fatto che intere nazioni acquistano e vendono quotidianamente energia elettrica. Allo stesso modo si può pensare a dei prototipi di Grid quando si considerano le reti di comunicazione stradali, ferroviarie, marittime e aeree o le reti telefoniche, postali e di trasporto, che sono tutte entità complesse ed interdipendenti. Allo stesso modo le Griglie Computazionali dipendono da altre infrastrutture, come ad esempio cluster di varia natura, estendendone le capacità e rappresentandone, quindi, la naturale evoluzione. Per continuare con l'analogia della rete elettrica, è possibile notare che lo stato attuale delle risorse di calcolo è tutto sommato quasi analogo a quello dell'elettricità nei primi anni del '900.

Infatti in quegli anni era possibile generare energia elettrica, venivano inoltre costruiti macchinari che dipendevano da questa per funzionare, ma la necessità, per ogni utilizzatore, di dover disporre oltre al macchinario, anche di un generatore, ne limitava fortemente la diffusione; la vera rivoluzione, in questo settore, si

1. Tecnologie di rete e paradigmi correlati

ebbe con lo sviluppo delle tecnologie per il trasporto e la distribuzione di energia elettrica su larga scala: questo processo evolutivo ha portato oggi ad una disponibilità capillare di energia elettrica affidabile ed a basso costo. Allo stesso modo oggi esistono, localizzate in tutto il globo, risorse di calcolo ed applicazioni che lavorano su di esse; tuttavia esse sono separate fra loro, disomogenee e spesso confinate alla soluzione di problemi specifici: potremmo infatti parlare di risorse computazionali "embedded". Una tecnologia atta alla condivisione di tali risorse, permetterebbe quindi di renderle disponibili, a chi necessita, o semplicemente a coloro che occasionalmente richiedano certe capacità, senza farsi carico della notevole spesa che comporterebbe una loro installazione in loco.

1.1.3 Il Cluster

Con il termine "Cluster" s'intende una particolare configurazione hardware e software avente per obiettivo quello di fornire all'utente un insieme di risorse computazionali estremamente potenti ed affidabili. Con l'avvento dei servizi di rete Internet e del web in particolare, i cluster sono cresciuti sempre più in popolarità, proponendosi come sistemi altamente scalabili ed efficienti.

Quando siti web o server che offrono servizi vengono acceduti da molti utenti, si può prevedere l'utilizzo di un cluster al fine di distribuire il carico complessivo sui diversi computer o per garantire la continuità del servizio a fronte dell'indisponibilità di uno dei nodi. In generale questo tipo d'organizzazione è assolutamente trasparente all'utente finale; ad esempio vengono adottate soluzioni per cui all'indirizzo web del sito in questione corrisponderanno fisicamente i diversi nodi del cluster.

Si può parlare propriamente di cluster quando un insieme di computer completi ed interconnessi ha le seguenti proprietà:

- I vari computer appaiono all'utente come una singola risorsa computazionale;
- Le varie componenti sono risorse dedicate al funzionamento dell'insieme.

Diverse motivazioni hanno contribuito alla crescente affermazione dei sistemi distribuiti:

- Un ambiente distribuito è scalabile, e può essere facilmente incrementato per venire incontro a nuove esigenze computazionali;
- La richiesta di risorse computazionali può essere distribuita fra un insieme di macchine invece di essere confinata ad un singolo sistema, eliminando i cosiddetti "colli di bottiglia";
- La disponibilità delle risorse e dei servizi viene fortemente incrementata.

1. Tecnologie di rete e paradigmi correlati

Di contro, è possibile individuare alcune problematiche che scaturiscono dalla scelta di adottare un ambiente distribuito:

- L'amministrazione di decine o centinaia di sistemi è molto più onerosa di quella del singolo sistema;
- La gestione della sicurezza cresce esponenzialmente all'aumentare degli host presenti nella rete;
- La gestione di risorse condivise via rete, è più complessa rispetto a quella locale.

Al fine di limitare l'effetto delle problematiche legate all'adozione di sistemi distribuiti e favorire la diffusione di sistemi di calcolo ad alte prestazioni implementati attraverso limitate risorse economiche è stato introdotto, dapprima in ambito accademico e poi anche in ambito aziendale, il concetto di cluster di personal computer. Questo ha fatto anche sì che il sistema operativo Linux giocasse un ruolo chiave per la definizione e l'implementazione di modelli di clustering, basati su tale concetto, atti ad affrontare e risolvere le diverse esigenze specifiche ed i vari macro-problemi.

Ovviamente la topologia del cluster e soprattutto la sua architettura vanno selezionate in base alle particolari esigenze degli utilizzatori ed al particolare contesto in cui si viene ad operare.

Si possono quindi individuare almeno quattro categorie distinte:

- Cluster per l'affidabilità dei servizi;
- Cluster per l'alta disponibilità dei servizi (High Availability, HA);
- Cluster per il bilanciamento del carico (Load Balancing, LB);
- Cluster per il calcolo parallelo.

Cluster per l'affidabilità dei servizi

Tale soluzione prevede la predisposizione di un insieme di sistemi che devono mantenere copie separate dei servizi e che rispondano comunque ad un unico nome di host virtuale con il quale i client interrogano il singolo servizio.

1. Tecnologie di rete e paradigmi correlati

Questo permette di avere una scalabilità dei servizi a costi irrisori rispetto ad un eventuale cambio architetturale (ad esempio da low a high level server).

Cluster per l'alta disponibilità dei servizi

Un cluster per l'alta disponibilità è costituito essenzialmente da due o più server gemelli dove meccanismi hardware e software permettono di far in modo che se il nodo principale del cluster HA si blocca, il nodo secondario, fino a quel momento in attesa, si riconfiguri per apparire come il nodo principale del cluster, con un'immagine congrua al nodo originale immediatamente prima della sua indisponibilità (ovvero senza perdite di dati sensibili).

Cluster per il bilanciamento del carico

Un load-balancing cluster può essere un incomparabile strumento di produttività. L'idea (che estende il modello di affidabilità) è quella di introdurre un sottosistema di schedulazione delle richieste applicative, ad esempio un sistema di code che sia in grado di reindirizzare la richiesta sul nodo più scarico, in grado di far fronte alla richiesta utente.

Cluster per il calcolo parallelo

La ragione principale per cui molti sforzi si sono concentrati verso lo sviluppo di architetture cluster per il calcolo parallelo basate su PC sono i costi. In questo modo, invece di rincorrere a colpi di svariati milioni di dollari l'ultimo supercomputer, si è sempre più affermata l'idea di assemblare grandi quantità di processori classe PC con una struttura di comunicazione a banda alta e bassa latenza. In alcuni casi viene utilizzato un meccanismo di Direct Memory Access (DMA) fra i nodi, fornendo una sorta di distributed shared memory che può essere acceduta direttamente da ogni processore su ogni nodo. Inoltre, è previsto un livello di comunicazione a scambio di messaggi (message passing) per la sincronizzazione dei nodi.

1.1.4 Le applicazioni delle Griglie

Gli ambienti operativi che presumibilmente utilizzeranno appieno il Grid Computing possono essere suddivisi in cinque classi.

Supercalcolo distribuito

Questo tipo di applicazioni possono utilizzare Grid per unificare le risorse di calcolo di alcuni, o molti, supercomputer al fine di risolvere problemi che, altrimenti, sarebbe oggi proibitivo risolvere su di un unico calcolatore, per quanto potente. Un ambito in cui il supercalcolo distribuito può portare notevoli innovazioni è quello scientifico; una simulazione accurata di un fenomeno spesso richiede risoluzioni spaziali e temporali molto elevate; in queste situazioni, utilizzare insieme diversi supercomputer o veri e propri insiemi di cluster può portare a superare le limitazioni esistenti.

High Throughput Computing

In questo tipo di applicazioni, Grid può essere utilizzato per organizzare il lavoro di un grande numero di programmi, che siano tra loro poco o per nulla collegati; lo scopo di questi sistemi è, in genere, quello di sfruttare il tempo macchina inutilizzato, ad esempio durante le ore notturne. Un esempio di software adatto ad implementare architetture di questo tipo è Condor, sviluppato dall'Università del Wisconsin, utilizzato per gestire pool di centinaia di workstation in università ed enti di ricerca di tutto il mondo.

On Demand Computing

Le applicazioni on demand (a richiesta) utilizzano Grid per rendere disponibili risorse di calcolo che, per il loro saltuario utilizzo, non sarebbe conveniente avere sempre a disposizione. Queste risorse possono essere processori, software, archivi di dati, strumenti molto specializzati e così via. Al contrario del supercalcolo distribuito, queste applicazioni sono spesso motivate dal rapporto costo-prestazioni, piuttosto che dalla necessità di performance. Applicazioni di questo tipo, come programmi di

1. Tecnologie di rete e paradigmi correlati

analisi numerica o acquisizione in tempo reale di immagini, rappresentano notevoli sfide per Grid, in quanto sono richiesti livelli di flessibilità e complessità che obbligano a trattare con molta cura le problematiche relative all'allocazione di risorse, gestione delle code, affidabilità e sicurezza.

Data Intensive Computing

Nelle applicazioni di questo tipo, il cui scopo è la gestione di enormi quantità di dati distribuiti geograficamente (diversi Terabyte di dati generati quotidianamente), si riscontrano spesso problemi legati all'alto carico computazionale ed alle modalità di trasferimento di tali moli di dati. Le sfide principali di queste applicazioni sono legate alle problematiche di scheduling ed alla gestione di flussi di dati numerosi e complessi.

Calcolo collaborativo

Questo tipo di applicazioni mirano soprattutto a favorire le comunicazioni e le collaborazioni tra le persone fisiche, pertanto sono spesso pensate in termini di spazi virtuali. Molte di queste applicazioni devono rendere disponibili risorse di calcolo condivise, ad esempio archivi di dati, e pertanto condividono tutti gli aspetti delle applicazioni già analizzate; in questo contesto, ciò che assume particolare rilievo è la necessità di fornire queste funzionalità in tempo reale.

1.1.5 La comunità delle Griglie

Dall'analisi svolta, si è evidenziato che, almeno in un primo momento, le categorie di persone che maggiormente saranno interessate all'utilizzo e allo sviluppo delle Griglie, saranno scienziati, tecnici, ingegneri e, comunque, tutte quelle categorie di persone direttamente interessate alle ricadute tecnologiche della scienza. All'esterno del mondo scientifico Grid si prospetta di notevole interesse anche per grosse multinazionali, che potranno beneficiare della conseguente capacità di aggregazione, oltre che scuole e università che potranno condividere sia archivi che risorse multimediali, o accedere in modo proficuo alle diverse banche dati.

Bisogna ricordare tra i beneficiari anche gli organi governativi dei vari stati, che potranno in futuro usufruire di Grid per l'integrazione di strumenti atti a gestire l'intera "cosa pubblica", come ad esempio la manutenzione di archivi catastali elettronici a larga diffusione, registri automobilistici o dati fiscali distribuiti lungo il territorio nazionale.

È anche interessante sapere che il management della Sony ha recentemente reso noto il suo obiettivo di rendere possibile la connessione in Griglia della prossima console della serie PlayStation: le Griglie Computazionali entrerebbero, dunque, di gran carriera, nelle case dei singoli utilizzatori.

Si avvicina un futuro in cui, come per l'energia elettrica, l'utente si farà carico solo dell'utilizzazione della potenza di calcolo, ma non della sua fornitura.

1.2 Peer-to-Peer e Griglie Computazionali

Reti Peer-to-Peer e Griglie si rifanno a modelli di computazione distribuita che introducono collaborazione decentralizzata e integrazione di computer in rete nel quale ogni utente può fornire o usufruire di servizi.

Il Peer-to-Peer (P2P) identifica una classe di sistemi autoorganizzati o applicazioni che ottengono un vantaggio dalla distribuzione delle risorse: spazio dati, informazioni, cicli di processamento e presenza umana, disponibili nella rete Internet. Una Griglia è una piattaforma geograficamente distribuita composta di un insieme di macchine eterogenee cui gli utenti possono accedere tramite una singola interfaccia. Oggi P2P e Griglie rappresentano due temi importanti della ricerca perché risultano essere promettenti paradigmi per lo sviluppo di efficienti sistemi distribuiti ed applicazioni [10].

Entrambi i modelli si discostano dalla classica configurazione client/server, dove i ruoli del client e del server sono ben separati.

Infatti, in un tale sistema, un insieme di potenziali macchine server offre informazioni e servizi generalmente ad un numero molto elevato di client. Invece, P2P e Griglie possono assegnare ad ogni nodo, eventualmente su livelli differenti, entrambi i ruoli secondo le operazioni che si intende fornire alla rete.

Il P2P ha attirato su di se un'enorme attenzione ed ha guadagnato popolarità negli ultimi anni grazie al supporto di due principali classi di applicazione:

- Applicazioni file sharing, nelle quali un peer può condividere le sue risorse con gli altri partecipanti del sistema. (Gnutella, Napster)
- Applicazioni di computing parallelo ad alte prestazioni, nelle quali un'applicazione parallela è in esecuzione su i nodi disponibili.

Separatamente da questi sistemi conosciuti, i modelli P2P emergono come il nuovo approccio ai paradigmi di computazione

1. Tecnologie di rete e paradigmi correlati

distribuita per la loro potenziale capacità di racchiudere la computazione, l'efficienza nell'aggregazione di dati ed il potere comunicativo tra gli hosts che compongono la rete, oltre a rendere disponibili le risorse di un utente a tutti i partecipanti.

Questo sistema di elaborazione condivide il primato assieme alle Griglie principalmente per la possibilità di accesso a risorse remote che non potrebbero esistere in ambienti domestici poiché prerogativa di applicazioni ad alte prestazioni e data-intensive.

Le Griglie computazionali [11] sono nate a supporto della creazione di ambienti per il computing integrato nel quale organizzazioni distribuite possono condividere dati, programmi e comunicare tra nodi implementando servizi decentralizzati.

Sebbene queste siano state introdotte originariamente per applicazioni scientifiche avanzate, il Grid computing è emerso come un paradigma per la coordinazione e la condivisione di risorse e problemi con la possibilità di una risoluzione dinamica, per l'organizzazione virtuale di industrie, al fine di migliorare i loro piani di business. Può, infatti, esser visto come la risposta agli inconvenienti dei sistemi a controllo centralizzato di tipo client/server come l'overloading, i failure, e la bassa Quality-of-Service (QoS).

I nodi di una Griglia tipicamente rendono disponibili le loro risorse e, allo stesso tempo, possono avere accesso a risorse disponibili dagli altri nodi. Perciò il modello di Griglia rimuove la distinzione tra macchine client e macchine server e, la corrente implementazione richiede che i nodi stessi abbiano una maggiore responsabilità per le funzioni specifiche ai quali sono delegati.

P2P comprende un grosso numero di applicazioni, progettate per soddisfare differenti richieste, fra le quali l'anonimità' (tipica del file sharing), scalabilità (tipica delle applicazioni ad alto calcolo parallelo), disponibilità (in entrambi i casi).

1. Tecnologie di rete e paradigmi correlati

A dispetto di ciò esistono numerose varietà di sistemi P2P: sistemi con indice delle risorse centralizzato (Napster), ricerche basate su flooding (Gnutella), strutture con un numero di indici di risorse distribuito (ancora in fase di sperimentazione, come Gridella) e reti ibride nel quale P2P e reti strutturate di tipo Client-Server sono combinate tra loro (come nei modelli Super-Peer [5]).

Analizzando entrambi i modelli possiamo scoprire che le Griglie sono in realtà dei sistemi P2P. Ma, in importanti aspetti gli odierni sistemi di elaborazione sono basati su una stratificazione gerarchica e, i prossimi obiettivi della ricerca consistono appunto nel cercare abolire questa forma implementativa.

Il problema maggiore risiede, infatti, nella scalabilità del numero di partecipanti del sistema che può crescere da decine a migliaia di nodi, e proprio in questo l'approccio P2P può aiutare ad effettuare la scalata verso questi sistemi, realizzando modelli di computazione in grado di essere gestiti in maniera distribuita perfettamente funzionale e stabile.

Verranno analizzate perciò le principali similitudini e differenze tra i due modelli:

a. Sicurezza: È un tema centrale nelle Griglie e la maggior parte degli sforzi di sviluppo odierni sono diretti verso la creazione e l'integrazione di meccanismi di autenticazione, autorizzazione, integrità e confidenzialità della piattaforma. I sistemi P2P, d'altro canto, nati nel contesto della "Open Community", nella quale ogni utente condivide più aspetti generici piuttosto che obiettivi specifici, non condividono la precedente filosofia, i meccanismi d'accesso, infatti, non ammettono autenticazione di indirizzi e validazione dei contenuti, ma offrono protocolli che assicurano anonimità e resistenza alle censure.

b. Connettività: oggi i sistemi di Griglia sono generalmente composti di risorse utilizzabili per la ricerca dalle istituzioni governative o dalle grosse imprese. Esse comprendono

1. Tecnologie di rete e paradigmi correlati

per lo più macchine abbastanza potenti, connesse in maniera statica attraverso reti ad alte prestazioni con un alto livello di disponibilità. I sistemi P2P invece, possono essere composti di semplici computer domestici connessi ad intermittenza alla rete e disponibili perciò per un limitato ammontare di tempo con un ridotto grado di affidabilità.

c. Servizi di accesso: L'accesso alle risorse remote è la principale motivazione per la costruzione delle Griglie ed è ancora oggi uno dei suoi punti di forza. I toolkit di architettura provvedono a fornire servizi sicuri per l'invio di processi o l'esecuzione di applicazioni interattive su macchine remote oltre che a meccanismi per condivisioni efficienti e spostamenti di dati tra i nodi. I correnti meccanismi P2P non supportano ad oggi metodi di allocazione esplicita di cicli di computazione e data storage, ma offrono protocolli per la condivisione e lo scambio di dati tra nodi.

d. Scoperta delle risorse e gestione della presenza: La scoperta delle risorse nei sistemi di Griglia è oggi principalmente basata su approcci centralizzati o gerarchici. Nel Globus Toolkit per esempio, le informazioni circa le risorse di ogni singolo nodo possono essere ottenute tramite la chiamata diretta all'applicazione server che al momento è in esecuzione su quel nodo, o tramite la richiesta ad un apposito server che si occupa del recupero e della pubblicazione delle informazioni per un dato insieme di nodi. Nei sistemi P2P invece, il meccanismo per la gestione della presenza è un elemento chiave dell'intera architettura: ogni nodo periodicamente notifica la sua presenza nella rete e allo stesso tempo scopre i suoi vicini. Differenti modelli usano diversi metodi per la scoperta: index server centralizzati, semplici meccanismi di flooding, o tipologie e algoritmi più complessi.

e. Tolleranza ai guasti: La natura dinamica delle Griglie insegna che è necessario un certo livello di tolleranza ai guasti, specialmente per casi in cui si utilizza codice altamente distribuito,

1. Tecnologie di rete e paradigmi correlati

perché potrebbero venire a crearsi un alto numero di processi indipendenti simili su più nodi. Il progetto della tolleranza ai guasti e le politiche utilizzate in Griglia considerando l'approccio decentralizzato degli algoritmi P2P può rappresentare uno sbocco per raggiungere altri livelli di affidabilità.

È necessario così definire le opportunità di integrazione tra P2P e Griglie ponendo l'attenzione su due principali ed importanti aspetti:

- Implementazione di Servizi di Griglia in accordo ai modelli P2P.
- Progetto e sviluppo di applicazioni di Griglia usando un modello P2P nell'accesso e la coordinazione delle risorse e dei servizi.

Recentemente la Grid community ha iniziato lo sviluppo di nuove tecnologie al fine di utilizzare le Griglie assieme ai Web services [12], utilizzando l'Open Grid Services Architectures (OGSA) [13]. Questo abilita l'integrazione di servizi e risorse attraverso ambienti distribuiti, eterogenei, dinamici e comunitari. Al fine di garantire questa integrazione, il modello OGSA adotta il Web Services Description Language (WDSL) [14] e definisce i concetti dei servizi di Griglia. OGSA definisce questi concetti basati su principi e tecnologie di entrambe le comunità. I Web Services e OGSA tentano di stabilire un'interoperabilità tra sistemi e servizi indipendenti debolmente accoppiati per implementazione, locazione e piattaforma. Vengono definiti infatti meccanismi standard per la creazione, il naming e la scoperta di istanze di servizi di Griglia transienti e persistenti, provvede alla trasparenza della locazione ed al bindings per istanze di servizi multiprotocollo oltre che, all'integrazione di supporti con particolari agevolazioni per le piattaforme native.

1.3 Web Services

Una Griglia computazionale è un insieme di componenti software che (a meno di dovute eccezioni), implementa il meccanismo dei Web Services per la creazione di sistemi distribuiti.

Un Web Service provvede a fornire un approccio per interoperare tra differenti applicazioni software in esecuzione su una varietà di piattaforme e/o Frameworks [11].

In sintesi è un sistema software progettato per supportare interazioni ed interoperabilità tra macchine sulla rete. Ha, infatti, un'interfaccia scritta in un formato processabile (più precisamente WSDL). Gli altri sistemi interagiscono con esso nella maniera prestabilita dalla sua descrizione utilizzando messaggi SOAP, tipicamente utilizzando HTTP ed una serializzazione XML in unione con gli altri standard web correlati [15].

I Web services, standardizzano i messaggi che le entità nei sistemi distribuiti possono scambiare al fine di realizzare svariate operazioni. Al livello più basso, la standardizzazione concerne il protocollo usato per il trasporto dei messaggi (di solito HTTP), la codifica (SOAP) e la descrizione dell'interfaccia (WSDL). Un client interagisce con il Web Services inviando a lui un messaggio SOAP; successivamente questo può ricevere in risposta una conferma di ricezione. Al livello più alto, altre definizioni stabiliscono convenzioni sulla sicurezza dello scambio dei messaggi, per la gestione e per le funzioni di alto livello come la scoperta e la gestione delle risorse.

Le tecnologie dei Web Services e delle Griglie in particolare, possono essere usate per costruire sia applicazioni orientate ai servizi che vere e proprie infrastrutture.

Dalla prospettiva del client, un Web Services [12] è semplicemente un'entità accessibile dalla rete che processa

1. Tecnologie di rete e paradigmi correlati

messaggi SOAP. Al fine di esplicitare la tipologia d'implementazione del servizio è necessario distinguere tra :

- L'hosting environment (o container), cioè il dominio logico usato per ricevere i messaggi SOAP e identificare e invocare il codice appropriato per maneggiare i messaggi, e potenzialmente per poter gestire le funzioni di amministrazione.

- L'implementazione del Web Service, cioè lo specifico codice per gestire i messaggi.

Questa separazione di concetti sta a significare che lo sviluppatore ha solo bisogno di ottenere l'opportuno codice per la manipolazione dei messaggi per implementare nuovi servizi.

È anche comune la divisione tra la logica dell'hosting environment da quello che concerne il trasporto dei messaggi SOAP (tipicamente utilizzando il protocollo HTTP, coadiuvato da un apposito application server) e quello che concerne il processamento dei messaggi (tipicamente ad opera di un SOAP engine).

Esistono in realtà più container differenti, con diverse proprietà e prestazioni, per il supporto di altri linguaggi di implementazione dei Web Services, sicurezza e molto altro.

Verranno perciò spiegate esclusivamente le caratteristiche tecniche che particolarizzano gli ambienti di Griglia. Questo comprende le specificazioni tecniche che definiscono l'architettura dei Web Services (XML, SOAP, WDSL, ecc).

XML è usato estensivamente con i WS come un formato dati standard, flessibile ed estensibile. In aggiunta alla sintassi XML, un'importante caratteristica di questo linguaggio sono l'XML-Schema [16] ed i Namespaces [17], che possono introdurre un'elevata efficienza nelle prestazioni dell'attuale tecnologia implementativa in quanto i WS tools odierni adottano serializzazione testuale.

SOAP [18] fornisce un Framework standard, flessibile ed estensibile per l'impacchettamento e lo scambio di messaggi XML

1. Tecnologie di rete e paradigmi correlati

tra fornitori e utilizzatori di servizio. SOAP è indipendente dalla tipologia del protocollo di trasporto, ma è comunemente utilizzato su HTTP.

WSDL [14] è un documento XML per la descrizione di Web Service. Convenzioni standardizzate hanno definito i modi di utilizzo di WSDL in congiunzione con SOAP e gli altri substrati del sistema.

Un'interfaccia può essere compilata per generare un proxy code che permette la costruzione di messaggi e gestisce la comunicazione per conto del client. Il proxy mappa automaticamente la struttura del messaggio XML in oggetti di linguaggio nativo che possono essere direttamente manipolati dall'applicazione, un ulteriore vantaggio del proxy è che questo esula il programmatore dalla comprensione e lo sviluppo di file XML.

La famiglia dei WS-Security specifica un range di funzioni relative all'autenticazione, autorizzazione e alle politiche di rappresentazione dei contesti su cui agiscono i WS. Il SAML (Security Authorization Markup Language) [19] è utilizzato per la protezione degli indirizzi nei messaggi, autenticazione, delegazione e autorizzazione [20].

Un certo numero di specifiche provvedono ad offrire funzionalità importanti orientate alle infrastrutture nelle quali possiamo avere una certa abilità nel rappresentare e manipolare entità con stato come risorse fisiche di vario tipo, componenti logiche come licenze software e attività transienti con processi o workflow.

Le specifiche del WS-Addressing [21] definiscono meccanismi di trasporto neutrale per indirizzare Web Service [22, 12] e messaggi. Nello specifico, esso definisce elementi XML per identificare endpoints di servizi Web e per assicurare la corretta identificazione dei destinatari/mittenti nei messaggi.

1. Tecnologie di rete e paradigmi correlati

Le specifiche del Web Service Resource Framework (WSRF) [12] definiscono un framework generico e aperto per modellare e accedere alle risorse usando i WS. Questo framework comprende i meccanismi per descrivere viste sullo stato delle risorse (WS-Resource properties), al fine di supportare la gestione dello stato attraverso le proprietà associate con i Web Services (WS-Resource Lifetime), per descrivere come questi meccanismi sono estensibili a gruppi di WS (WS-Service Group) e come reagire agli errori di sistema (WS-Base Faults).

La famiglia del WS-notification definisce un modello di approccio per permettere ai WS di disseminare informazioni fra di essi [23]. Questo framework comprende meccanismi per la notificazione base (WS-Notificacion), la notificazione Topic-Based (WS-Topics) oltre che quella fra agenti del sistema (WS- Brokered Notification).

Il gruppo di lavoro Open Grid Services Architectures (OGSA) [13] ha completato un documento che provvede ad una descrizione di alto livello delle funzionalità richieste per le future applicazioni e infrastrutture orientate ai servizi, ed un framework che definisce come queste funzionalità possono essere utilizzate. Questo gruppo di lavoro sta inoltre definendo una serie di Profili come il WS-I Profiles per identificare le specifiche tecniche che possono essere usate tramite i WS nello scenario di Griglia.

1.4 Swarm Intelligence e Sistemi ad Agenti

In uno sciame di insetti, l'interazione tra gli elementi del gruppo in tutte le attività individuali non sembra richiedere alcun tipo di supervisione, ma, preso nell'insieme, tutto il gruppo risulta essere ben organizzato.

Un insetto è una creatura complessa, sa riconoscere e utilizzare un buon numero di stimoli sensoriali, modificare il proprio comportamento come conseguenza degli stimoli subiti, interagire con i compagni di colonia e prendere decisioni sulla base di un largo numero di informazioni acquisite.

Ma la complessità di un singolo insetto non è sufficiente per spiegare l'alto livello di strutturazione che riesce a raggiungere il lavoro di una colonia di insetti sociali.

Molti aspetti delle attività collettive degli insetti sociali sono basati sull'auto organizzazione.

La teoria dell'auto organizzazione (Self Organization - SO) [3], sviluppata originariamente nel contesto della fisica e della chimica per descrivere l'affiorare di modelli macroscopici al di fuori di processi e interazioni definite a livello microscopico, può essere estesa agli insetti sociali per dimostrare che un complesso comportamento generale può emergere dall'interazione tra individui che hanno fondamentalmente un semplice modello comportamentale.

Recenti ricerche hanno verificato che l'auto organizzazione è veramente la componente maggiore di una vasta scala di fenomeni collettivi negli insetti sociali.

I modelli basati su SO non precludono complessità individuale: essi mostrano che, a qualsiasi livello di descrizione è possibile spiegare la complessità del comportamento collettivo assumendo che gli insetti sono entità che compiono interazioni relativamente semplici.

1. Tecnologie di rete e paradigmi correlati

Tutto ciò è logicamente corretto: si assume, infatti, che è sempre possibile spiegare qualcosa di apparentemente complesso in termini di interazioni tra semplici processi.

La scoperta che la teoria dell'auto organizzazione può essere applicata al lavoro degli insetti sociali ha non solo conseguenze sullo studio degli stessi ma ci fornisce uno strumento molto potente per trasferire queste conoscenze nel campo della progettazione di sistemi intelligenti.

La modellazione del comportamento degli insetti per mezzo dell'auto organizzazione può aiutare nella creazione di congegni artificiali per la soluzione di problemi distribuiti che in maniera autonoma riesce a risolvere il dato problema (Swarm Intelligence System).

Ma i sistemi basati sull'uso delle intelligenze di massa sono difficili da programmare, richiedendo una buona conoscenza non solo di come il comportamento individuale va implementato, ma anche di come le interazioni tra entità sono necessarie per produrre un corretto comportamento collettivo.

Questo approccio è simile agli altri che imitano il modo in cui la natura (intesa come sistema fisico o biologico) risolve i suoi problemi.

L'espressione Swarm Intelligence (SI) fu usata per la prima volta da Beni, Hackwood, e Wang [24] nel contesto dei sistemi per la robotica cellulare dove, molti semplici agenti occupando uno o due ambienti dimensionali riuscivano a generare modelli e ad auto organizzarsi tra loro attraverso i rapporti di vicinanza.

Usare l'espressione SI solo per rappresentare il suddetto tipo di lavoro è chiaramente restrittivo: è perciò che si estende la definizione per includere ogni tentativo di realizzazione di congegni per la soluzione di problemi distribuiti o algoritmi prendendo spunto dal comportamento collettivo delle colonie di insetti sociali e degli altri animali che fanno vita sociale.

1. Tecnologie di rete e paradigmi correlati

La modellazione è molto differente dalla progettazione di un sistema artificiale, perché nella modellazione si cerca di scoprire cosa realmente accade nei sistemi naturali.

Il modello deve dare una rappresentazione testabile, e idealmente tutte le variabili ed i parametri dovrebbero risultare accessibili dagli esperimenti.

L'indiscutibile successo degli insetti sociali può essere utile come punto di partenza per nuove metafore ingegneristiche e della Computer Science.

L'auto collaborazione è un insieme di meccanismi dinamici dal quale le strutture appaiono al livello globale del sistema attraverso l'interazione fra i suoi componenti di basso livello.

Le regole specificando le interazioni tra le unità costituenti il sistema, sono eseguite sulla base di pure informazioni locali, senza riferimenti a disegni globali, il quale risulta essere una proprietà in via di sviluppo del sistema piuttosto che una proprietà imposta al di sopra dello stesso da un ordine esterno.

La SO è basata su quattro principi basilari :

a. Feedback positivo (Amplificazione), è una semplice regola comportamentale che promuove la costruzione di strutture. Quando un ape trova una fonte di nettare, torna subito indietro all'alveare e avvisa le altre api circa la scoperta appena avvenuta. Dopo di che può agire in più modi: inizia a volare nella direzione della fonte di nettare in modo da indirizzare le altre api nella direzione e sulla giusta distanza dalla fonte del cibo o continua a presiedere nella zona appena scoperta senza avvisare della scoperta l'alveare o ancora può decidere di abbandonare la sua fonte di cibo e diventare essa stessa un'ape esploratrice delle altre risorse scoperte. Se la colonia offre due identiche fonti di cibo collocate alla stessa distanza dall'alveare, le api esplorano le due fonti simmetricamente. A quel punto se una delle due fonti è migliore dell'altra, tutte le api si spostano sulla fonte prescelta per

1. Tecnologie di rete e paradigmi correlati

sfruttarla o altrimenti tutte si spostano sull'ultima migliore fonte scoperta.

b. Feedback negativo: controbilancia il feedback positivo e aiuta a stabilizzare il disegno collettivo. Può assumere la forma di saturazione, esaurimento o competizione. Nell'esempio della ricerca delle risorse, il feedback negativo tiene sotto controllo il numero limitato di risorse disponibili, la saturazione, l'esaurimento e l'affollamento delle fonti di cibo, o la competizione per la scoperta delle suddette fonti.

c. SO dà assistenza nell'amplificazione delle fluttuazioni (cammini casuali, errati, ecc). La casualità non emerge soltanto nella creazione di strutture, ma a volte risulta spesso cruciale perché permette di scoprire nuove soluzioni, e le fluttuazioni possono agire come un seme da cui la struttura prende forma e riesce a crescere.

d. Tutti i casi di SO sono basati su interazioni multiple. Un singolo individuo può generare una struttura auto organizzata utilizzando una sorta di traccia che genera un sentiero. Questo è possibile in natura tramite il feromone, una sostanza chimica usata dagli animali per comunicare che ha ovviamente un certo tempo limitato di efficacia, necessario per realizzare lo scopo da perseguire. Le strutture generalmente richiedono un insieme minimo di individui mutuamente tolleranti. Questi individui devono essere capaci di utilizzare al meglio le proprie capacità per perseguire i loro compiti e soprattutto devono saper interpretare i risultati ed i segni che gli altri individui ottengono. Ad esempio, grazie all'utilizzo del feromone, una colonia può seguire le tracce lasciate dagli individui del proprio gruppo e sviluppare il proprio lavoro secondo la situazione che gli si presenta.

Quando una SO è basata su utilizzo di feromone si caratterizza solitamente da poche proprietà chiave:

1. Tecnologie di rete e paradigmi correlati

I. La creazione di strutture spazio temporali inizialmente di omogeneità media. Queste strutture includono l'architettura del nido, i cammini da seguire per la scoperta delle risorse e l'organizzazione sociale.

II. La possibile coesistenza di più condizioni di stabilità. Visto che le strutture emergono dall'amplificazione di deviazioni casuali, ogni deviazione può essere amplificata, facendo convergere il sistema verso uno dei possibili stati stabili, dipendendo ovviamente dalle condizioni iniziali. Per esempio, quando due identiche fonti di cibo, A e B vengono scoperte alla stessa distanza dal nido da una colonia di formiche che ricorrono alla ricerca di massa, una delle due verrà pesantemente sfruttata mentre l'altra sarà abbandonata: entrambe le fonti avevano la stessa probabilità di venire utilizzate ma solo una delle due viene scelta, ed è la colonia di formiche che opera la scelta.

III. L'esistenza di biforcazioni quando alcuni parametri vengono a variare. Il comportamento di un sistema auto organizzato cambia drasticamente quanto ci si trova di fronte alle biforcazioni. L'effetto che viene a generarsi è il cosiddetto effetto "palla di neve", quando, superata una certa barriera o al momento della scelta di una data strategia viene a realizzarsi una netta predominanza di eventi tali da realizzare un comportamento totalmente opposto agli altri proponibili nel dato momento.

L'auto organizzazione negli insetti sociali spesso richiede interazioni fra insetti: queste interazioni possono essere dirette o indirette.

Le interazioni dirette sono le interazioni con scambio di liquidi, contatto mandibolare, contatto visivo e chimico. Le indirette sono concettualmente più fini: due individui interagiscono indirettamente quando uno dei due modifica il comportamento e l'altro risponde al nuovo comportamento in breve tempo.

1. Tecnologie di rete e paradigmi correlati

Questa interazione è un esempio di Stigmergia. La stigmergia è spesso trascurata e non serve a spiegare il dettagliato meccanismo con il quale gli individui coordinano le loro attività.

In ogni caso provvede a spiegare un meccanismo generale che lega i comportamenti degli individui e delle colonie: il comportamento collettivo modifica l'ambiente circostante e questa modifica influenza il comportamento degli altri individui.

Questo è un significativo passo in avanti per progettare gruppi di agenti artificiali con il quale risolvere problemi: rimpiazzando la coordinazione attraverso la comunicazione diretta in comunicazione indiretta è utile per realizzare semplici agenti, riducendo il carico comunicativo fra gli stessi.

Infine, la stigmergia è spesso associata con la ricerca di flessibilità, quando il comportamento cambia, a causa di una perturbazione interna, l'insetto risponde appropriatamente a tale perturbazione come se vi fosse una sorta di modificazione del comportamento da parte dell'attività della colonia.

In altre parole, la colonia può rispondere collettivamente alle perturbazioni modificando negli individui presi singolarmente il loro comportamento.

Avere la dote della flessibilità negli studi con agenti artificiali è una proprietà senza prezzo, significa, infatti, che gli agenti possono rispondere alle eventuali perturbazioni senza dover essere riprogrammati per reagire alla particolare perturbazione verificatasi.

1.5 Clustering e Sorting

Clustering e sorting basati su agenti derivati da insetti sociali (prevalentemente formiche), sono forme euristiche ispirate dalla natura, introdotte inizialmente come modelli per spiegare due tipologie di comportamento emergenti dall'osservazione di colonie di insetti reali [25].

Più recentemente questi concetti sono stati applicati alle tecnologie di data-mining al fine di realizzare clustering e mapping topografico.

La dimensione dei dati raccolti oggi giorno dall'uomo in campi quali la bioinformatica ed il recupero di informazioni da documenti elettronici ha deviato le mete degli studi sulla conoscenza verso una vera e propria svolta.

Le nuove tecnologie, come il recupero di documenti da internet, esperimenti su micro-array in bioinformatica, simulazione scientifica su calcolatori elettronici e molto altro, hanno dato luogo alla nascita di enormi depositi di dati, che sfuggono ormai, date le loro dimensioni, al controllo umano e possono essere gestite solo tramite i computer.

Mentre la sempre crescente capacità di salvataggio dati in memoria e l'incremento delle velocità di processamento ed elaborazione facilitano lo sviluppo di questi filoni scientifici, non bisogna sottovalutare lo sviluppo parallelo delle tecnologie di archiviazione e di sviluppo delle basi di dati e delle tecniche di realizzazione dei sistemi informativi.

L'analisi automatizzata dei dati e la loro rappresentazione sintetica all'uomo sono cruciali in questo contesto in quanto, solo dopo esser stati interpretati dalla mente umana diventano significativi e possono contribuire alla generazione di conoscenza sugli argomenti trattati. Il campo della ricerca che maggiormente si interessa di questa tipologia di lavori si chiama Data-mining ed

1. Tecnologie di rete e paradigmi correlati

alcuni dei problemi fondamentali che si viene a trovare in questo campo sono appunto l'analisi dei cluster e il mapping topologico.

L'analisi dei cluster concerne nella divisione dei dati in gruppi omogenei nel quale ogni caratteristica di un particolare cluster è simile a quelli della stessa tipologia, mentre risultano ovviamente dissimili dalle altre tipologie presenti nel sistema.

Algoritmi di clustering sono stati studiati per molti anni e sono ancora una parte viva della ricerca. Al momento sono disponibili numerose tipologie di algoritmi, differenziati non soltanto nei principi fondamentali quali efficienza e scalabilità, ma anche per altre caratteristiche, come il tipo di attributi che sono in grado di trattare ed elaborare e gli scopi prefissati che ognuno è in grado di raggiungere.

I principali algoritmi per questa tipologia di studi si suddividono in quattro principali classi: metodi di partizionamento, metodi gerarchici, clustering basato su densità e basato su Griglia (Grid-based) [26].

Gli algoritmi che effettuano operazioni di mapping topografico sono collocati un passo avanti rispetto alle tipologie classiche sopra citate.

Essi non sono limitati alla scoperta di gruppi di dati omogenei, ma, puntano a catturare le relazioni di vicinato tramite una visualizzazione bidimensionale (ma se possibile anche tri o pluri dimensionale) di uno spazio dati di grossa entità.

Questo è realizzato tramite la localizzazione di informazioni addizionali sotto forma di relazioni tra cluster individuali e relazioni tra classi di dati presenti all'interno dei cluster. In questo senso, si ottiene una forma di "sorting" dei dati in due o più dimensioni.

In letteratura il mapping topografico è associato anche ad altre simbologie come il neighborhood preserving, mapping topologico o sistematico. Possiamo distinguere tra due importanti tipologie di algoritmi utilizzati per produrre mappe topologiche. Su un fronte

1. Tecnologie di rete e paradigmi correlati

abbiamo procedure basate su calcoli ben fondati teoricamente e matematicamente corretti, i quali sono usualmente basati sulla computazione esplicita di un mapping fra dati e locazione spaziale.

Sfortunatamente questa tecnica non genera necessariamente la migliore rappresentazione visuale-informativa e può spesso richiedere una computazione dispendiosa.

Sull'altro fronte esistono metodi che approssimativamente e gradualmente migliorano i criteri di ottimizzazione (che siano essi impliciti o espliciti). È indubbio che anche questi ultimi algoritmi possono ottenere soluzioni subottimali e venire usati con successo in una moltitudine di applicazioni reali.

Clustering e sorting [27] basati su sistemi ad agenti sono approcci locali e distribuiti che possono venire applicati entrambi in una moltitudine di applicazioni. Sono stati compiuti su essi molti studi circa l'applicazione di questi sistemi su argomenti inerenti l'auto organizzazione delle mappe [28, 29], e si è visto che è possibile ottenere grazie agli agenti, un utilizzo simultaneo di questi processi con buoni risultati [29, 30]. Ma, l'analisi dell'efficacia di questi algoritmi in continuo aggiornamento, non supporta ancora pienamente questa tesi.

Gli esperimenti relativi a questa tipologia di problemi e soprattutto legati all'analisi di dati artificiali riguardano un ambito molto ristretto e una specifica di problemi molto semplici. La valutazione di questi risultati è soprattutto basata su un'osservazione visiva e, solo una piccola parte dei risultati viene ricavata tramite metodi matematici.

Sull'altro fronte invece, questi risultati analitici che ci vengono resi disponibili, tendono a far crescere dubbi concernenti l'efficacia euristica del mapping topografico [31]. Gli esperimenti riportati su dati reali sono limitati e risultano scarsamente sufficienti per valutazioni analitiche esaustive.

1. Tecnologie di rete e paradigmi correlati

Infine, per entrambi i test con dati sia reali che artificiali, non esistono confronti convincenti con metodi alternativi.

Gli aspetti critici da migliorare per ottenere algoritmi effettivamente funzionali e soddisfacenti implicano il cambiamento di alcuni parametri fondamentali.

Innanzitutto migliorie per quanto riguarda la separazione spaziale dei cluster e l'irrobustimento dell'algoritmo per avviare esecuzioni su grosse moli di dati.

Il clustering ed il sorting basati su agenti (formiche), fu introdotto per la prima volta da Deneubourg [6] per spiegare differenti tipologie di fenomeni emergenti che accadevano naturalmente.

Questa branca di studio è una sottofamiglia di quella degli algoritmi sugli agenti, che consiste nello sviluppo di modelli che emulano il comportamento reale di insetti sociali. Nel caso della nostra tipologia di algoritmo, sono stati modellati principalmente due tipi di comportamento esistenti in natura [6]. Sul clustering, si è seguito l'esempio delle formiche che raccolgono materiale a loro utile formando cataste, per esempio nella costruzione di formazioni cimiteriali (aggregazione di corpi morti) osservato nella specie *Pheidole pallidulam*, e sul sorting si è osservato il comportamento di discriminazione fra differenti tipologie di risorse e la loro collocazione spaziale in accordo alle specifiche proprietà.

Questo tipo di comportamento può essere osservato per esempio nei nidi di *Leptothorax unifasciatus* dove le larve sono disposte in differenti locazioni del formicaio secondo la loro taglia.

Nel suo lavoro Deneubourg [6] propone un modello continuo per descrivere questi comportamenti e da ciò riesce a derivare un modello discreto da cui estrarre informazioni ritenute valide sperimentalmente.

Nelle simulazioni informatiche le formiche vengono rappresentate da semplici agenti che vagano con cammini casuali

1. Tecnologie di rete e paradigmi correlati

nell'ambiente virtuale, una Griglia sferica con condizioni periodiche di movimento. I materiali (risorse o indici) che vengono scoperti dal movimento degli agenti nella mappa possono essere presi da questi e trasportati in giro per venire depositati in altre posizioni. Queste operazioni di prelievo e deposito sono basate sulla distribuzione di risorse nelle posizioni di vicinato in cui l'agente si trova. In punti in cui le risorse del vicinato scarseggiano, un'agente sarà maggiormente tentato nel prelevare queste e depositarle in punti dove ne è già presente un tasso più elevato. Il risultato di queste operazioni porta così alla realizzazione di una Griglia clusterizzata.

Il lavoro principale di Deneubourg [6] era principalmente indirizzato per la realizzazione di modelli applicabili alla robotica collettiva, ma, si è riusciti ad applicare le sue teorie anche all'analisi dei dati. Lumer e Faceta [30] hanno invece introdotto un certo numero di modificazioni al modello esistente rendendo possibile la manipolazione di dati numerici e migliorando la qualità delle soluzioni oltre che il tempo di convergenza dell'algoritmo. E per la prima volta appare la definizione di processo di sorting come una primordiale approssimazione di mapping topografico che risulta subito in contrasto con la definizione di clustering puro. Nel sorting le principali misure utilizzate per l'analisi dei risultati e la modellazione del comportamento del sistema risultano essere la densità locale media e l'entropia di Griglia ed è importante notare che queste misure non catturano l'integrità delle relazioni tra cluster di tutto il sistema, ma si riflettono esclusivamente su una scelta limitata all'interno dei cluster.

Successivamente a queste scoperte, al lavoro dei due ricercatori susseguirono un grosso numero di applicazioni reali basate su agenti applicando le metodologie del clustering e del sorting per il mapping topografico, parzialmente motivate dalla notevole scalabilità posseduta dall'algoritmo locale e dalle operazioni distribuite.

1. Tecnologie di rete e paradigmi correlati

Nelle loro ricerche si è utilizzato principalmente una visualizzazione bidimensionale di collezioni di documenti sotto forma di mappe topologiche al fine di esaminare la loro funzionalità tramite il partizionamento grafico nell'ovvia unità di misurazione spazio-dati.