

UNIVERSITÀ DI PISA

DIPARTIMENTO DI INFORMATICA

TECHNICAL REPORT: TR-06-04

La Ricerca su Grid di prossima generazione

Prof. Marco Vanneschi

10 Aprile 2006

ADDRESS: Largo B. Pontecorvo 3, 56127 Pisa, Italy. TEL: +39 050 2212700, FAX: +39 050 2212726

La ricerca su Grid di prossima generazione

Marco Vanneschi

Dipartimento di Informatica, Università di Pisa

Rapporto Tecnico-scientifico del Dipartimento di Informatica, gennaio 2006

1. Contesto scientifico, tecnologico e applicativo

La rapidissima crescita delle innovazioni tecnologiche nel settore delle comunicazioni e dell'informatica diffusa stanno producendo scenari applicativi completamente nuovi, il cui requisito fondamentale di tali scenari applicativi è la *pervasività*, cioè la capacità di far operare e cooperare individui e comunità (industriali, di governo, sociali, scientifiche) “in qualunque momento e in qualunque luogo” (“ubiquitous computing”). Nella realizzazione di applicazioni secondo tali scenari, specie a livello industriale, un ruolo fondamentale riveste il raggiungimento di elevati livelli di *Qualità del Servizio*, specialmente in termini di scalabilità e affidabilità. Ne sono esempi, pur non esaustivi, il controllo ambientale e la prevenzione di disastri utilizzando dati rilevati via satellite e via terra, la progettazione cooperativa di apparati in campo automobilistico e aeronautico, la tele-sanità (“e-Health”), il telelavoro (“e-Work”), la teledidattica (“e-Learning”), la gestione integrata del ciclo di vita di prodotti industriali, applicazioni di e-Business e Supply Chain Management, la gestione della conoscenza per il cittadino, ecc. In tutti questi casi, è richiesta la completa virtualizzazione e l'accesso ubiquo, in modo distribuito e su larga scala, a risorse computazionali, basi di dati e basi di conoscenza, nonché a reti e mobili, sistemi “embedded”, dispositivi specializzati e di identificazione a radio frequenze (RFID), assistenti digitali personalizzati (PDA), “wearable computer”, ecc.

In altri termini, come analizzato in [1], in un contesto che si muove sempre di più verso la Ambient Intelligence, una larga varietà di industrie, non legate tipicamente al mondo ICT, si sta sempre più muovendo verso grosse applicazioni che richiedono notevoli sviluppi di software applicativo, spesso con requisiti stringenti di pervasività, scalabilità ed affidabilità, e che comportano l'uso di risorse distribuite, eterogenee e diversificate.

Per rispondere ai suddetti requisiti, nei prossimi 5-10 anni è necessario lo sviluppo e la messa in opera di sofisticate piattaforme abilitanti complesse *su larga e larghissima scala*, costituite cioè da un elevatissimo numero (dell'ordine dei miliardi) di nodi e di dispositivi eterogenei interagenti e cooperanti, supportate da architetture, strumenti e ambienti software non ancora disponibili e in parte non ancora completamente specificati. Tali sistemi hanno come riferimento il paradigma *Grid* che, abilitando l'accesso, la selezione e l'aggregazione di una varietà di risorse e servizi, realizza una infrastruttura di elaborazione collaborativa che può rappresentare il punto di partenza per la realizzazione di una piattaforma distribuita diffusa capace di supportare le richieste dei nuovi scenari applicativi. Tuttavia, la tecnologia attuale di Grid non è ancora in grado di fornire adeguati ambienti di sviluppo e middleware capaci di supportare le esigenze delle nuove applicazioni in termini di *cooperazione*, *adattatività*, *ubiquità*, *auto-organizzazione*, *interoperabilità*, *gestione semantica della conoscenza*, *sicurezza e privacy*, anche nel rispetto di determinati aspetti sociali, etici

e legali. Inoltre, le attuali infrastrutture Grid (basate su Globus o altri middleware di prima generazione) non sono in grado di soddisfare l'esigenza, fondamentale dal punto di vista dell'applicazione industriale, della *scalabilità e performance* in sistemi costituiti da un numero molto grande di nodi e dispositivi.

Il processo tecnologico e scientifico, già in atto, che vede la convergenza dei paradigmi Grid, Service Oriented Architectures, Web Services, Semantic Web, ambienti di sviluppo software e sistemi embedded distribuiti, è un chiaro indice della tendenza ad affrontare questa complessa e strategica problematica attraverso innovative soluzioni a carattere globale e pervasivo.

In questo contesto un ruolo cruciale è rivestito dallo sviluppo di una adeguata *tecnologia software* e dal rilancio di attività di ricerca nel medio-lungo termine in questo settore dell'Information Technology.

Genericamente parlando, il prodotto della tecnologia software è il *Middleware*, che va inteso come *il complesso* di tutti gli strumenti e servizi interposti tra l'infrastruttura di rete e le applicazioni. Più precisamente, all'interno di questo complesso si riconoscono specifici livelli, ognuno con i propri strumenti e servizi; partendo dal basso (infrastruttura di rete) e procedendo verso l'alto (applicazioni) si riconoscono i seguenti livelli:

- I. Supporti base ("foundations") del Middleware
- II. Servizi del Middleware
- III. Ambiente generale di sviluppo applicazioni
- IV. Problem Solving Environment per domini applicativi specifici.

È ampiamente riconosciuto che la ricerca nel medio-lungo termine deve fornire soluzioni innovative in tutti i suddetti livelli secondo i requisiti indicati precedentemente.

In particolare, un interessante approccio, risultante dall'analisi condotta in [2], consiste nel concentrarsi su *ricerca e sviluppo di prodotti e servizi ad elevato valore aggiunto*, caratterizzati cioè dal raggiungimento e mantenimento di elevata Qualità del Servizio (funzionalità, performance, affidabilità, sicurezza), *basati sull'esistenza di un robusto strato standard comune open-source*. Le caratteristiche dello strato, che viene quindi assunto come un supporto comune a tutte le attività di ricerca e sviluppo, emergerà dall'elaborazione e sperimentazione di gruppi di standardizzazione internazionali, come già da alcuni anni sta avvenendo nel Global Grid Forum ed iniziative europee congiunte.

I prodotti e servizi ad elevato valore aggiunto, oltre che nelle applicazioni, saranno quindi posizionati a tutti i livelli I-IV sopra menzionati.

2. Finalità e caratteristiche della ricerca nel medio-lungo termine

Una caratteristica generale della ricerca su piattaforme Grid su larga scala riguarda la possibilità di definire, progettare e mettere in opera un ambiente robusto e flessibile che abiliti il coordinamento fra i membri di gruppi di lavoro in condizioni di elevata ubiquità e adattività, inclusa la possibilità che alcuni agenti siano in movimento.

Più specificamente, dal punto di vista funzionale e computazionali, è necessario:

- prendere in considerazione la gestione di computazioni che sono tipicamente *dinamiche* e richiedono la modifica di un certo numero di elementi, e tener conto dei servizi *context-aware* che modificano il loro comportamento in base ai parametri di un utente,
- gestire *operazioni parallele, concorrenti e nondeterministiche* su computazioni e strutture dati distribuite o condivise per organizzare l'interazione tra servizi, sorgenti di dati e hardware remoti,
- supportare la possibilità che *nuovi dispositivi e servizi* possono facilmente essere incorporati negli ambienti esistenti ovunque e in qualsiasi momento.
- permettere l'accesso *on demand* a servizi, non solo computazionali, per affrontare lo studio di sistemi applicativi adattivi rispetto alla variabilità di dati correlati dinamicamente agli algoritmi tipici dei vari settori applicativi e relativi modelli matematici.

Dal punto di vista non funzionale e prestazionale, e specificamente in termini del *raggiungimento e mantenimento della Qualità del Servizio*, è necessario:

- permettere di ottenere e controllare dinamicamente *applicazioni ad alte prestazioni* (performance, throughput, tempo di risposta), anche in contesti di elaborazione in tempo reale, sfruttando sia il parallelismo naturale dei sistemi distribuiti e l'evoluzione delle reti ad alta velocità, che l'architettura dei nodi i quali, al loro interno, sono tipicamente strutturati in modo parallelo (dai cluster ai supercalcolatori, dai multiprocessor a memoria condivisa ai dispositivi e sistemi paralleli embedded);
- garantire elevata *affidabilità, disponibilità e tolleranza ai guasti*;
- garantire il desiderato livello di *sicurezza e privacy* nell'accesso alle risorse e nella cooperazione e comunicazione, secondo approcci che degradino il meno possibile il livello di prestazioni.

La sfida consiste nel progettare, realizzare e sperimentare una nuova generazione di piattaforme su larga scala adattive che potranno essere utilizzate in ambito scientifico, industriale, sociale e commerciale, in grado di reagire dinamicamente - in modo scalabile, affidabile e sicuro - a flussi di dati online in diversi campi di applicazione.

A questo scopo, le Grid di nuova generazione dovranno avere la caratteristica di fornire alle applicazioni una interfaccia del tutto amichevole, ad alto livello e completamente indipendente dall'infrastruttura sottostante. La messa in opera di tale concetto, detto anche "*Invisibile Grid*" [3], è assolutamente indispensabile alla luce della complessità dei problemi e dei requisiti delle Grid di prossima generazione.

Come per qualsiasi tipo di piattaforma abilitante, di fondamentale importanza per le piattaforme Grid risulta infatti l'aspetto di "*dominare la complessità*" nella progettazione del sistema e delle applicazioni; la ricerca su Grid di prossima generazione deve quindi concentrarsi su modelli, strumenti ed ambienti

- a) capaci di *nascondere, a utenti e sviluppatori di applicazioni, le caratteristiche delle infrastrutture di Grid* attraverso la completa virtualizzazione delle risorse e dei servizi;
- b) *facili da programmare* e che permettano di ottenere assoluta *interoperabilità e integrazione* tra infrastrutture, strumenti e applicazioni, nonché *riuso* di software esistente, incluso software legacy;

- c) che permettano *lo sviluppo di software di alta qualità, con tempi di progettazione, prototipazione e realizzazione drasticamente ridotti*, almeno dello stesso ordine di grandezza rispetto a quanto l'utenza è finora abituata utilizzando le piattaforme tradizionali;
- d) che permettano il *mantenimento, a tempo di esecuzione, del desiderato livello di Qualità del Servizio*, e che quindi siano caratterizzati da nuovi e potenti controlli dinamici, facendo evolvere tutte le tecniche di configurazione e ottimizzazione, finora applicate staticamente nei sistemi tradizionali.

Solo raggiungendo queste finalità sarà fattibile una reale applicazione delle tecnologie ICT Grid a livello industriale, con un profondo impatto sulla produttività in vari settori strategici.

In questo contesto, è necessario che vengano perseguite, ed adeguatamente sviluppate ed innovate in diversi aspetti specifici, molte metodologie proprie dell'Information Technology per quanto riguarda:

- *modelli computazionali, di parallelismo e di controllo di computazioni decentralizzate,*
 - *modelli di costo per la predizione/valutazione delle prestazioni ed altri aspetti non funzionali,*
 - *modelli, linguaggi e strumenti per lo sviluppo di applicazioni, interoperabilità e riuso,*
 - *modelli e architetture a componenti e/o a servizi,*
 - *definizione della semantica rispetto alla gestione della conoscenza e dei servizi,*
 - *interazione tra modelli di simulazione e di business e modelli computazionali,*
- oltre che metodologie proprie di altri settori interdisciplinari.

3. Tematiche di ricerca sulla tecnologia delle Grid di prossima generazione

Questa sezione è dedicata ad una descrizione delle principali tematiche di ricerca sugli aspetti delle tecnologie Grid nel medio-lungo termine. Oltre alle tematiche tecnologiche, sono ovviamente della massima importanza tematiche di ricerca sulle applicazioni, la cui discussione è sono oggetto di altri capitoli della proposta di piattaforma tecnologica; si veda anche [1, 2, 3].

Come detto nella sezione precedente, l'innovazione, rispetto all'attuale stato della tecnologia Grid, deve riguardare tutti i livelli della piattaforma:

- I. Supporti base ("foundations") del Middleware*
- II. Servizi del Middleware*
- III. Ambiente generale di sviluppo applicazioni*
- IV. Problem Solving Environment (PSE) per domini applicativi specifici.*

3.1. Ambienti e strumenti per lo sviluppo di applicazioni su Grid

Le applicazioni dovranno essere sviluppate, in modo completamente indipendente dall'infrastruttura, utilizzando interfacce ad alto livello, eventualmente specifiche dei domini applicativi, come attraverso PSE (IV). Il livello dell'ambiente generale di sviluppo applicazioni (III) deve permettere sia di progettare PSE, che di progettare e mettere a punto direttamente (quando non siano previsti PSE) applicazioni, utilizzando nuovi modelli e metodologie di servizi e componenti, che costituiscano importanti evoluzioni dell'ingegneria del software verso le piattaforme Grid su larga scala. Le più recenti tendenze nel settore hanno portato a individuare le seguenti linee guida lungo le quali sviluppare la ricerca destinata a produrre risultati disponibili agli utenti e sviluppatori nel medio-lungo termine:

- a) *sviluppo composizionale di applicazioni e servizi* mediante una rigorosa metodologia a componenti distribuite, capace di garantire completa *interoperabilità, integrazione e riuso* di software, anche di tipo legacy;
- b) *modello di componenti dinamiche ad alte prestazioni*, parallele, distribuite e non deterministiche. Questa tematica è di fondamentale importanza in un contesto adattivo e dinamico, in quanto affronta l'aspetto, tutt'ora non risolto in generale, della realizzazione di "versioni" alternative di componenti software (*software versioning*): una stessa componente, quindi con lo stesso insieme di interfacce, può avere diverse realizzazioni tra loro distinte in termini di prestazioni e quindi di qualità del servizio, tipicamente agendo sul grado di parallelismo e/o sulle forme di parallelismo adottate. Le diverse "versioni" possono essere disponibili separatamente e staticamente, oppure essere integrate in una stessa realizzazione adattiva;
- c) *portabilità ed adattabilità, anche dinamica, delle applicazioni, attraverso tutte le principali piattaforme di base che concorrono a realizzare una piattaforma su larga scala*: cluster omogenei e macchine parallele, cluster eterogenei, Virtual Private Networks, piattaforme ad oggetti distribuiti, piattaforme mobile/ubique, fino alle Grid intese nel significato più vasto del termine (organizzazioni virtuali come cooperazione di domini amministrativi distinti), ma capace di prevedere come casi particolari tutte le altre piattaforme. In altri termini, per garantire il desiderato livello di QoS, la stessa applicazione deve poter essere configurata di volta in volta su piattaforme con caratteristiche diverse, a seconda delle esigenze dell'utenza, e questa configurazione può variare anche dinamicamente durante la vita dell'applicazione stessa;
- d) definizione, progettazione e realizzazione di applicazioni ad alte prestazioni secondo *paradigmi cooperativi*, che, oltre ad essere gli unici in pratica utilizzabili per alcune tipologie di applicazioni naturalmente distribuite e/o ubiquitous, rappresentano la base necessaria per la portabilità e l'adattività di cui al punto precedente;
- e) definizione di strumenti automatici basati su *modelli di costo* capaci di supportare le suddette forme di cooperazione, adattività e dinamicità. Occorre ribadire come, in un contesto come quello prefigurato per le piattaforme Grid su larga scala di prossima generazione, sia necessario poter realizzare applicazioni costituite da un elevatissimo numero di componenti, in generale tra loro parallele e cooperanti, e come la loro composizione non possa essere espressa mediante strumenti che presuppongono la conoscenza della piattaforma sottostante, l'allocazione e la scelta delle risorse, il grado di parallelismo, ed altri parametri che non possono essere previsti staticamente;

- f) supporto alla *progettazione di software Grid robusto e di qualità, in termini di affidabilità, disponibilità e tolleranza ai guasti*. Questo aspetto, che è stato solo parzialmente risolto nei sistemi tradizionali a livello di strumenti di sviluppo applicazioni (in generale, nei sistemi attuali esistono solo alcuni meccanismi automatici a livello di sistema operativo), necessita di individuare e sperimentare nuove soluzioni in termini di metodologie di programmazione e di politiche in contesti altamente dinamici, come nel caso di approcci basati su Autonomic Computing;
- g) supporto alla realizzazione di strumenti di *workflow management*, secondo approcci collaborativi, di produzione, orizzontali e verticali.

3.2. Middleware e strumenti per la gestione della conoscenza

Attualmente, le tecnologie Grid hanno dimostrato la loro efficacia per gli aspetti computazionali (“Computational Grid”) e per la gestione di grandi moli di dati (“Data Grid”) in domini scientifici specifici. La comunità internazionale si sta ora concentrando sulla realizzazione di *Knowledge Grid* e *Semantic Grid* al fine di rendere accessibili a tutti tali tecnologie, nonché sulla transizione dalle applicazioni scientifiche a quelle per l’industria e la società con accesso facile, economico e generalizzato. Per potere essere realmente adottate con successo in vari campi applicativi, specie di tipo industriale, queste metodologie necessitano ora di una intensa attività di ricerca e sviluppo nel medio-lungo termine.

- a) Il Middleware deve contenere tutti i servizi essenziali per la gestione delle risorse, identificazione e scoperta, il loro monitoraggio, lo scheduling e co-allocazione, la configurazione e riconfigurazione, la gestione dei dati e metadati e la loro virtualizzazione, lo streaming dei dati in applicazioni ad alte prestazioni, nonché servizi per la tolleranza ai guasti e per l’amministrazione. Tutti questi aspetti, che in qualche misura già fanno parte di infrastrutture Grid, necessitano di essere affrontati in modo completamente nuovo alla luce dei requisiti di dinamicità, adattività, scalabilità ed affidabilità delle Grid di prossima generazione in configurazioni su larga e larghissima scala. In sostanza, il complesso dei servizi del Middleware dovrebbe costituire la macchina astratta del modello con cui vengono sviluppate le applicazioni secondo le caratteristiche suddette (“Grid Abstract Machine”). Questo approccio, da una parte permette di individuare un insieme strettamente indispensabile, e quindi efficiente, di servizi che il Middleware deve possedere, dall’altra rende le applicazioni completamente svincolate dall’infrastruttura e dal modo con cui essa viene gestita: una l’applicazione deve limitarsi a contenere indicazioni di lato livello sulla Qualità del Servizio che si desidera garantire, mentre come garantirla è completamente delegato al Middleware visto come supporto a tempo di esecuzione del modello di programmazione.
- b) Come conseguenza del punto precedente, è necessario individuare approcci caratterizzati da una chiara *semantica* dei servizi e delle modalità con cui sono invocati e forniti e, trattandosi in generale di servizi complessi, delle modalità con cui sono composti, confederati e creati, indipendentemente dai dati attraverso cui la cooperazione ha luogo. La ricerca riguarderà anche *ontologie* per rilevanti aree applicative allo scopo di garantire interoperabilità tra diversi settori.

- c) Nuovi metodi per il Sistema Informativo di Grid, o “Grid Search Engine”, saranno basati su tecniche di *KDD* (Knowledge Discovery in Data Bases) e *data mining*, basati sull’uso di metadati e ontologie per navigare nelle risorse software e hardware di Griglia, come ad esempio nella realizzazione di servizi di Publishing, Browsing, Discovery, e Querying.
- d) Diversi approcci al Data Mining Distribuito, proposti recentemente, necessitano di essere studiati e sperimentati nel contesto di applicazioni per piattaforme Grid su larga scala:
 - ensemble learning,
 - meta-learning framework,
 - collective data mining.
- e) La problematica della gestione della conoscenza può essere accoppiata a quella di individuare e realizzare *modelli economici* per il supporto ad applicazioni Grid secondo una visione che prelude alla costituzione di un “mercato aperto di componenti, servizi e risorse”, definibili anche come “Eco-sistema di componenti software”. Questo potrà essere reso via via disponibile agli sviluppatori di applicazioni Grid, i quali avranno la possibilità di utilizzare sia risorse di calcolo che soluzioni software “preconfezionate” offerte da altri “providers”, soluzioni che potranno essere proposte a differenti prezzi in funzione dei requisiti di Qualità del Servizio richiesti dal utente. Allo scopo, la ricerca deve occuparsi di metodi per individuare i componenti che meglio rispettano i requisiti specificati dall’utente in termini di migliore rapporto prezzo/prestazioni.
- f) La ricerca sui livelli di Middleware deve prevedere anche adeguato spazio per nuovi modelli e strumenti di controllo della *sicurezza*, capaci di superare i limiti delle soluzioni attuali, molto gravose in termini di flessibilità, programmabilità e performance. Sono da individuare e sperimentare nuovi approcci che incorporino i meccanismi di sicurezza dalle politiche di sicurezza, e che quindi si prestino ad una completa integrazione con strumenti di programmazione e di gestione delle risorse in un contesto altamente dinamico e adattivo.

3.3. Middleware e Sistemi Operativi di Grid

Il supporto base (livello I, o “foundation”) del Middleware deve provvedere ad elevare l’interfaccia di ogni sistema operativo a quella richiesta dai Servizi del Middleware (livello II).

Gli attuali sistemi operativi (sia di tipo generale che embedded) non prevedono i servizi necessari per interoperare con i Servizi del Middleware, non tanto a causa del tipo di architettura (esistono ovviamente sistemi operativi per multiprocessor e per sistemi distribuiti), quanto a causa dell’approccio basato sulla “gestione esclusiva” (sicurezza, scheduling) delle risorse di ogni singolo nodo o dispositivo: questo gap deve essere colmato mediante nuove soluzioni appartenenti al livello I del Middleware.

Si tratta di una tematica di ricerca e sviluppo estremamente critica in termini di successo delle prossime generazioni di Grid, in quanto da essa possono, in ultima analisi, dipendere le prestazioni del sistema (performance, affidabilità, sicurezza).

Il livello I del Middleware deve ovviamente poter essere configurato in modo da adattarsi a più sistemi operativi che coesistano nella stessa piattaforma su larga scala, inclusi sistemi embedded e per dispositivi mobili e PDA.

Molte delle problematiche di un sistema operativo per Grid (sia che si tratti di un sistema operativo realizzato ex-novo, oppure che si tratti di più sistemi operativi pre-esistenti e interfacciati via il livello I del Middleware) sono state studiate, prevalentemente, in via teorica, a proposito dei sistemi decentralizzati (sistemi senza alcun punto di centralizzazione e alcuna forma di stato globale). Le Grid di prossima generazione impongono di riprendere tali ricerche in una visione di utilizzazione concreta nel medio-lungo termine. Problematiche critiche da affrontare riguardano

- a) meccanismi e politiche di sincronizzazione e check-pointing,
- b) aggiungimento di adeguate prestazioni circa il tempo di risposta dei servizi, quando a questi sia richiesto anche un elevato livello di affidabilità,
- c) scalabilità in piattaforme su larga e larghissima scala,
- d) supporti base decentralizzati alla sicurezza.

3.4. Modelli e strumenti per la pianificazione, gestione, esecuzione, controllo e manutenzione dei processi applicativi

In diversi domini applicativi occorre sviluppare specifiche politiche adatte allo sfruttamento adeguato delle potenzialità delle piattaforme Grid su larga scala. Nel rispetto dei principi dell'“Invisibile Grid”, tali politiche devono poter essere espresse ad alto livello, senza utilizzare direttamente i livelli del Middleware e del Sistema Operativo. Ad esempio, esse devono poter essere espresse a livello III (ambiente generale di sviluppo applicazioni), oppure costituire il supporto di costrutti di altissimo livello di un *PSE* (livello IV: Problem Solving Environment) orientato allo specifico dominio.

Ad esempio, in applicazioni *e-business* (come nel caso del Supply Chain Management), occorre permettere che, in contesti applicativi altamente dinamici, siano espresse le “richieste degli utenti”, siano identificati potenziali “fornitori virtuali” dei servizi richiesti, questi siano selezionati mediante negoziazione dinamica e in accordo a contratti circa la Qualità del Servizio, sia possibile accedere ai servizi scelti, gestire situazioni di guasto e malfunzionamento, monitorare l'esecuzione, gestire aspetti amministrativi nell'uso delle risorse e servizi.

Questa ampia problematica, con forti ricadute applicative, richiede un forte investimento di ricerca e sviluppo per riuscire a fornire ricadute industriali innovative nel medio-lungo termine.

In questo filone di ricerca si inseriscono i seguenti approcci, tutti da studiare e sperimentare come potenziali soluzioni selezionate per il problema di “dominare la complessità” di applicazioni Grid su larga scala:

- a) metodologie per il supporto all'*adattività negli algoritmi*. In molti dei punti precedenti (ambiente di sviluppo applicazioni, Middleware visto come Grid Abstract Machine dell'ambiente di sviluppo applicazioni) è stato evidenziato l'aspetto dell'adattività di una stessa computazione rispetto alla disponibilità delle risorse allo scopo di garantire il desiderato livello di Qualità del Servizio; in questo caso l'adattività è quindi implementata *nel supporto* del modello di programmazione secondo politiche invisibili al programmatore. Un altro

approccio (eventualmente coesistente con quello ora ricordato) consiste nella progettazione di algoritmi adattivi, cioè di algoritmi nei quali il programmatore esprime *esplicitamente* il comportamento adattivo in funzione di determinati eventi sotto il suo controllo, sempre utilizzando un approccio ad lato livello indipendente dalla infrastruttura. Si tratta di un nuovo e fertile campo di ricerca, a supporto del quale un ruolo importante è giocato da modelli di programmazione parallela strutturata caratterizzati da chiari modelli di costo;

- b) tecniche di *workflow mining*, capaci cioè, in una computazione espresse mediante un certo schema di workflow fissato, di fornire strumenti per controllare il comportamento reale del sistema di allo scopo di predire le più probabili modalità e le anomalie di esecuzione. Nelle applicazioni di e-business vanno effettuate molte scelte durante l'esecuzione di workflow; alcune scelte possono condurre ad un beneficio, altre dovrebbero essere possibilmente evitate in avvenire. Allo scopo sono state proposte tecniche di data mining che, sulla base delle tracce delle esecuzioni già effettuate, raccolte in opportuni log, permettono di estrarre conoscenza inattesa ed utile sul processo (ad esempio una certa sequenza di esecuzione di attività porta prima o poi a uno stato indesiderato, esecutori in continuo ritardo, attività svolte raramente, coppie di attività mai svolte nella stessa istanza) e per prendere le opportune contromisure e compensazioni nelle esecuzioni dei casi futuri. Questo problema di identificazione di attività critiche e di caratterizzazione di successi/insuccessi è stato recentemente affrontato tramite opportune rappresentazioni grafiche sia della schema che delle istanze di workflow (una istanza è un sottografo dello schema e un pattern è, a sua volta, un sottografo di una istanza) e ricorrendo ad algoritmi di data mining con vincoli che specializzano l'algoritmo classico a-priori per l'estrazione di pattern significativi e frequenti;
- c) *tecniche di process mining*, in un certo senso il problema duale del workflow mining, sono meccanismi di astrazione e induzione che, partendo da istanze, derivano lo schema che le descriva tutte. Recenti lavori hanno affrontato il problema di scoprire uno schema "sconosciuto" di workflow di un dato processo, partendo dai log di un certo numero di sue esecuzioni, sfruttando approcci basati su tecniche di clustering delle esecuzioni, in cui ogni cluster di esecuzioni aventi la stessa struttura e lo stesso comportamento non catturato dai vincoli locali viene utilizzato per derivare vincoli globali e per sintetizzare uno schema di processo che contenga una descrizione articolata dei vari casi possibili di esecuzione.

3.5. Sommario

Ricapitolando, si possono individuare i seguenti temi specifici di ricerca:

1. ***Modelli e strumenti di sviluppo di applicazioni adattive ad alte prestazioni:*** approcci ad alto livello basati su componenti ad alte prestazioni e su adeguati modelli di costo, allo scopo di privilegiare il raggiungimento di adeguati trade-off tra interoperabilità, modularità, riusabilità, affidabilità del software e performance;
2. ***Modelli e strumenti di knowledge management,*** non solo per lo sviluppo di componenti per la costruzione delle applicazioni, ma soprattutto per la realizzazione delle politiche di controllo dell'adattività e della dinamicità nella fruizione delle risorse in funzione della Qualità del Servizio richiesta

(scoperta, allocazione, scheduling) e per il supporto a nuove funzionalità di tolleranza ai guasti basate sui concetti dell'Autonomia;

3. ***Modelli e strumenti per Middleware***, sia a livello di supporti base ("foundations") che di Servizi, anche in accordo all'evoluzione degli standard internazionali, con particolare riguardo a: transazioni; scheduling, gestione delle risorse e ottimizzazione; interfacce per utenti, dispositivi e codici mobili; modelli e strumenti di controllo della security, capaci di superare i limiti delle soluzioni attuali, molto gravose in termini di flessibilità, programmabilità e performance, mediante approcci che incorporino i meccanismi dalle politiche di security, e che quindi si prestino ad una completa integrazione con strumenti di programmazione e di gestione delle risorse in un contesto altamente dinamico e adattivo;
4. ***Modelli e strumenti per Grid Operating Systems***, che, eventualmente adattando e interfacciando sistemi operativi tradizionali, mettano a disposizione un supporto tollerante ai guasti e scalabile su cui possano essere realizzati i Servizi di Middleware attraverso l'interfaccia "foundations";
5. ***Modelli e strumenti per la pianificazione, gestione, esecuzione, controllo e manutenzione dei processi applicativi***, in particolare nel dominio *e-business*, per permettere, in contesti applicativi altamente dinamici, di esprimere le richieste degli utenti, identificare potenziali "fornitori virtuali" dei servizi richiesti, selezionarli mediante negoziazione dinamica e in accordo a contratti circa la qualità del servizio, accedere ai servizi scelti, gestire situazioni di guasto e malfunzionamento, monitorare l'esecuzione, gestire aspetti amministrativi nell'uso delle risorse e servizi.

4. Tematiche di ricerca in relazione alla situazione nazionale ed europea

Il posizionamento nei confronti della situazione internazionale è obiettivo di altri capitoli della proposta di piattaforme tecnologica. Questa sezione prende in considerazione specificamente il posizionamento nei confronti delle tematiche di ricerca indicate in precedenza.

Sulle tematiche di ricerca delle sezioni 3, l'Italia e l'Europa possono contare su significative competenze ed esperienze, che contribuiscono a creare la base per future iniziative per piattaforme tecnologiche nel campo delle Grid di prossima generazione:

- in Italia è attualmente in corso il progetto di ricerca di base FIRB *Grid.it* [4] ("Piattaforme abilitanti per griglie computazionali ad alte prestazioni orientate ad organizzazioni virtuali scalabili"), nel quale sono studiate e sperimentate soluzioni innovative ai suddetti livelli, partendo dall'esistenza degli standard disponibili per i servizi base di Middleware. In particolare, sono indagate nuove soluzioni ai livelli dell'ambiente generale di programmazione e dei servizi di Knowledge Management, soluzioni che vengono sperimentate su varie infrastrutture di Grid (CNR, ASI) e sulla Grid di Produzione (INFN), anche sperimentando l'utilizzazione di reti fotoniche ad alte bande (CNIT);
- diversi importanti progetti di ricerca industriale sono in fase di definizione;

- in Europa, nell'ambito del 6th Framework, è anzitutto da citare il progetto NoE *CoreGrid* ("European-wide virtual laboratori for longer term Grid research – creating the foundations for the next generation Grids"), oltre ai progetti IP NextGrid ("EU-driven Grid services architecture for business and industry") ed AKOGRIMO ("Mobile Grid architecture and services for dynamic virtual organizations"), e ad alcuni STREP. Inoltre, i progetti *XtreemeOS* e *GridComp* sono di prossima attivazione;
- nel panorama europeo, 5th Framework, di particolare rilievo nel settore delle applicazioni non scientifiche è il progetto *GRASP* ("Grid Application Service Provision"), i cui risultati sono valicati mediante tre test-bed, due nel settore del biomedicale e uno in quello finanziario.

Queste iniziative sono importanti anche alla luce dell'analisi condotta in [2], in base alla quale in alcuni dei settori scientifici e tecnologici suddetti l'Europa può già contare su consolidati punti di forza (semantic web, ontologie, data/web mining, modelli e linguaggi di programmazione ad alte prestazioni, modelli di costo, embedded systems, mobile computing), che possono e devono essere ulteriormente rafforzati nell'ottica di una reale leadership, mentre in settori tradizionalmente più deboli dal punto di vista prettamente tecnologico e commerciale (sistemi operativi, basi di dati, ambienti integrati, architetture hardware-software) lo sviluppo della ricerca nel settore Grid rappresenta una occasione concreta per colmare, almeno parzialmente, il gap esistente nei confronti di US e Giappone.

Questo obiettivo può essere realisticamente affrontato secondo l'approccio rivolto a prodotti e servizi ad alto valore aggiunto realizzati a partire da standard comuni open-source. Questo comporta una transizione dell'industria europea del software, ed una interessante caratterizzazione delle PMI specializzate in piattaforme abilitanti, verso la customizzazione del software a partire da un minimo di Middleware standard. La stima riportata in [2] è che l'industria europea sia potenzialmente ben posizionata nei confronti di questo approccio, e quindi potenzialmente in grado di recepire gli avanzamenti della ricerca e sviluppo nel campo delle Grid di prossima generazione, a condizione che venga costituita la necessaria massa critica attualmente sottodimensionata.

L'analisi di queste problematiche è anche un obiettivo primario del progetto *GridCoord* [5], una SSA del 6th Framework ("ERA Pilot on a Coordinated Europe-wide Initiative in Grid Research"), sul coordinamento delle iniziative europee su Grid Computing e sull'elaborazione di strategie comuni alla ricerca e sviluppo nel settore.

Riferimenti

1. Gruppo di Lavoro ICT coordinato dalla Fondazione Ugo Bordoni: *La ricerca nazionale nel settore Information and Communication Technology - Analisi dell'area e proposta di temi di ricerca*. MIUR, 2003.
2. EC Information Society Technologies Programme Advisory Group, Working Group 8: *Grids, Distributed Systems and Software Architectures*. August 2004.

3. EC Information Society, Unit F2, Expert Group Report: *Next Generation Grids 2 - Requirements and Options for European Grids Research 2005-2010 and Beyond*. July 2004. *Next Generation Grids 3*, 2005.
4. *Grid.it - Piattaforme abilitanti per griglie computazionali ad alte prestazioni orientate ad organizzazioni virtuali scalabili*. Progetto MIUR-FIRB, 2002-2005, www.grid.it.
5. *GridCoord - ERA Pilot on a Coordinated Europe-wide Initiative in Grid Research*. SSA, 6th Framework Programme, Technical Annex, July 2004.