



Enabling Grids for
E-science in Europe



Διημερίδα

«Προχωρημένα Θέματα
σε Τεχνολογίες Πλέγματος»

Βόλος 14-15 Μαΐου 2007



Εφαρμογές e-Science στο GRID

Μανόλης Βάβαλης
Παν. Θεσσαλίας - ΚΕΤΕΑΘ



Περιεχόμενα

- Εισαγωγή - Ηλ-επιστήμες (e-science)
- Υπάρχουσες Τεχνολογίες
- Πιθανά Προβλήματα
- Τυπικές Εκδοχές Χρήσης
- Ειδικά Παραδείγματα
 - NetSolve, GasTurbine Lab, ARION, ...

Καλές Υποψήφειες Εφαρμογές για GRID

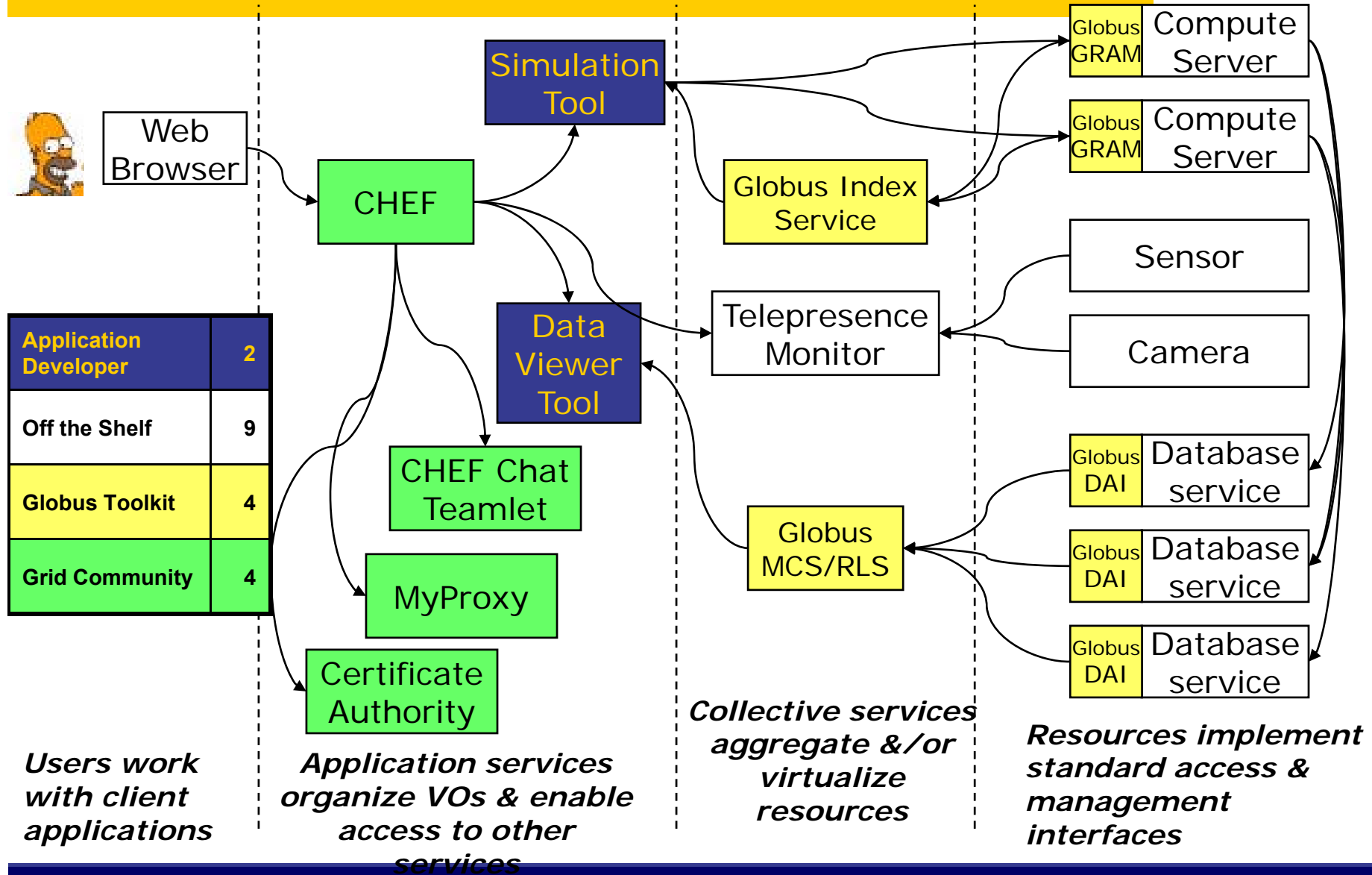
- **Δεδομένα**
 - Μεγάλου όγκου (πολλά)
 - Κομματισμένα (διαφορετικές πηγές, τοποθεσίες, διατάξεις, μετα-δεδομένα)
 - Σε επαναλαμβανόμενα πολλά αντίγραφα
- **Υψηλές Υπολογιστικές Απαιτήσεις**
 - Τα προβλήματα της περιοχής μπορούν να παραλληλισθούν (με περιορισμένη αλληλεπίδραση μεταξύ των διεργασιών).
 - Υπάρχει όφελος από την ύπαρξη πληθώρας CPUs
- **Επιθυμία για διαμοιρασμό**
 - Πόρων
 - Ανθρώπινου δυναμικού
 - Γνώσης

Επιπρόσθετο κέρδος: Απροσδόκητα μεγάλες απαιτήσεις σε κάποιο από τα παραπάνω.

- Παράλληλες διεργασίες σε πολλαπλές τοποθεσίες:
 - Δεν είναι δυνατόν να εγγυηθούμε την ταυτόχρονη έναρξη όλων των διεργασιών.
 - Δεν είναι εύκολο να ελέγξουμε το εύρος ζώνης και τις λανθάνουσες καταστάσεις των διεργασιών.
 - Μπορούμε βεβαίως να εκθέσουμε μια MPI-εφαρμογή σαν πόρο ενός grid.
- Διεργασίες που απαιτούν απόκριση πραγματικού χρόνου.
 - Όπως και παραπάνω.

Το GRID Computing ΔΕΝ ΕΙΝΑΙ Cluster Computing

Πώς Μοιάζει μια Εφαρμογή Grid;



Τι εννοούμε με τον όρο e-science;

e-Science is about global collaboration in key areas of science and the next generation of infrastructure that will enable it.

- **Άλλοι Ορισμοί:**

- E-science is the science that can be achieved through the use of computers
 - to connect different sources of data about a subject, usually collected independently,
 - to extract new information beyond that which is in each data set taken separately,
 - to generate new knowledge and understanding
- Large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet

- **Απλοϊκός Ορισμός (δικός μου):**

- Επιστήμη σε μέτα επίπεδο με χρήση τεχνολογιών GRID

Κίνητρα: Σε ποια εικοσαετία βρίσκεσθαι;

- Γράφω το λογισμικό τελείως μόνος μου (50-60)
- Χρησιμοποιώ το λογισμικό που υπάρχει στο υπολογιστικό μου σύστημα (70-80)
- Χρησιμοποιώ βιβλιοθήκες από το διαδύκτιο και κάποιο υπολογιστικό σύστημα δικό μου ή γνωστού μου (90-00)
- Κάνω μέτα-υπολογισμούς χρησιμοποιώντας υπηρεσίες GRID

Μετα υπολογίζω ΔΕΝ σημαίνει δεν ασχολούμαι με προγραμματισμό.

- Δυστυχώς ο χρόνος που έως τώρα ξοδεύαμε για προγραμματισμό δεν προβλέπεται να μειωθεί σημαντικά.
- Αλλάζει όμως η φύση του και είναι πολύ πιο αποδοτικός

Κίνητρα: Παραδείγματα από Επιστημονικούς Υπολογισμούς

- Έχετε να λύσετε ένα (μεγάλο) γραμμικό αλγεβρικό σύστημα. Ποιόν αλγόριθμο/λογισμικό θα χρησιμοποιήσετε;
- Προτίθεστε να κάνετε ανάλυση Πεπερασμένων Στοιχείων. Θα χρησιμοποιήσετε πολλά στοιχεία και σχήμα χαμηλής τάξης ή λιγότερα στοιχεία και σχήμα υψηλής τάξης;
- Κατασκευάσατε τον X αλγόριθμο για την λύση του Y προβλήματος. Πώς θα μπορέσετε να τον συγκρίνετε με τους υπάρχοντες αλγορίθμους που επιλύουν το συγκεκριμένο Y πρόβλημα;
- Εκτελείτε μια μεγάλη σειρά προσομοιώσεων ενός φυσικού προβλήματος. Πώς θα καταγράψετε τις πληροφορίες σχετικά με την κάθε εκτέλεση;

- Εισαγωγή - Ηλ-επιστήμες (e-science)
- **Υπάρχουσες Τεχνολογίες**
- Πιθανά Προβλήματα
- Τυπικές Εκδοχές Χρήσης
- Ειδικά Παραδείγματα
 - NetSolve, GasTurbine Lab, ARION, ...

Υπάρχουσες Τεχνολογίες Λογισμικού: Περιγραφή (μετα) δεδομένων

- XML (eXtensible Markup Language)
 - Μια απλή, γενική & επεκτάσιμη γλώσσα περιγραφής (μετα) δεδομένων
 - για (δεδομένα & προγράμματα)
 - Έχει καθιερωθεί σαν W3C standard με πολλές εφαρμογές. Π.χ.
 - Scalable Vector Graphics –SVG
 - Resource Description Framework –RDF
 - Mathematical ML –MathML
 - Chemical ML –CML
 - Xpath, Xlink, Xpointer, ...
- Απαραίτητη για την οριζόντια ενοποίηση προγραμμάτων και δεδομένων

Υπάρχουσες Τεχνολογίες Λογισμικού: Κατανεμημένοι Υπολογισμοί & Πρωτόκολλα

Προσφέρουν ένα τυποποιημένο τρόπο επικοινωνίας μεταξύ διαφορετικών εφαρμογών λογισμικού που τρέχουν σε διαφορετικές υπολογιστικές πλατφόρμες & πλαίσια εργασίες.

- **SOAP, XML-RPC, ...** : Πρωτόκολλα για την ανταλλαγή δεδομένων μεταξύ εφαρμογών / υπηρεσιών δικτύου
- **WSDL (Web Service Description Language)**: για την περιγραφή υπηρεσιών δικτύου.
 - Μας επιτρέπει να διαχωρίζουμε την περιγραφή της ουσιαστικής λειτουργικότητας που μας προσφέρει μια υπηρεσία από την περιγραφή των ειδικών λεπτομερειών της (το *πώς* και το *πού*).
- **OGSA (Open-Grid-Services Architecture)**: μια πρόταση για υπηρεσίες στο διαδύκτιο
 - Συνεργασία μεταξύ W3C & GGF

Υπάρχουσες Τεχνολογίες Λογισμικού: Γλώσσες Προγραμματισμού

- **Java**
- **Python**

- Εισαγωγή - Ηλ-επιστήμες (e-science)
- Υπάρχουσες Τεχνολογίες
- **Πιθανά Προβλήματα**
- Τυπικές Εκδοχές Χρήσης
- Ειδικά Παραδείγματα
 - NetSolve, GasTurbine Lab, ARION, ...

Πιθανά Προβλήματα: Προβλήματα κουλτούρας/συνεργασίας

Ξεκάθαρο παράδειγμα δι-επιστημονικής συνεργασίας μεταξύ επιστημόνων εφαρμογών και επιστημόνων πληροφορικής

Αναπόφευκτα δημιουργεί τριβές

- Μεταξύ έρευνας και ανάπτυξης (ένα ακραίο παράδειγμα):
 - Οι επιστήμονες των εφαρμογών θέλουν να αναπτύξουν έρευνα στην περιοχή τους
 - Οι επιστήμονες της πληροφορικής θέλουν να αναπτύξουν έρευνα στην δική τους περιοχή
 - Οι τεχνικοί των εταιριών επιθυμούν την δημιουργία προϊόντων
 - Κανείς δεν θέλει να παράγει εύχρηστο και εύρωστο λογισμικό για το GRID

Απαραίτητοι και πρωταρχικοί στόχοι στο e-Science:

1. Η ταυτόχρονη ικανοποίηση των επιθυμιών όλων των εμπλεκομένων επιστημόνων
2. Η διατήρηση του επιστημονικού ενδιαφέροντος όλων στον κοινό στόχο.

Απαιτείται δράση εκλεπτυσμένης ισορροπίας

Πιθανά Προβλήματα: Προβλήματα Ασφάλειας

Η ασφάλεια των συστημάτων έχει πολλές συνιστώσες. Ας δούμε μόνον δύο από αυτές:

1. Συνήθως οι διαχειριστές υπολογιστικών συστημάτων των διαφόρων οργανισμών

- Εγκαθιστούν ένα σύστημα με firewalls για να αποτρέψουν την είσοδο εισβολέων
- Κλειδώνουν όσες πόρτες (και παράθυρα) μπορούν

Στην περίπτωση των Ιδεατών Οργανισμών (στα GRIDS) πρέπει να πεισθούν οι διαχειριστές να εμπιστευθούν τα πιστοποιητικά των ΙΟ και να επιτρέψουν τους χρήστες που δεν ανήκουν στον οργανισμό τους να χρησιμοποιήσουν τους πόρους τους.

2. Προστασία των προσωπικών δεδομένων.

- Παράδειγμα: Ανωνυμία δεδομένων ασθενών.

Πιθανά Προβλήματα: Προστασία Πνευματικής Ιδιοκτησίας

Όταν διάφοροι επιστήμονες που ανήκουν σε διαφορετικούς οργανισμούς συνεργάζονται είναι σημαντικό (και δύσκολο) να καθορισθεί ποιος ανακάλυψε τι.

Το πρόβλημα είναι ιδιαίτερα σύνθετο όταν εμπλέκονται εταιρίες.

- Εισαγωγή - Ηλ-επιστήμες (e-science)
- Υπάρχουσες Τεχνολογίες
- Πιθανά Προβλήματα
- **Τυπικές Εκδοχές Χρήσης**
- Ειδικά Παραδείγματα
 - NetSolve, GasTurbine Lab, ARION, ...

Τυπικές Απλές Εκδοχές Χρήσης: Χρήση Batch

- Χρησιμοποιούμε το grid σαν μια πελώρια δεξαμενή υπολογιστών.
- Στην ουσία πολλές σειριακές εκτελέσεις.
- Τυπικές θεματικές περιοχές
 - Θεωρία Αριθμών
 - Προσομοιώσεις Στατιστικού Τύπου
 - Παρεμβολή ...
- Παράδειγμα

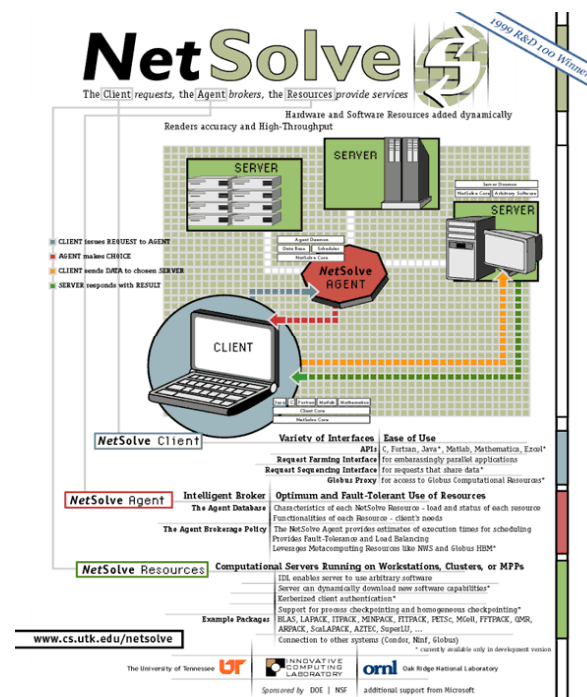


Search for Extraterrestrial Intelligence (SETI)

- Ένα επιστημονικό πείραμα χρήσης GRID.
- Προς το παρόν η «μεγαλύτερη προσπάθεια καταναμεμένων υπολογισμών» με περισσότερους απο 3 εκατομμύρια χρήστες.
- Μπορείς να συμμετάσχεις και εσύ εκτελώντας ένα πρόγραμμα που κατεβάζει και αναλύει ραδιο-τηλεσκοπικά δεδομένα ...

Τυπικές Απλές Εκδοχές Χρήσης: Στατικά Περιβάλλοντα Επίλυσης Προβλημάτων

- Στατικές υλοποιήσεις πάνω σε GRID
- Παρέχονται σαν υπηρεσίες GRID
- Τυπικές θεματικές περιοχές
 - Επιστημονικοί Υπολογισμοί Ευρείας Κατανάλωσης
- Παράδειγμα



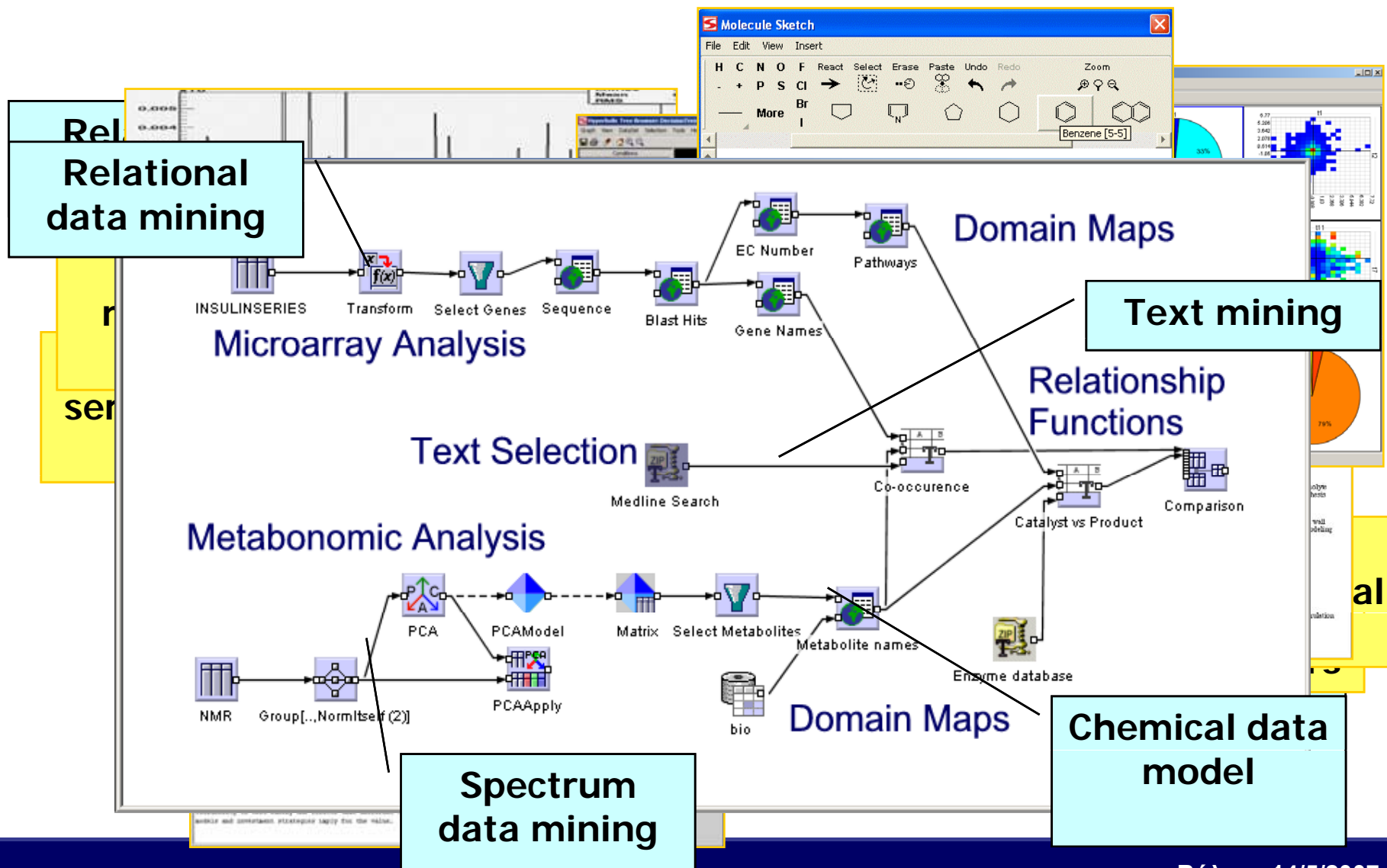
Λεπτομέρειες
σε λίγο ...

Παράδειγμα:

Ανάλυση Γονιδιωμάτων με εργαλεία e-Science

- Μικρής κλίμακας έργο
- Συμμετέχοντες
 - Computer Science Department, University of Manchester
 - St Mary's Hospital, Manchester
 - European Bioinformatics Institute
- Αντικείμενο
 - Ανάπτυξη & χρήση ροών εργασίας (workflows) χρησιμοποιώντας συνιστώσες του myGrid για την απεικόνιση περίπλοκων γονιδιακών περιοχών σχετιζόμενες με τρία γενετικά σύνδρομα.
- Αποτελέσματα
 - Παρήχθησαν ενδιαφέροντα αποτελέσματα σε χρόνο ρεκόρ.

Παράδειγμα: Εφαρμογή στο DataGRID



Τυπικές Απλές Εκδοχές Χρήσης

Περιβάλλοντα επίλυσης με δυναμικές ροές εργασίας

- Το GRID είναι στην ουσία στατικό
- Ο χρήστης μπορεί να δημιουργήσει νέες ροές εργασίας συνθέτοντας υπάρχουσες υπηρεσίες GRID
- Παράδειγμα:



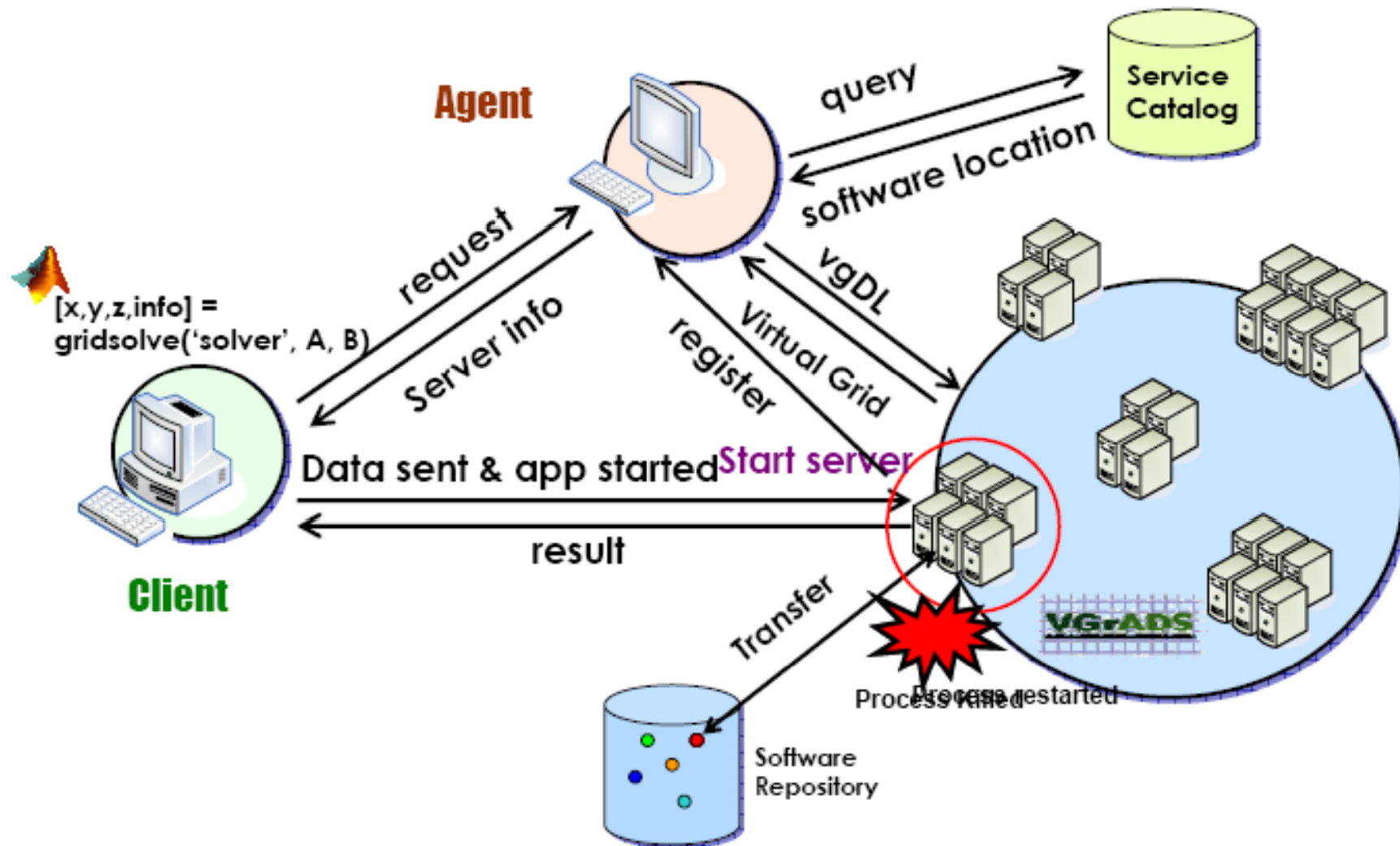
- Μεσαίας κλίμακας έργο
- Ένα e-Science έργο για Επιστημονικές Συλλογές
- Σενάρια για κυματισμούς στην θάλασσα
- Λεπτομέρειες σε λίγο ...

- Εισαγωγή - Ηλ-επιστήμες (e-science)
- Υπάρχουσες Τεχνολογίες
- Πιθανά Προβλήματα
- Τυπικές Εκδοχές Χρήσης
- **Ειδικά Παραδείγματα**
 - NetSolve, GasTurbine Lab, ARION, ...

Τι είναι το NetSolve/GridSolve

- Ένα Client-server σύστημα τύπου-RPC
 - Αυξημένης ευχρηστίας
- Αλληλεπιδράσεις μέσω ενός διαμεσολαβητή (agent)
 - π.χ scheduling, tracking, fault tolerance
- Δυναμικές δεσμεύσεις υπηρεσιών
 - Ο Client δεν χρειάζεται να έχει πιστοποιητικό πρόσβασης στην συγκεκριμένη υπηρεσία που επιθυμεί να χρησιμοποιήσει
- Πολλαπλοί clients
 - C, Fortran, Matlab, Java, Mathematica, Octave
- Επεκτάσιμο για την υποστήριξη του GridRPC API
 - Μέρος του GGF working group για τον καθορισμό πρότυπου API

Η Αρχιτεκτονική του GridSolve



Σύνθετα προβλήματα: Multidomain – Multiphysics

- Gas Turbine engine simulation
- Air pollution modeling
- Weather prediction problem
- Aerodynamic and aeroacoustic simulation
- Nuclear reactor simulation
- Underwater acoustic problems
- Modelling of semi- and super-conducting devices
- etc.

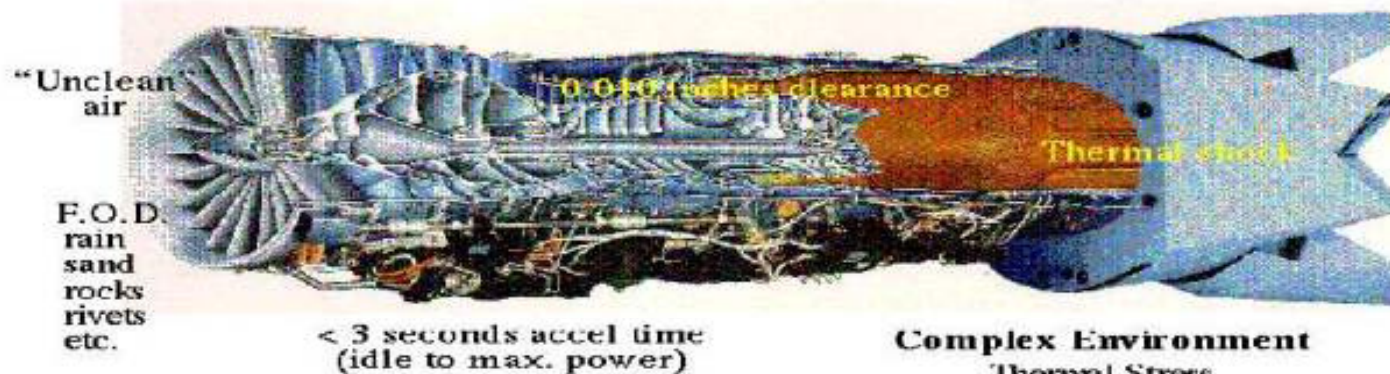
Gas Turbine Engine

31,000 PARTS

1,300 parts rotate to 50,000 rpm
Up to 10 "g" maneuver loads (repeatedly)
Up to 40 atmospheres of pressure

Constant changes of throttle

Constant flight envelope changes
Max maneuvering to meet mission
-- each flight and pilot is different



High Strength Materials Are Sensitive

Notches (< 0.002 inch)
Rubs (small kisses)
Chemical attack (oxidation/corrosion)
Surface defects (natural & service)
Long term exposure (creep)
Cyclic mechanical & thermal loading order
Synergistic loads

Complex Environment

Thermal Stress

up to ~1400F above melt temp
rich in oxygen & sulfur

Mechanical Stress

~80% ultimate strength
steady + vibratory loads
fluctuating pressure fields
mech. & thermal deflections
acoustics
impact loads

- **Φυσική:** Gas flow, combustion, highly stressed materials, high temperatures, multi-scale in time/space
- **Εξαρτήματα:** Fans, compressor, combustor, high-pressure turbine, low-pressure turbine
- **Γεωμετρία:** 31,000 parts, 1,300 of them rotate at 50,000 rpm

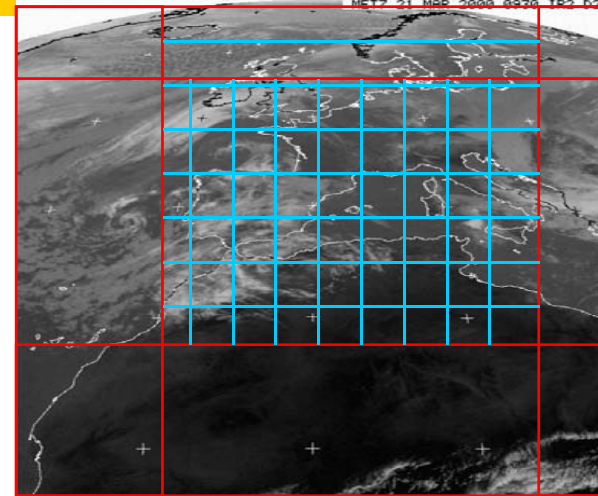
Προβλήματα Πρόγνωσης Καιρού - Μόλυνσης της Ατμόσφαιρας

- **Πρόγνωση Καιρού**

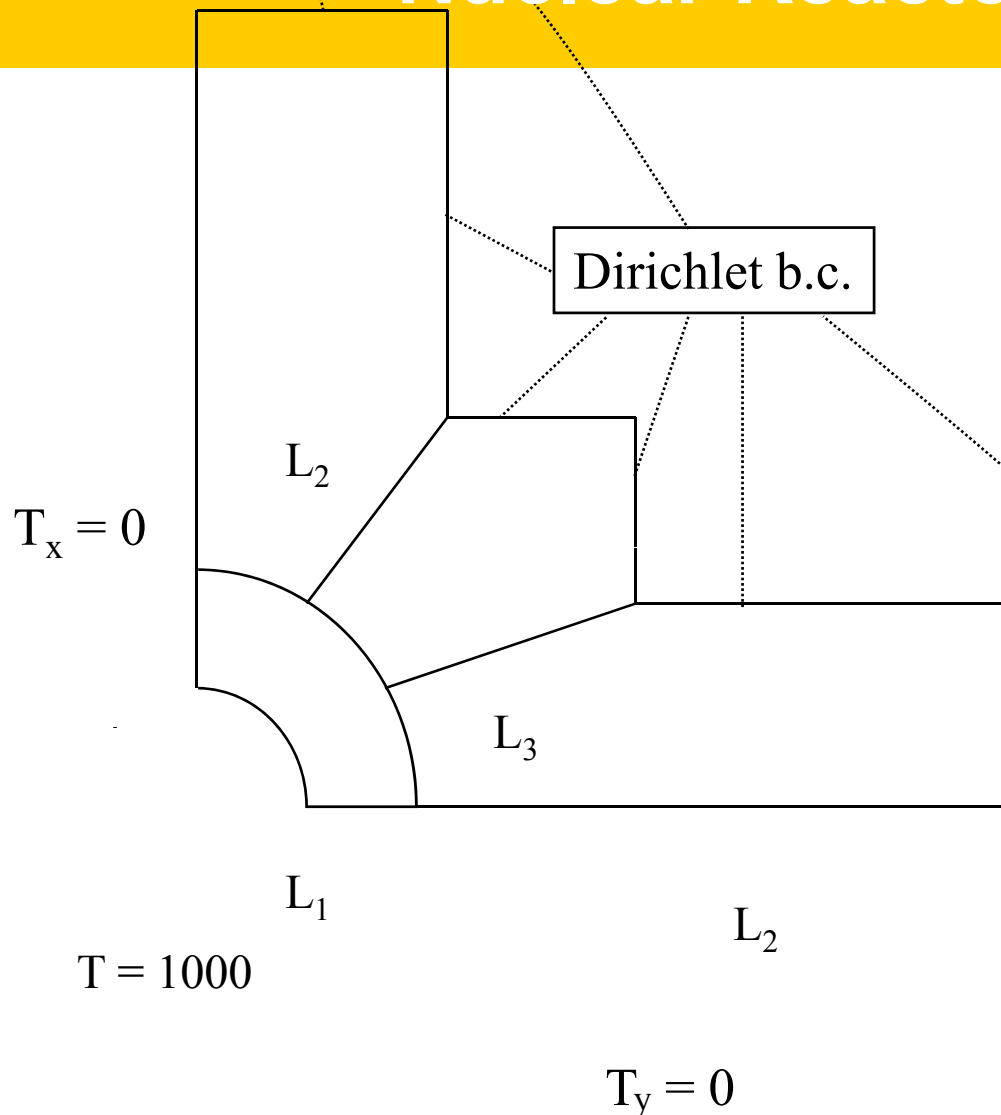
- Απλό σχετικά πρόβλημα ορισμένο σε πελώριο χωρίο

- **Ατμοσφαιρικοί Ρύποι**

- Η χημική συμπεριφορά του κάθε ρύπου είναι απολύτως γνωστή. Η αλληλεπίδρασή του με άλλα στοιχεία του περιβάλλοντος είναι ουσιαστικά ακόμα άγνωστη.
- Διαφορετικά μοντέλα ανάλογα με:
 - Το υψόμετρο
 - Εθνικά σύνορα
 - Γεωγραφικά σύνορα
 - Τα μικροκλίματα
 - ...
- Χημικές αντιδράσεις σε διαφορετικές χρονικές κλίμακες



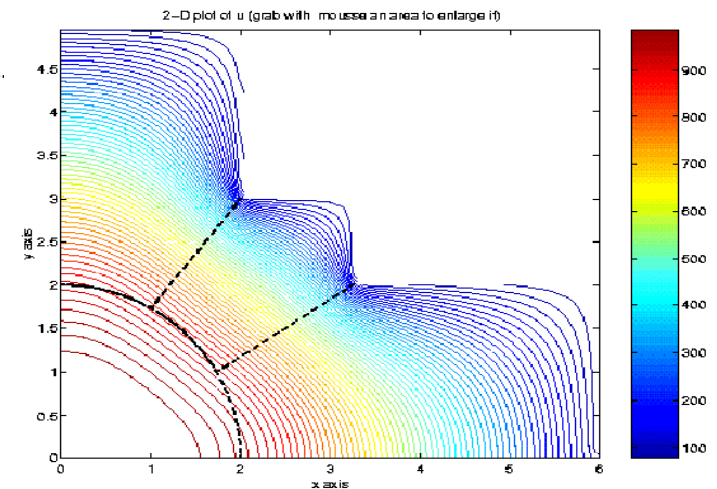
Nuclear Reactor Problem



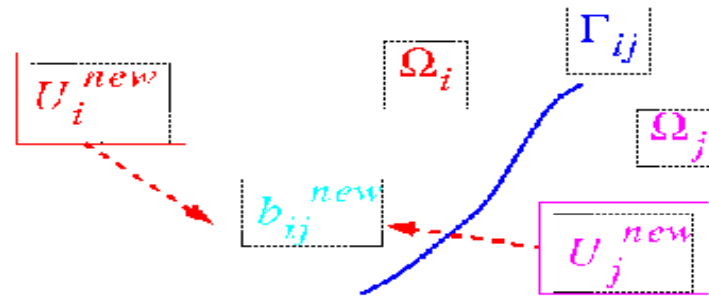
$$L_1 = T_{xx} + T_{yy} + \alpha_1 T - \beta_2 (x^2 + y^2 - 2)$$

$$L_2 = T_{xx} + T_{yy} + \alpha_2 T$$

$$L_3 = T_{xx} + T_{yy} - \gamma_3 (T_x + T_y) + \alpha_3 T$$



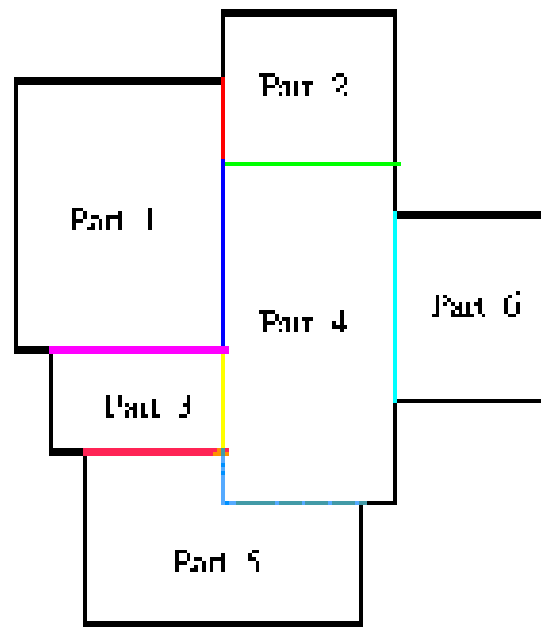
New Schemes are Needed



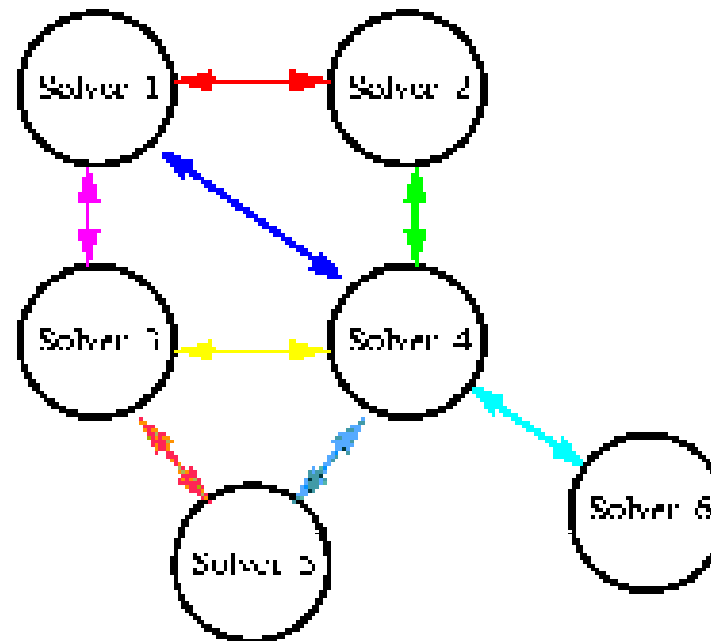
Relaxation:

$$g_{ij}(U_i^{new}, U_j^{new}, \frac{\partial U_i^{new}}{\partial n}, \frac{\partial U_j^{new}}{\partial n}) = 0$$

Collaborative Solvers

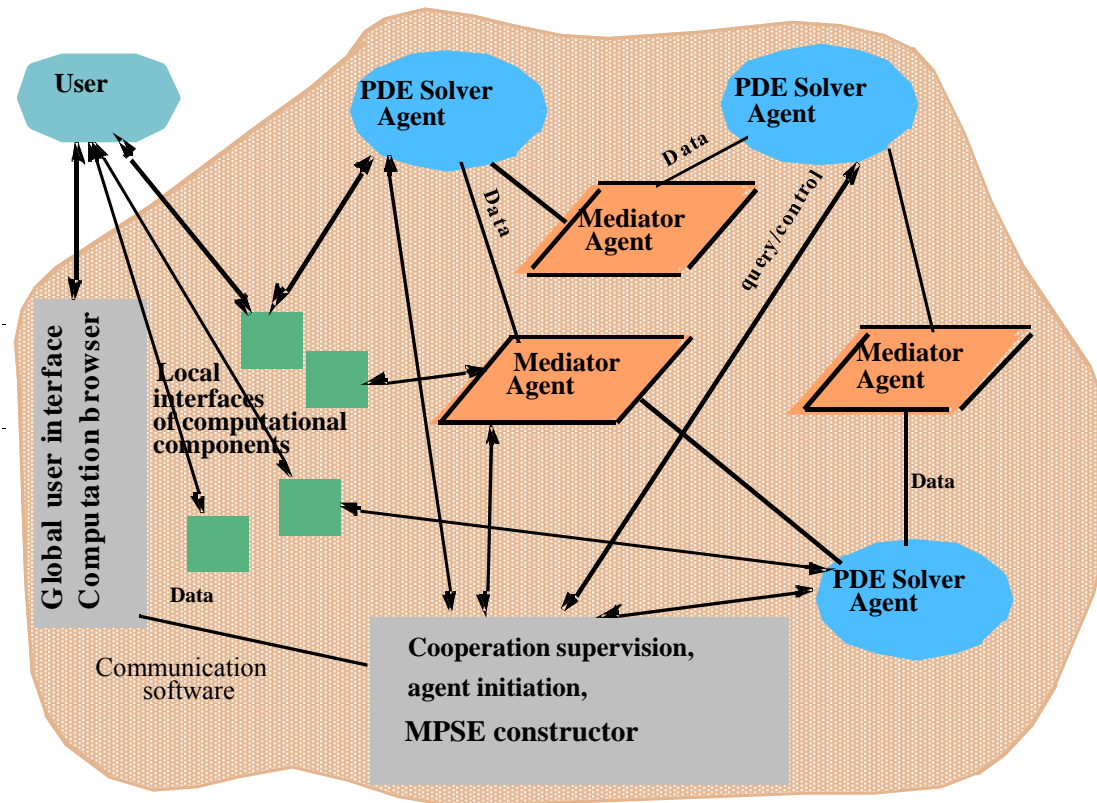


A physical phenomenon



A network of collaborating solvers

PSEs for Multidomain – Multiphysics Problems



Μια επίδειξη χρήσης

