



Comunidad GRID y Supercómputo del CUDI: Principales Iniciativas Nacionales

Coordinador

Dr Luis A. Trejo, ltrejo@itesm.mx

Abril 2010



Objetivos de la Comunidad

MORELIA • Reunión de Primavera

- El grupo de Grid y Supercómputo tiene como objetivo fomentar la colaboración y el desarrollo de proyectos conjuntos entre las instituciones miembros de CUDI, sustentados en una plataforma nacional basada en tecnología GRID y cómputo de alto desempeño, utilizando a su vez la red Internet 2 para el apoyo del desarrollo de la e-ciencia en México.

Miembros de la Comunidad



MORELIA • Reunión de Primavera

- Cerca de 400 miembros registrados en el portal de CUDI
- Grupo de trabajo de aproximadamente 50 integrantes de más de 10 instituciones de educación superior del país.



Principales Iniciativas y Proyectos de Colaboración Internacional

MORELIA • Reunión de Primavera

- LA-GRID. Latinamerican GRID
- LNGSeC
- PRAGMA Institute
- Delta Metropolitana
- IBM-Google Academic Initiative
- EELA-2 (E-Science grid facility for Europe and Latin America).
- Creación del JRU-MX. Junio 2009



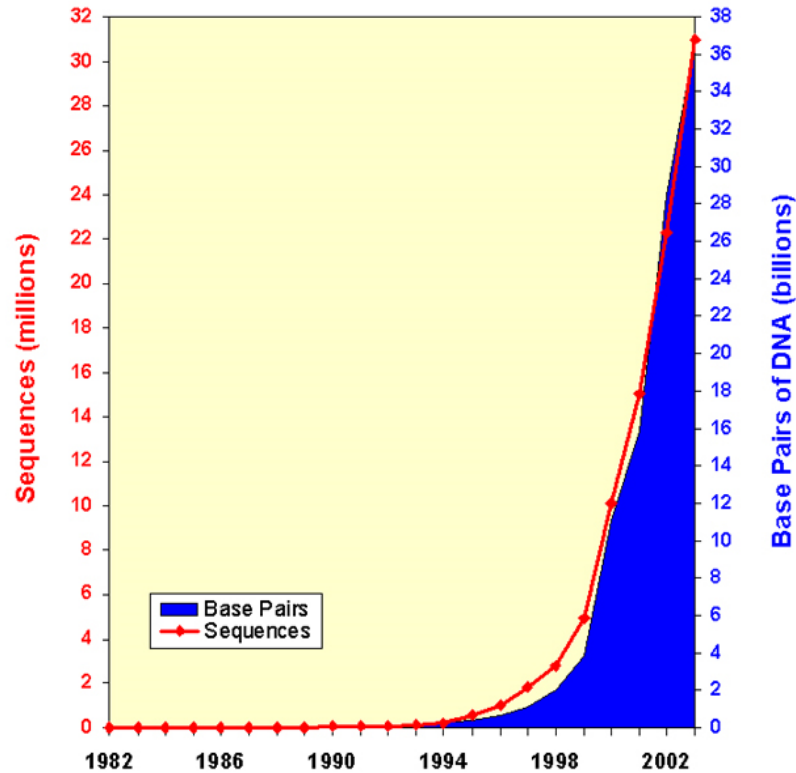
Principales Iniciativas y Proyectos de Colaboración Internacional

MORELIA • Reunión de Primavera

- GISELA (Programa FP7)
- LA-ROC
- Red temática en Cómputo GRID (CONACyT)
- ISUM 2010

Explosion of genomic data

Growth of GenBank



Making sense of genomic data



MORELIA • Reunión de Primavera

PHASE TWO: INTERPRETATION

SHERMAN the Star Ledger



LA Grid



2010

CUDI

M C Unión de Primavera



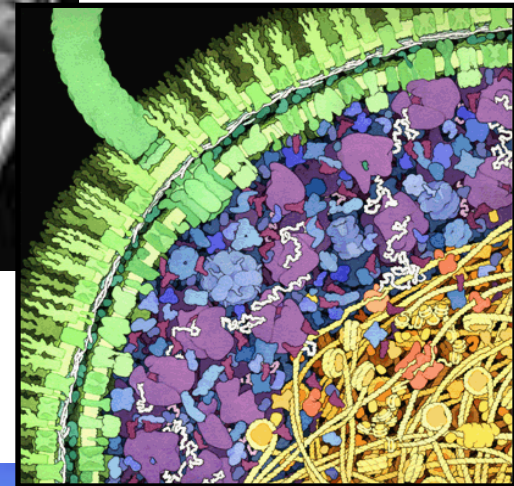


LA GRID

- **Healthcare:** Delivering secure, flexible, real-time collaboration and healthcare information.
- **Life Sciences:** Using new pattern recognition and data mining technics to make medical breakthroughs.
- **Hurricane Disaster Mitigation:** Quickly produce mitigation information based on projected storm parameters, enable data sharing, and scale to meet user demand

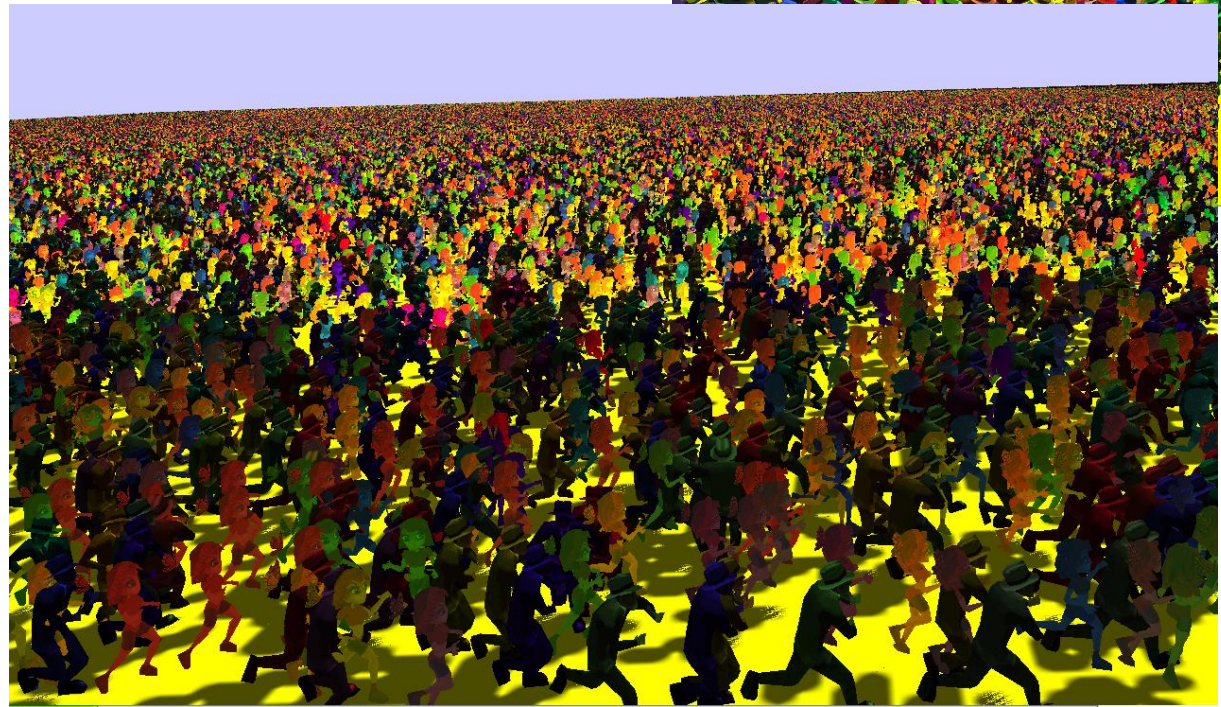
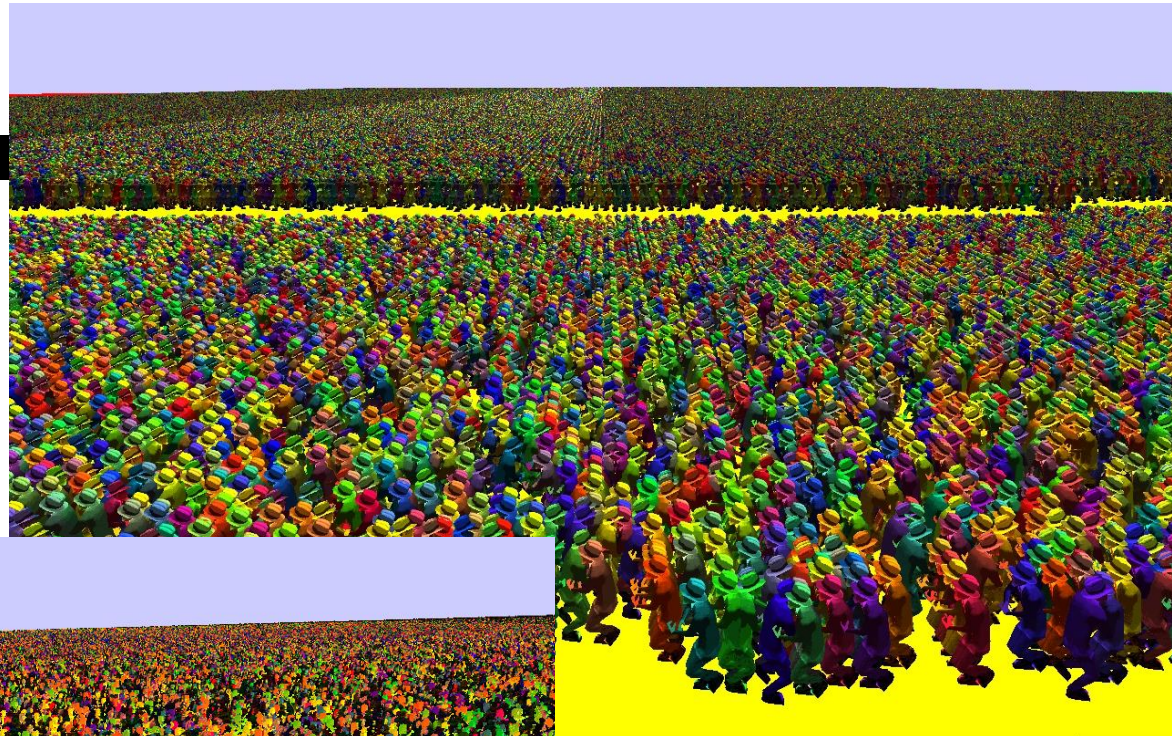


I A • Reunión de Primavera





GPU for large scripted crowds



LNGSeC



MORELIA • Reunión de Primavera

- Proyecto Laboratorios Nacionales, coordinado por CICESE.
 - Laboratorio Nacional de Grids de Supercómputo para el Soporte de Aplicaciones de e-Ciencia (LNGSeC).
 - Modelos Climatológicos
- Universidades involucradas:
 - CUDI, IPICyT, IPN, UDLAP, UDG, UNISON, CICESE, UANL, UCOL, ITESM-CEM

LNGSeC



MORELIA • Reunión de Primavera

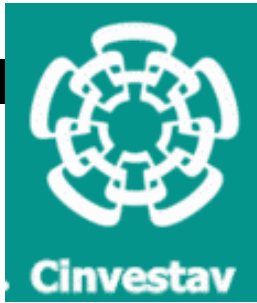
- Proyecto Laboratorios Nacionales, coordinado por CICESE.
 - Laboratorio Nacional de Grids de Supercómputo para el Soporte de Aplicaciones de e-Ciencia (LNGSeC).
 - Modelos Climatológicos
- Universidades involucradas:
 - CUDI, IPICyT, IPN, UDLAP, UDG, UNISON, CICESE, UANL, UCOL, ITESM-CEM

Delta Metropolitana de Cómputo de Alto Rendimiento.



2007

MORELIA • Reunión de Primavera



362 nodes

10 Gps

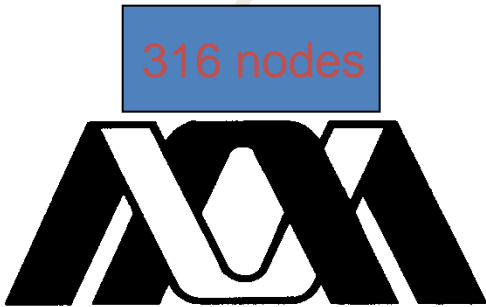
10 Gps

GRID, or "virtual center" with initial capacity of 300 processors per node reaching a minimum of 1,000 processors per node in three years

High performance computer centers connected through high capacity dedicated links

KanBalam. 126th largest supercomputer in the world

"Virtual center" with proper middleware to take advantage of the delta resources,



10 Gps

Casa abierta al tiempo



Cloud Computing

IBM-Google Academic Initiative

MORELIA • Reunión de Primavera

- The Google-IBM Cloud Computing Initiative to support scientific research.
- There are several universities involved in the project, mainly in the USA, and two in Mexico (ITESM, UAA)



Cloud Computing

IBM-Google Academic Initiative

MORELIA • Reunión de Primavera

- By belonging to this initiative universities have access to the following resources:
- The actual cloud, **cluster of 457 processors** running an open source implementation of Google's cloud computing infrastructure including **MapReduce** and the Hadoop Distributed File System.



Cloud Computing Support for e-Science Applications

- **Application 1:**
Statistical methods for timely detection of attacks to the Domain Name System
- **Application 2:**
Quantum algorithms for solving the protein folding problem



REDES DE INTERACCIÓN

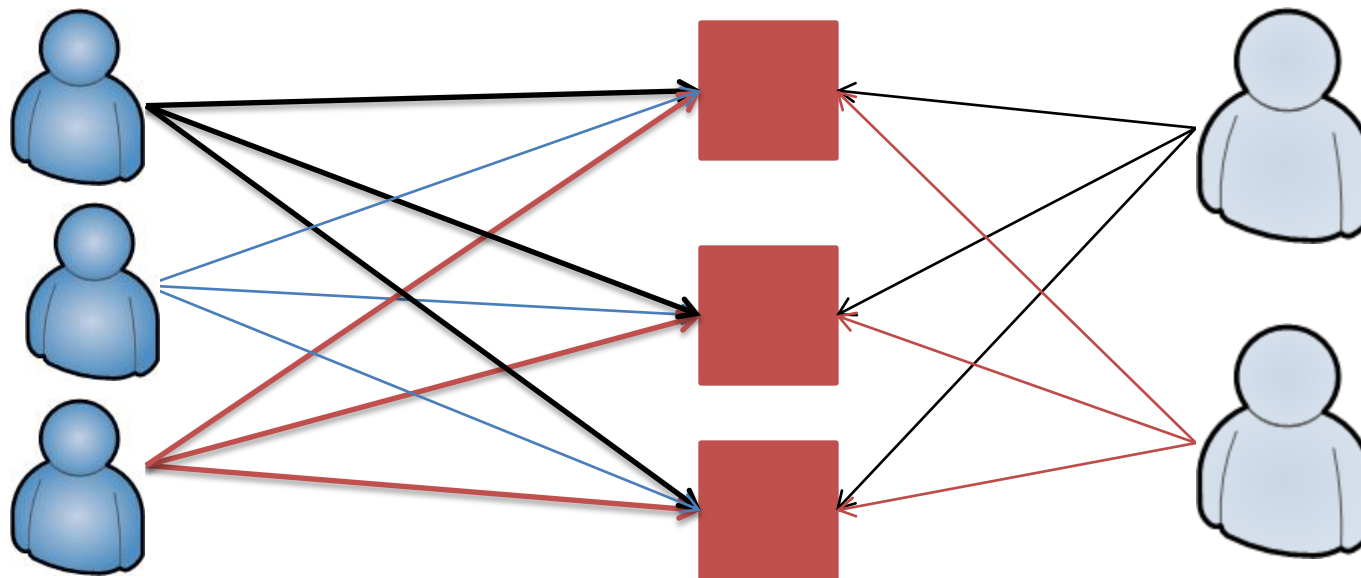


Redes de Interacción

MORELIA • Reunión de Primavera

- Modelo basado en el comportamiento social de los usuarios.
- Se definen características como la cantidad de visitas que un usuario realiza a cierto dominio, que tan relacionado son estas visitas con respecto a las de otro usuario.
- Representado por un grafo acíclico dirigido.

Por ejemplo



Grupo tamaño 3 , peso 5



Variables globales

MORELIA • Reunión de Primavera

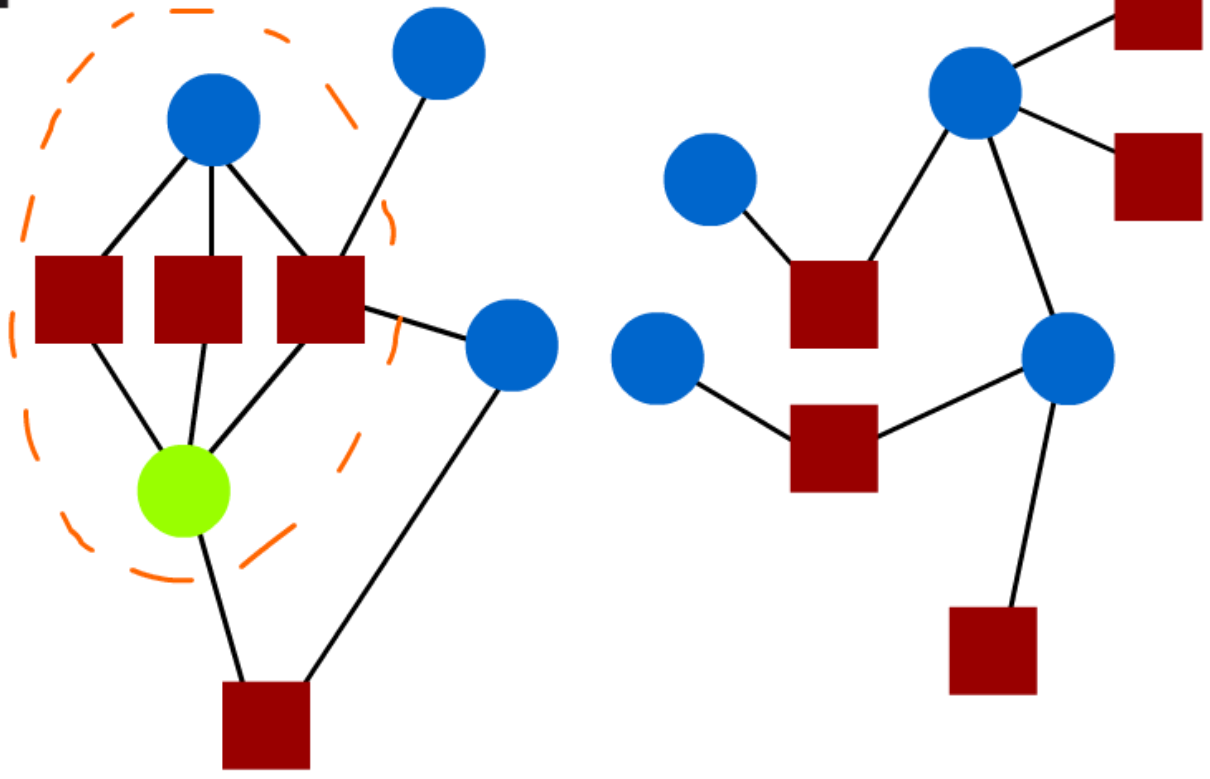
- Símbolos que representan a la red de interacción.
- IP más activa
- Dominio más popular
- Etc.



REACCIÓN DE LAS REDES DE INTERACCIÓN ANTE UN ATAQUE DDOS

Comportamiento normal

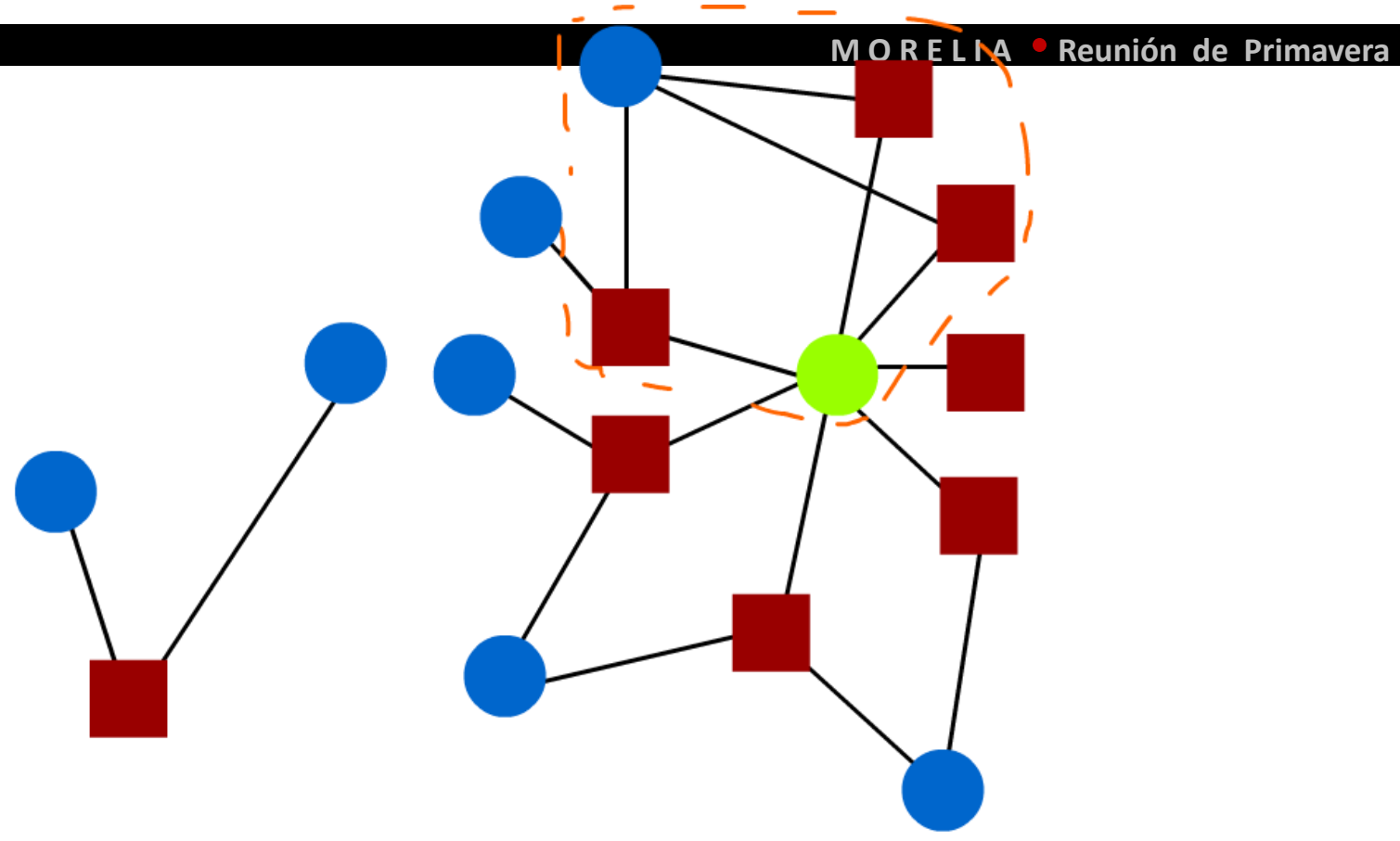
MORE Reunión de Primavera



Peticiones DNS



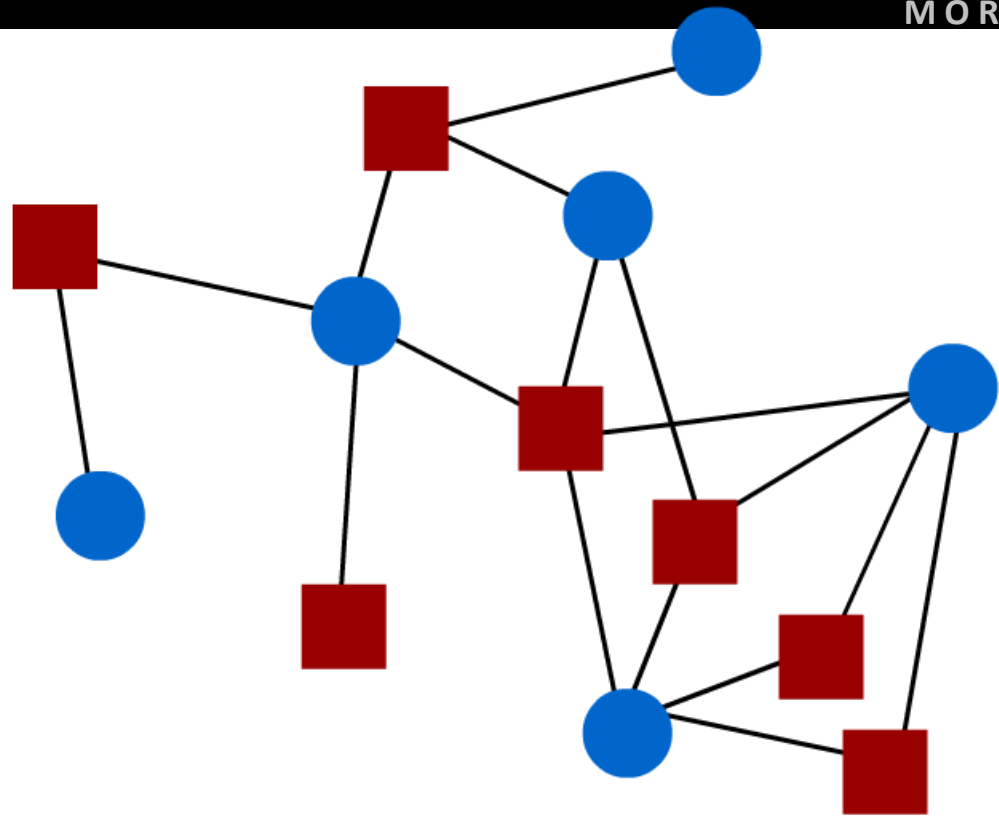
Comportamiento normal



Peticiones DNS



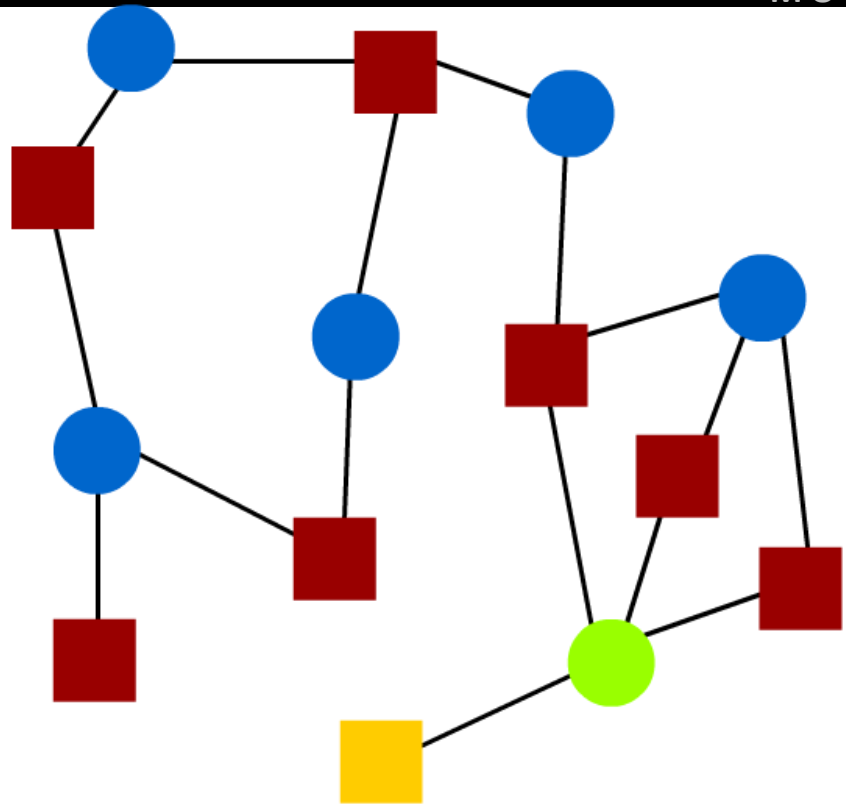
Comportamiento normal



Peticiones DNS



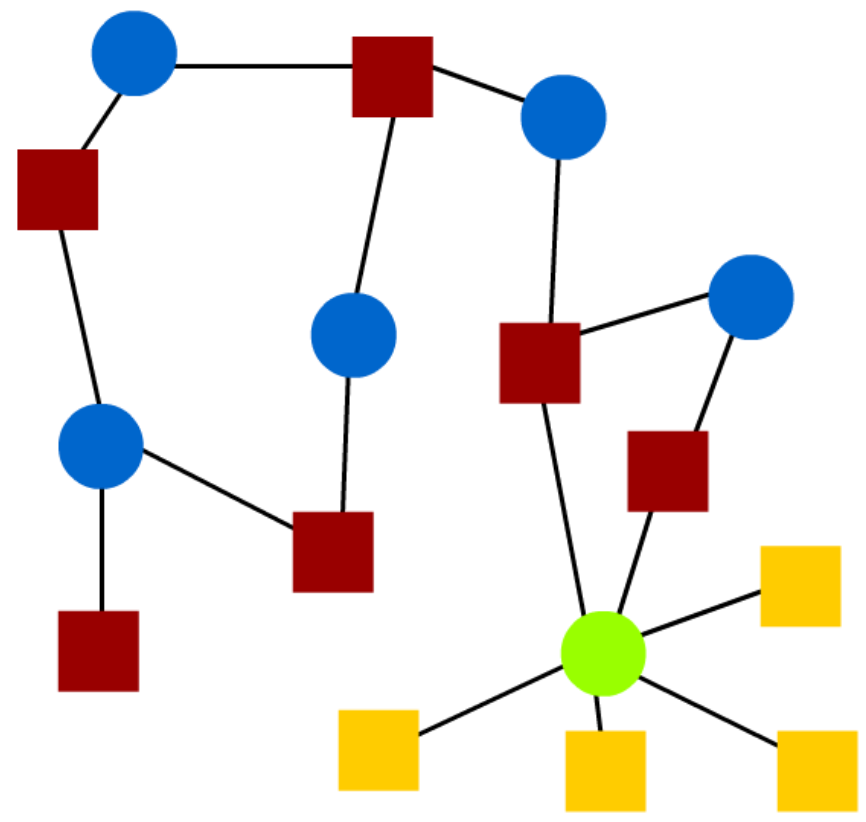
Comienza comportamiento anormal



Peticiones DNS



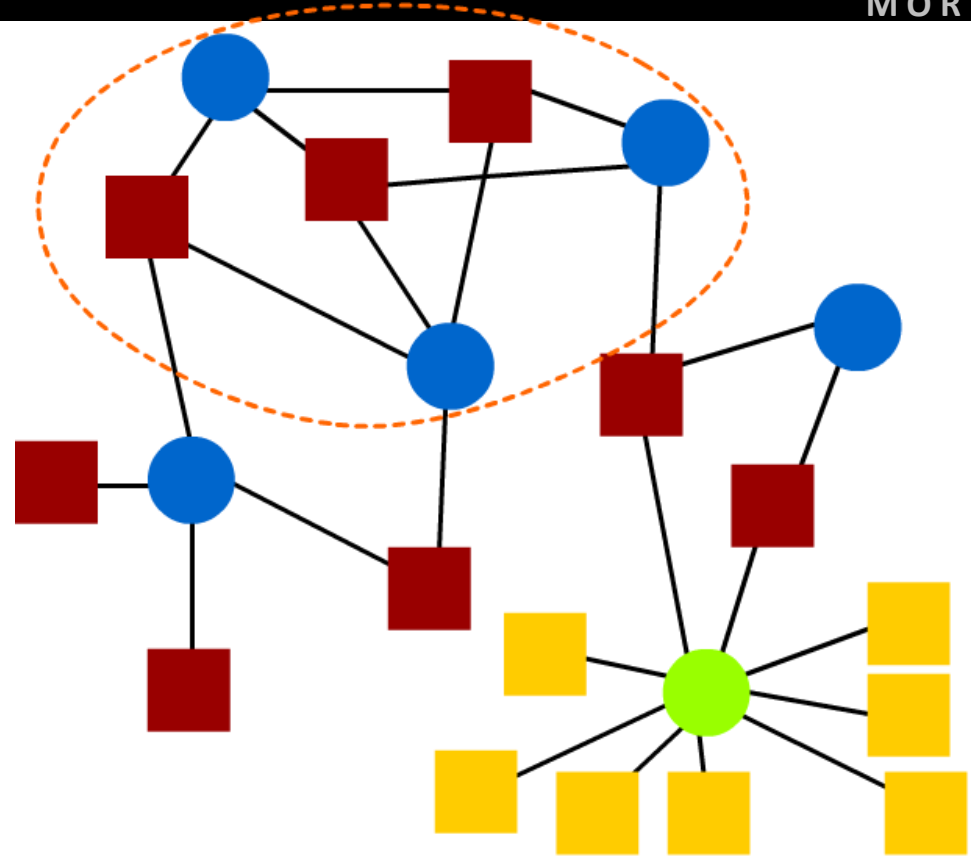
Continua comportamiento anormal



Peticiones DNS



Continua comportamiento anormal

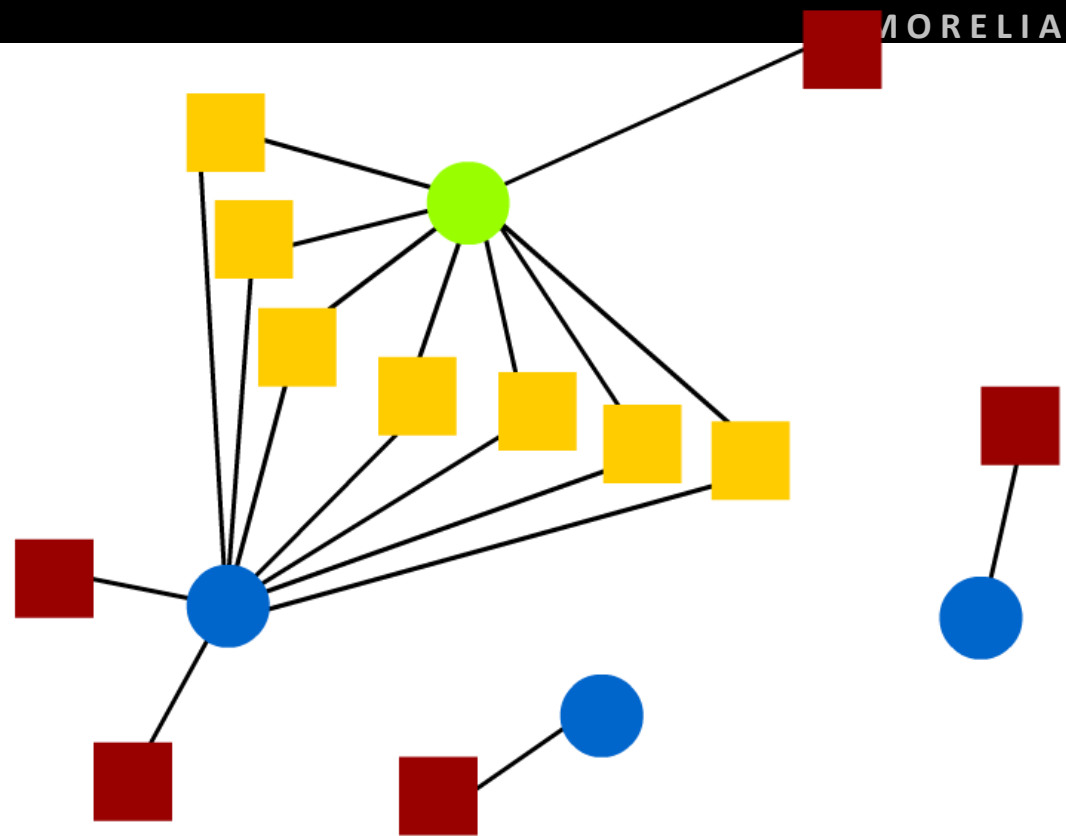


Peticiones DNS



Comportamiento claramente anormal

MORELIA • Reunión de Primavera



Peticiones DNS






Cuantificación de la Red de Interacción

MORELIA • Reunión de Primavera

- Representación numérica de cada Red de Interacción.
- Finalidad: Aplicación de herramientas estadísticas, minería de datos, etc.
- Valores a considerar:
 - Número de IPs
 - Número de Dominios
 - Número de Puertos Fuente
 - Número de Grupos
 - Número de IPs sin Grupo
 - Número de Dominios sin Grupo
 - Frecuencia del dominio más popular
 - Frecuencia de la IP más activa
 - Número de visitas del dominio menos popular
 - Número de solicitudes de la IP menos activa
 - Peso del Grupo más grande
 - Tamaño máximo de los grupos
 - Peso del grupo más pequeño
 - Tamaño del grupo más pesado
 - Entropía de IPs
 - Entropía de Puertos
 - Entropía de Dominios



Google/IBM Virtual Infrastructure Support

-  Login
-  Register








Login To Your Account

Username:

Password:

Have a signup token? [Register for your account](#)





Google/IBM Virtual Infrastructure Support

-  Home
-  My Account
-  Documents
-  Clusters
 -  **Main Cluster**
-  Help Tickets
-  Logout




Main Cluster

Namenode: mainclusternn.hipods.ihost.com
Jobtracker: 10.160.3.154
Number of nodes: 457
Hadoop version: 0.20.1 [download](#)
Useful files:

-  [Cygwin/UNIX SSH connect script](#) [Right click and Save As to download]
-  [Foxyproxy Config XML File](#) [Right click and Save As to download]
-  [Hadoop core-site.xml config file](#) [Right click and Save As to download]
-  [Hadoop mapred-site.xml config file](#) [Right click and Save As to download]

Related Documents

-  [Cluster Client Setup Instructions](#)

Status:	warning
Last Update:	4 minutes ago
HDFS:	131430.4 GB / 278937.6 GB
JobTracker:	0 / 820 Maps 8 / 411 Reduces



Hadoop job_201002100150_3552 on vm-10-160-3-154 - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Yahoo! Herramientas Ayuda

http://10.160.3.154:50030/jobdetails.jsp?jobid=job_201002100150_3552&refresh=30 traductor

Foro de elhacker.net ... AusCERT - Australia's... Herramientas inseguri... http://www.securityf... Kriptópolis | Criptogra... milw0rm - exploits : v... SecuriTeam™ - About... Hispasec - Seguridad ...

Hadoop job_201002100150_355... HDFS:/tmp/hadoop-n3o07/bitacora

Hadoop job_201002100150_3552 on vm-10-160-3-154

User: n3o07
Job Name: ITESM-CEM, Calcular Variables
Job File: hdfs://mainclusternn.hipods.ihost.com/tmp/hadoop-hadoop/mapred/system/job_201002100150_3552/job.xml
Job Setup: [Successful](#)
Status: Succeeded
Started at: Sat Mar 06 10:15:00 EST 2010
Finished at: Sat Mar 06 10:15:54 EST 2010
Finished in: 53sec
Job Cleanup: [Successful](#)

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00% 	29	0	0	29	0	0 / 0
reduce	100.00% 	29	0	0	29	0	0 / 0

	Counter	Map	Reduce	Total
Job Counters	Launched reduce tasks	0	0	29
	Rack-local map tasks	0	0	26
	Launched map tasks	0	0	29
	Data-local map tasks	0	0	3

Terminado

EELA-2



MOBILIA • Desarrollo Primavera



- It provides its users with a stable, well supported Grid Infrastructure based on **16 Resource Centres (RCs)** summing up to over **730 CPU cores** and **60 Terabytes** of storage space
- Deja la consolidación de 11 JRUs en México y 61 aplicaciones corriendo en distintas áreas

EELA-2 Projected Infrastructure

- 41 Resource Centres
- ~ 3000 CPU Cores
- ~700 TB of Storage

vera



Conformación del JRU-MX



MORELIA • Reunión de Primavera

- JRU Nacional conformado el 4 de Junio 2009
- Incorporación a la comunidad EELA-2

Conformación del JRU-MX



MORELIA • Reunión de Primavera

- Beneficios:
 - Oportunidades de colaboración en investigación
 - Acceso al uso de la plataforma
 - Capacitación para administradores
 - Gridificación/paralelización de aplicaciones
 - Consolidación de una plataforma GRID Nacional
 - Primer paso hacia una NGI (National Grid Initiative)

Conformación del JRU-MX



MORELIA • Reunión de Primavera

- Miembros iniciales:
 - UNAM, ITESM, USON, IPN-CIC, CUDI, UAEM, UMSNH, IPICyT, CICESE
- Requisitos:
 - Firma de MoU
 - Carta de reconocimiento por parte de CONACyT

Conformación del JRU-MX



MORELIA • Reunión de Primavera

- Sigüientes Pasos
 - Primer paso hacia una NGI (National Grid Initiative) con apoyo de la NREN (CUDI) mexicana.
 - La formación de NGIs en países de América Latina (Brasil, Argentina, Chile, Colombia, México,...) permitirá la creación de una iniciativa Latinoamericana en GRIDs (LGI) similar a EGI (European Grid Initiative), en estrecha relación con RedCLARA.

Congresos



MORELIA • Reunión de Primavera

- **ISUM 2010 (1st International Supercomputing Conference in México (ISUM))**, es el Congreso Internacional de Supercómputo en México, propuesto para llevarse a cabo en la ciudad de Guadalajara, Jalisco, del 2 al 5 de marzo de 2010.

Visión a corto plazo



MORELIA • Reunión de Primavera

- Infraestructura estable capaz de soportar aplicaciones de e-ciencia.
- Crecimiento de Infraestructura
- Aplicaciones (e-ciencia)
- Apoyo a la Investigación
- Docencia (Capacitación)



Proyectos Críticos para el 2010 - 2011

MORELIA • Reunión de Primavera

- LA-ROC, Nace de la comunidad HEP
- GISELA (Programa FP7), JRU-MX
- Red temática en Cómputo GRID (CONACyT), JRU-MX



Comunidad GRID y Supercómputo del CUDI: Principales Iniciativas Nacionales

Coordinador

Dr Luis A. Trejo, ltrejo@itesm.mx

- The Pacific Rim Application and Grid Middleware Assembly (PRAGMA) was formed in 2002 with NSF funding to establish sustained collaborations and advance the use of grid technologies in applications among a community of investigators working with leading institutions around the Pacific Rim.