Computational Science at Pacific Northwest National Laboratory:
Research Overview and Opportunities for Academia

Deborah Frincke, PhD
Deborah.Frincke@pnl.gov
CyberSecurity Chief Scientist Initiative Lead, Information and Infrastructure Integrity Initiative
PNNL is the Office of Science’s Most Diversified Laboratory

Business Volume ($M)

<table>
<thead>
<tr>
<th></th>
<th>FY05</th>
<th>Est. FY06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dept. of Energy</td>
<td>423</td>
<td>488</td>
</tr>
<tr>
<td>DHS</td>
<td>133</td>
<td>113</td>
</tr>
<tr>
<td>Other Agencies</td>
<td>93</td>
<td>109</td>
</tr>
<tr>
<td>Battelle Private</td>
<td>77</td>
<td>90</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>726</strong></td>
<td><strong>800</strong></td>
</tr>
</tbody>
</table>

Department of Energy
FY06 Est. 61%
(FY05 Actual 58%)
National and Homeland Security

Science-based solutions for:

- Protecting U.S. homeland, international borders, air and sea ports
- Monitoring nuclear treaties, preventing and detecting proliferation, countering WMD terrorism.

Training border agents to detect smuggling of weapons of mass destruction

Facilitating military transformation with innovative technologies

Field-deployable tools for early detection of chemical and biological weapons attacks

Visual analytics tools for intelligence, counter-intelligence, counter-terrorism and cyber security

High-sensitivity, high-precision instrumentation for proliferation detection
Computational and Information Sciences Directorate (CISD) has a key role in enabling PNNL mission-critical R&D to meet national challenges.
Focus of R&D within CISD

Data-Intensive Computing
Develop a new, data-intensive computing paradigm to extract knowledge from large, data-driven problems

Visualization and Analysis
Acquire knowledge through analytical reasoning facilitated by interactive visual interfaces

CyberSecurity
Advance the science behind sound and reliable software, systems architectures and infrastructures

Factoid: PNNL operates the largest open scientific computer in the DOE system – a Linux-based supercomputer from Hewlett Packard with 11.7+ teraflops peak performance and 6.9 terabytes of memory
Data-Intensive Computing

Definition

Data-Intensive Computing is required for any computational task where data intensity and the timeliness of the data are the rate-limiting factors to producing a time critical solution. This may be caused by a variety of factors:

- Data source is large and distributed
- Data volume is too great or complex
- Data structures being operated on are too large
- Data is complex and from varied sources.
CyberSecurity

Integrating Security within Systems and Processes

Defense & Response

Cyber Analytics & Policy

Trusted Engineering

Training & Education
Sensor Analytics—the Next-Generation Event Identification System uses teleseismic signatures and mathematical statistics to identify earthquakes and explosions.

Visualization and Analysis—the National Visualization & Analytics Center is setting the agenda for visual analytics with a new publication, *Illuminating the Path*.

- NVAC is expanding its influence through a new regional and government centers (RVACs and GVACs).
- Starlight™ and IN-SPIRE™ continue to innovate their technologies and lead as advanced visualization tools.
How to build in defenses
How the unit or system performs under attack
Prediction and forecasting

Selected activities:
- Security in depth suite of tools
- Vulnerability Assessment
- Insider Threat
- Static Visualization and Forecasting
- Real-time prediction
- Real-time massive sensor fusion

Active and proposed partnerships:
Sample Collaboration: Extended CyberCIEGE
Naval Postgraduate School and Pacific Northwest National Laboratory

- Educational game – originally devised by NPS and partnering company
- Extensions – joint funding from NCASSR
- NPS emphasis:
  - Educational content
  - Game design
- PNNL emphasis
  - Effectiveness and usability
  - Specialized scenarios

Prepared by C. Irvine 11 August 2005, Version 1.0
CyberSecurity Academic Outreach Program

Partnering with Academia

- “Semester at PNNL”
- CyberCorps students as interns
- Faculty Sabbaticals
- Joint Lecture Series
- Workshop sponsorship
- Joint research
A Sampling of Programs ...

- Computational Science Graduate Fellowships

- DOE SULI (Science Undergraduate Laboratory Internships)
  - [http://www.scied.science.doe.gov/scied/ERULF/about.html](http://www.scied.science.doe.gov/scied/ERULF/about.html)

- PNNL National Security Internship Program

- PNNL National Visualization and Analytics Center

PNNL Internship Programs for High School and Undergraduate Students

- 12 workforce and diversity programs for students
  - 3 for high school students
  - 9 for undergraduate students

- Students are placed with mentors for 8 weeks to over a year.

- Funding is from DOE Office of Science, Laboratory projects, Department of Homeland Security, National Science Foundation, etc.

Contact:
Karen Wieda
Team Lead - Workforce Development Programs for Students Science and Engineering Education
kj.wieda@pnl.gov
509-375-3811
Getting Involved

Specific Research Areas:
- Information or Visualization and Analysis: Kris Cook
- CyberSecurity: Deborah Frincke
- Data Intensive Computing: Deb Gracio

Internships and Student Placements
- University Partnerships: Kelly Sullivan
- Student Internships (general): Karen Wieda
- CyberSecurity Academic Outreach Program: Kristy Huston
- Student placement: Mariah Zabriskie

Sabbaticals, PostDocs and Research Partnerships
- Specific research areas (above)
IN-SPIRE

- Knowledge Signatures - Steerable Vector space
- Visualizations - Family of Interrelated Visualizations
  - Interaction and Discourse
    - Retrieval
    - Structuring
    - Evidence and Hypothesis Reasoning
    - Support for Repeating Activity
- Engineering / Deployment Suitability
  - Windows Platform
  - Approvals to Operate
  - Client/Server Lightweight System
  - Full Auditing, etc.
  - No Data Caching
  - Freedom for Questionable Infrastructure

Stand-alone version available
Data-Intensive Computing

- **The Morning Report** — data-intensive approach to airline safety through analysis of flight data—won the 2005 R&D 100 Award and the R&D 100 Editors Award for the Highest Impact Application for Safety.

- **Active Storage** — technology won the StorCloud Challenge at Supercomputing 2004 and 2005.

- **Computational Biology** — robust software solutions that address a number of key large-scale data-intensive computational problems in biology and bioinformatics.
Starlight Information Visualization System

Starlight - The Starlight software automatically organizes, characterizes, and integrates a variety of structured and unstructured information types, then generates easily interpretable, graphical representations of relationships among the data. This approach enables exciting and powerful new forms of information access, exploitation, and control.

Key Features:
- Visualization-oriented GUI
- Advanced information model
- Sophisticated query tools
- Information extraction tools
- Integrated GIS
- XML-based
- Windows NT/2000/XP platform

Principle Benefits:
- Information Integration
- Complexity Management
- Holistic Analysis
- Workflow Continuity
- Accelerated Interpretation
- Improved Understanding

http://starlight.pnl.gov