NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

INFORMATION TECHNOLOGY LABORATORY

Investigating Reliability and Robustness of Standards-Based Grid Computing Systems

Chris Dabrowski & Kevin Mills
National Institute of Standards and Technology

Presented at GGF15 in Boston, USA
October 6, 2005

Motivation

Vision: Future global information infrastructure will rely on emerging standards for Web Services and Open Grid Services Architecture

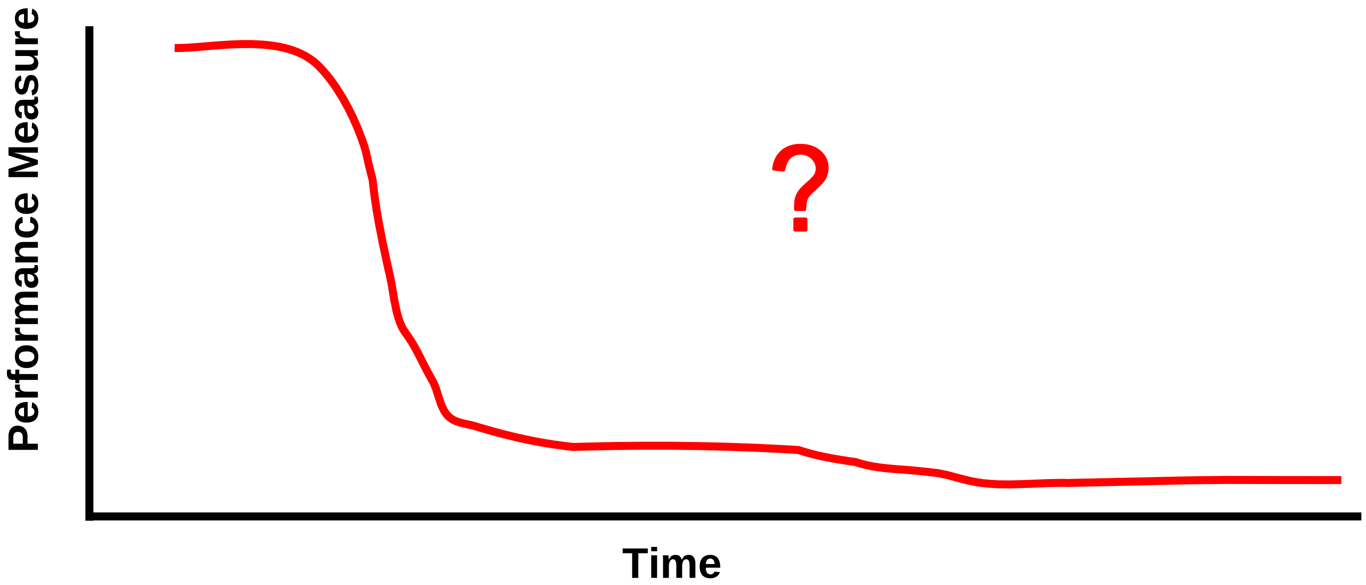
Question #1: – Will future distributed systems designed in conformance to Web Services and Grid standards achieve levels of robustness, scalability, and performance required for critical enterprise applications?

Question #2: – As industrial grid systems grow in size, can unplanned interactions among distributed components lead to emergence of undesirable patterns of behavior?

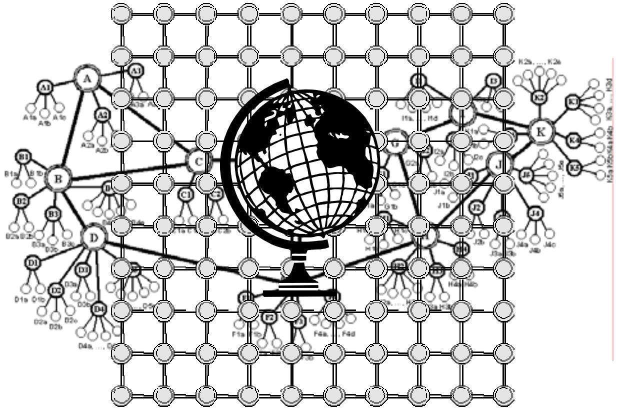
Question #3: – Can we identify areas in GGF specifications that might lead to implementation of operational Grids that are unreliable or that experience unexpected failures?

Possibility of emergent behaviors

- **Possible Concerns:** System designs may lead to interactions under failure conditions that result in emergent behaviors and unexpected performance degradations. The scaling of grid systems may, in and of itself, result in emergent behavior that adversely affects system behavior.



Investigating Emergent Properties in Standards-based Grid Systems



Customers

- Relevant industry standards groups (GGF, W3C, OASIS)
- Government users and backers of Grid technology (DoE, DISA, and NSF)

Goals

- Understand behavior of scaled SOA grids
- Investigate emergence in large-scale SOAs
- Improve related consortia specifications wrt reliability and robustness
- Investigate control mechanisms for shaping emergent behaviors

Technical Approach and Plans

Develop models of large-scale Grid systems

- Define architectures and components based on WS and GGF specs, use cases, failure scenarios, and recovery mechanisms; implement in SLX (Wolverine Software)
- Define metrics to reveal reliability, robustness, & scalability of Grid applications
- Execute experiments for large topologies and provide results to relevant standards consortia

Project phases

- ***Micro-model experiments***: 10^3 nodes 10^4 processes with components based on selected WS and GGF specs
- ***Macro-model experiments***: 10^4 nodes 10^5 processes of selected abstractions validated against micro-model.
- ***Decentralized Feedback Control Algorithms***: Experiments to evaluate candidate control algorithms that produce desirable overall system behaviors & apply to scenarios that exhibit un-desirable emergent behaviors

Model processes and components based on selected web and grid standards

Layered Component Architecture

- ***Network Layer:*** sites located in (x,y,z)-space; z axis indicates distance in hops to a simulated inter-site transmission delay;TCP-like simulated transport protocol; model node CPU delays, buffer & port capacity
- ***Basic Web Services:*** WS- Addressing, Messaging, Reliable Messaging (to be added).
- ***WSRF:*** WS- Resource Property, Lifetime, Notification, Topics, Service Group
- ***Grid Services:*** MDS v4, WS Agreement, DRMAA, *Grid security/access not included*

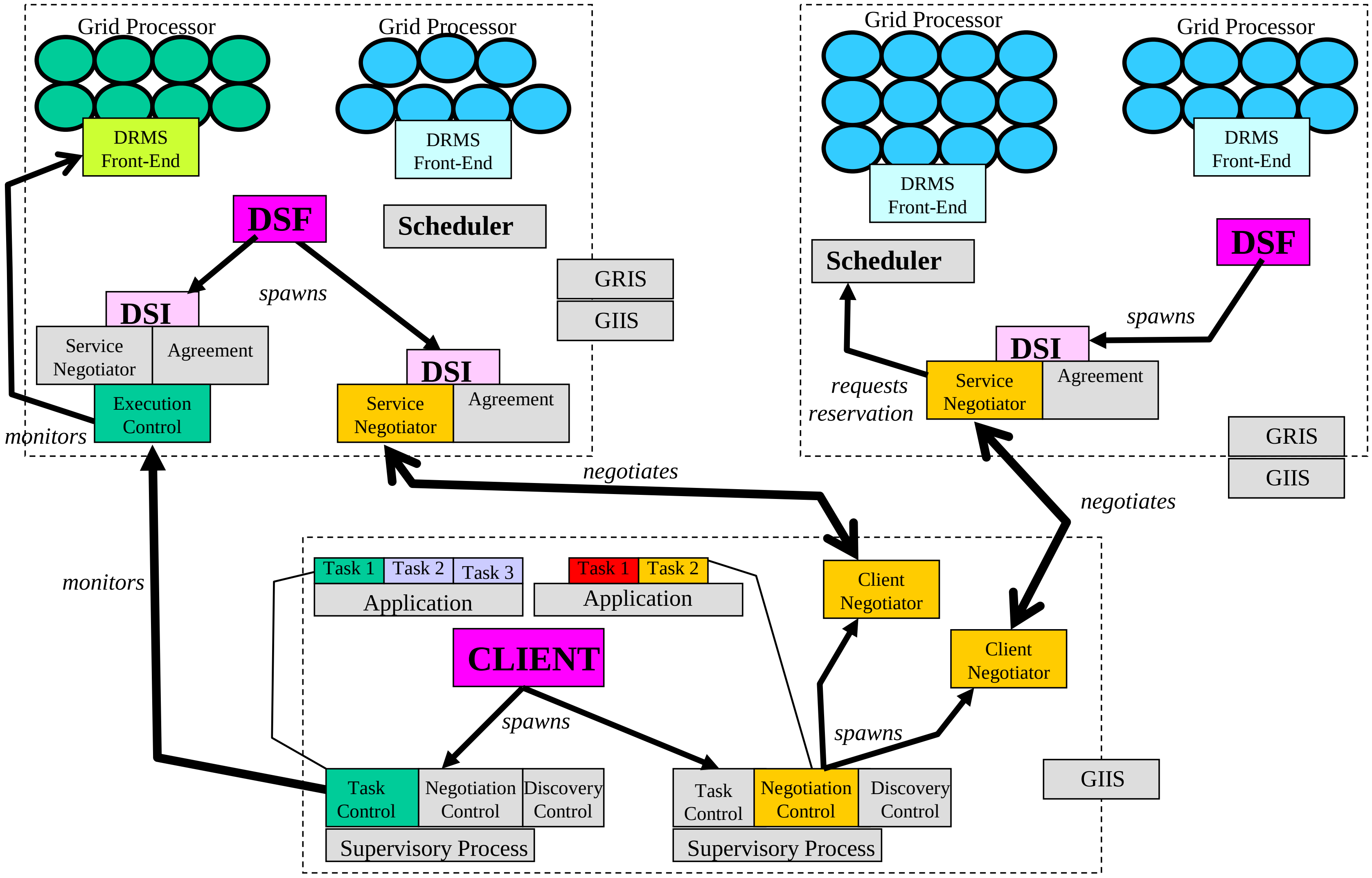
Major Grid Entities

- ***Service Providers:*** combine service & Agreement Factory WS resource (WS Agreement)
  - processor components (DRMS front-end): vectored S-computer or cluster
  - each site has scheduler for processors, GRIS, & GIIS
- ***Clients:*** discover providers by querying GIIS for required service types, rank discoveries by earliest availability (no economy scheduling, yet)**,** spawn WS resource negotiators to seek agreements, submit & monitor jobs.

Client Grid Applications

- ***Application types (5):*** workflows of 1-4 tasks, each with parallelizable sub-computations;
- ***Tasks:*** 3 types defined by required service type, task parallelism, & compute cycles
  - matched to processor component with suitable parallelism
  - Assume single agreement for resource requirements (no co-reservation)
- ***Workload:*** regulated by initial assignment of applications to clients; capacity determined by *client application requirements / (total processor capacity * time)*

Schematic showing operation of simulated grid



How does system respond to DNS spoofing under different negotiation strategies?

Negotiation Strategies

- ***Single-reservation request*** (SRR) - WS Agreement (sec. 9.2, Create Pending Agreement)
- ***Multiple-reservation request*** (MRR) - non-obligating offers, follow-up offers, and agreement superseding; based on draft WS-Agreement-Negotiation; no related agreements for co-allocation, etc. - yet?

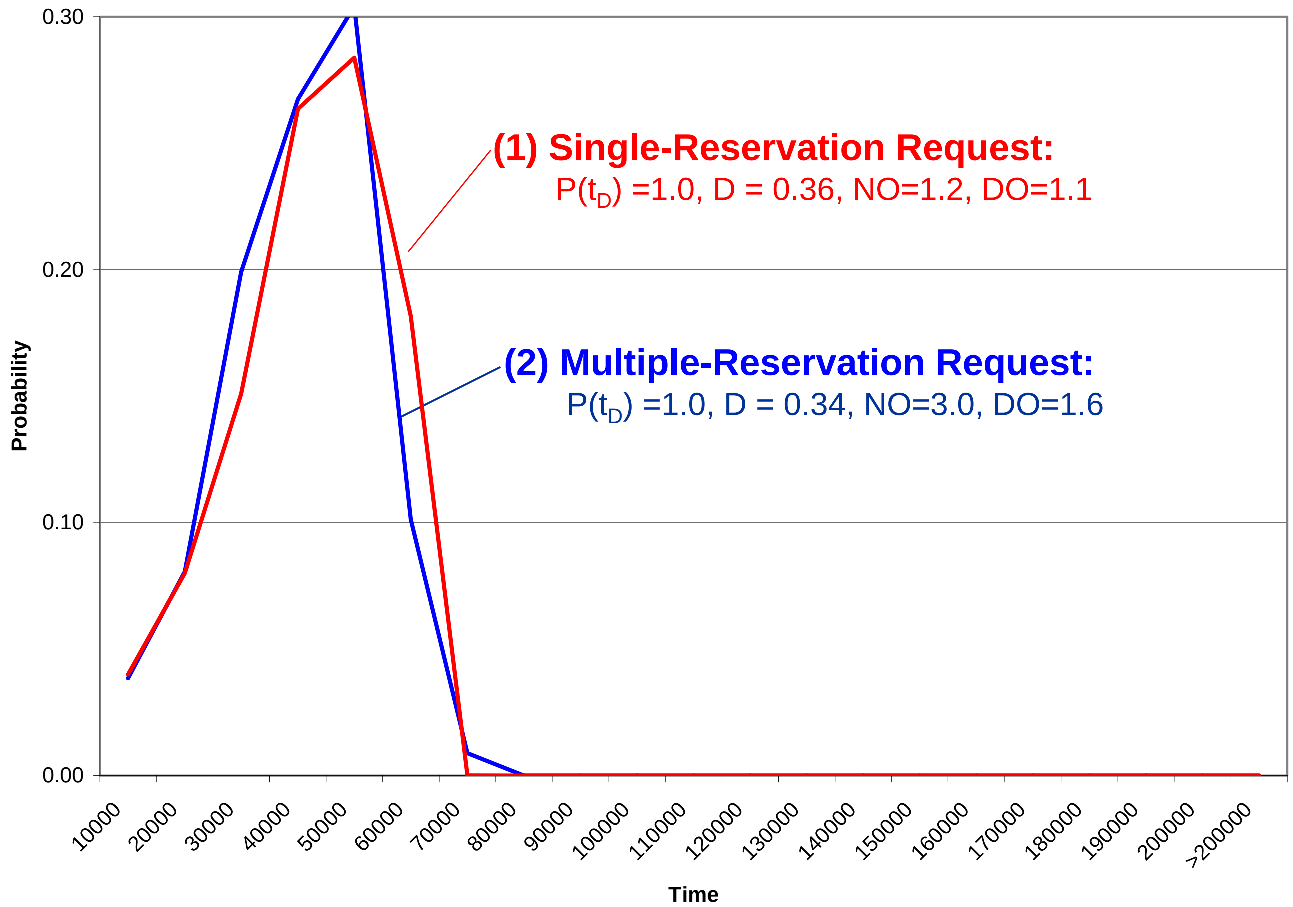
Introduction of Spoofing

- Miscreant alteration of DNS to redirect messages to false addresses
- ***Failure Response strategy:*** identify spoofers and do not repeat

Experiment

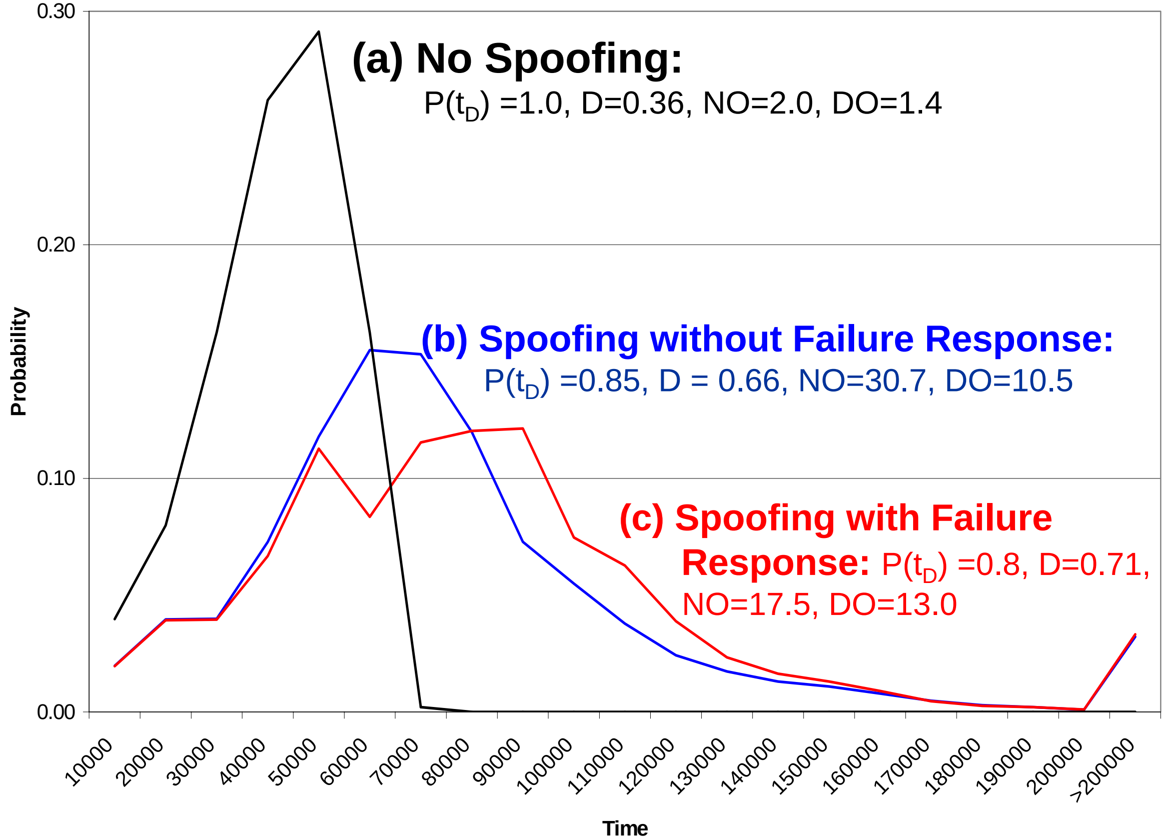
- At 50% capacity, 30 providers, 12 clients, 200 nodes - simulate 200,000s period (2+ days) with 24 hour deadline t_D using identical seed generation
- Model spoofing of service factories (p=50%); record performance with & without failure response
  - ***Primary metrics:*** probability of completion $P(t_D)$, application duration (D), negotiation (NO) & discovery overhead (DO) computed as multiple of min number of messages
  - ***Multidimensional time series analysis*** – select variables (number reservations created, number completions, etc.) to monitor over time.

Performance of Single-Reservation Request (SRR) and Multiple-Reservation Request (MRR) with no spoofing



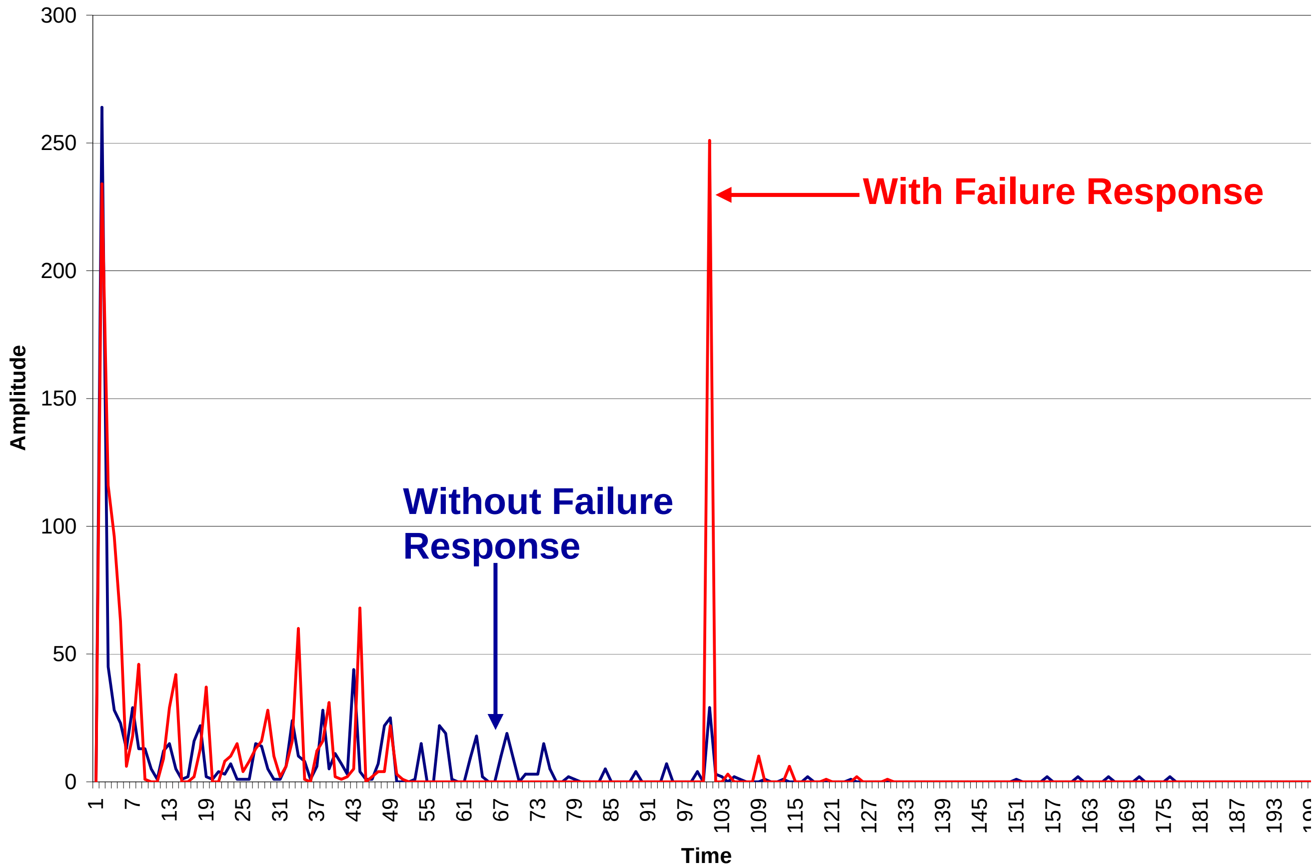
Comparative Probability Density Functions (PDFs) for application completion times and selected primary metrics for two negotiation strategies (over 200+ repetitions)

Performance degradation caused by spoofing activity in simulated Grid with 50% clients SRR and 50% MRR



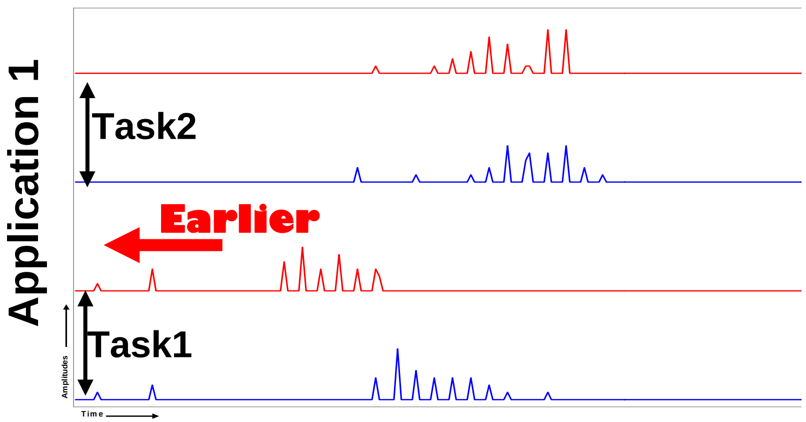
Comparative Probability Density Functions (PDFs) for application completion times and selected primary metrics given: (a) No spoofing (b) spoofing without failure response, and (c) spoofing with failure response. (200+ repetitions)

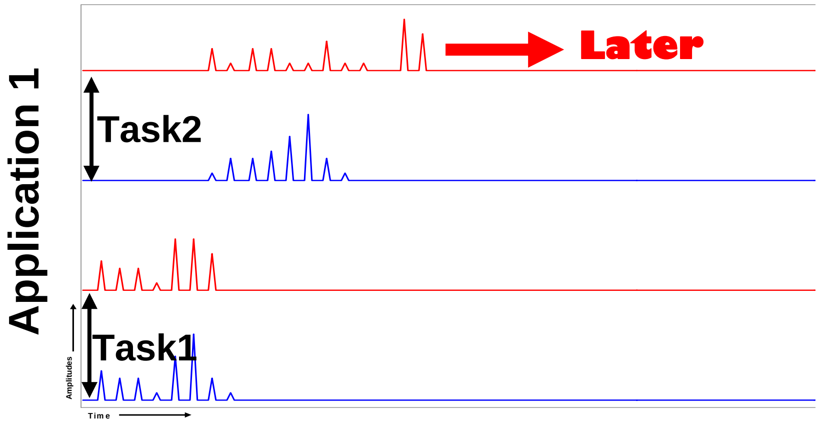
Time series in simulated Grid for reservations created with and without failure response strategy



Two Time Series: (a) Reservations Created without Failure Response and (b) Reservations Created with Failure Response for simulated grid with 50% clients SRR and 50% MRR.

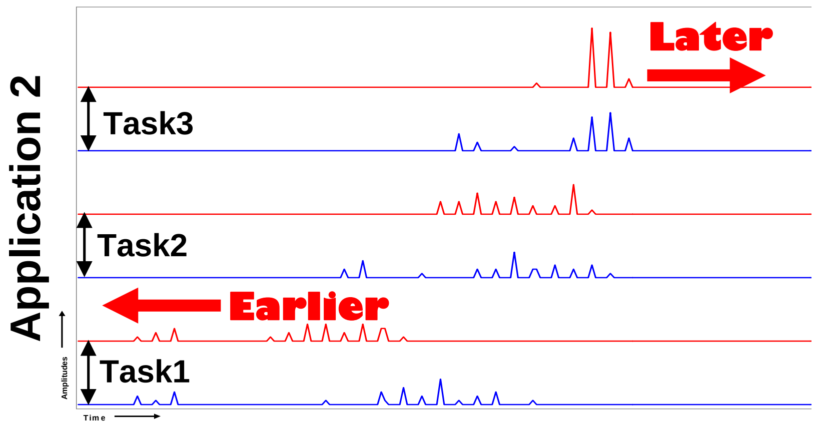
Time series for application/task completion for two application types with no failure response (lower blue) and with failure response (upper red)

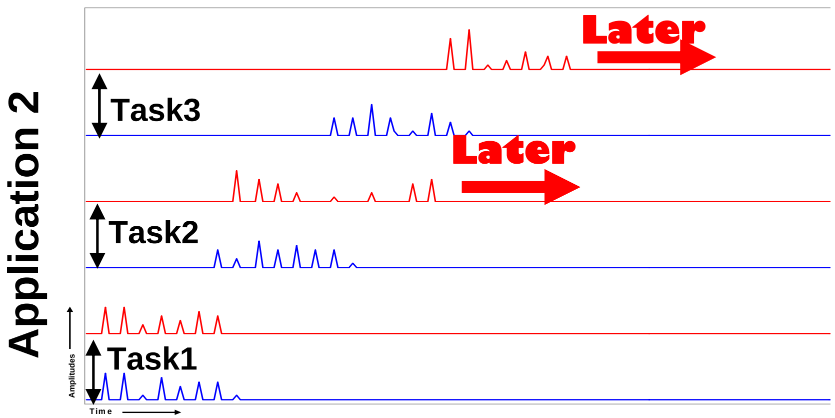




Single-Reservation Request (SRR)

Multiple-Reservation Request (MRR)





Conclusions and continuing/future work

- Under “normal” operating conditions SRR and MRR exhibit comparable performance with expected differences in overhead
- Not surprisingly, spoofing causes overall performance to degrade; both SRR and MRR degrade predictably
- However, unexpected further degradations occur when reasonable failure response action is introduced
  - “Reordering” of schedule results in overall increase in run times, (SRR clients are helped more at expense of MRR clients).
  - Emergent phenomena are difficult to explain and resolve with traditional metrics; require more sophisticated techniques such as multidimensional analysis to discern and explain causes
- Possible next steps
  - Incorporate additional GGF and WS specs
  - Formalize multidimensional analysis approach
  - Experiment with additional scenarios & scheduling algorithms to create economic model of grid computing, multiple/related agreements?

→ Should GGF have an RG on Grid Reliability and Robustness?

Web Services

[1] Daoi, Y. et al., “Managing Web server performance with AutoTune agents”, IBM SYSTEMS JOURNAL, VOL 42, NO 1, 2003, pp. 136-149

[2] Karp, A. “E-speak E-xplained”, HP Labs Technical Report, HPL-2000-101 August 7, 2000

[3] Chandra, P. et al. “Darwin: Resource management for value-added customizable network service”, Proceedings of the 6th IEEE International Conference on Network Protocols (ICNP ’98), 1998

[4] Huang, A.C. and Steenkiste, P. “Hierarchically-Synthesized Network Services”, unpublished draft, Department of Computer Science, Carnegie-Mellon, 2003

[5] Curbera, F., Silva-Lepe I. and Weerawarana,S., “On the Integration of Heterogeneous Web Service Partners”, Proceedings of the Workshop on Object-Oriented Web Services (OOPSLA ‘2001)

[6] Kıcıman, E., Melloul, L., and Fox, A. “Position Summary: Toward Zero-Code Service Composition”, position paper, Stanford University.

[7] Machiraju, V., Rolia, J., van Moorsel, A. “Quality of Business Driven Service Composition and Utility Computing”, HP Labs Technical Report 66, 2002

[8] Li, J., Yarvis, M., and Reiher, P. “Securing Distributed Adaptation”, Computer Networks, Vol. 38, No. 3, 2002, pp. 347-371

[9] Tosic, V., Mennie, D., and Pagurek, B. “Dynamic Service Composition and Its Applicability to E-Business Software – The ICARIS Experience”, Proceedings of the WOOBS (Workshop on Object-Oriented Business Solutions) Workshop (at ECOOP 2001), Budapest, Hungary, June 18, 2001, pp. 95-108

[10] Chandrasekaran S., Madden S., and Ionescu, M. "Ninja Paths: An Architecture for Composing Services over Wide Area Networks", CS262 class project, UC Berkeley (2000)

Survey References

Web Services (cont.)

[11] Felber, P., et al. Failure Detectors as First Class Objects, Proceedings of the International Symposium on Distributed Objects and Applications (DOA'99), IEEE Computer Society Press, September 5-7, 1999, p. 132

[12] Frolund S., et al. "Building Dependable Internet Services with E-speak", *Proceedings of the Workshop on Dependability of IP Applications, Platforms, and Networks*, June 26, 2000, held in conjunction with the 2000 International Conference on Dependable Systems and Networks, IEEE Computer Society, and also available as Hewlett-Packard Labs Technical Report 2000-78

[13] Chen, M., Kiciman, E., Fratkin, E., Fox, A., and Brewer, E. "Pinpoint: Problem Determination in Large, Dynamic Internet Services", Proceedings of 2002 International Conference on Dependable Systems and Networks *(DSN),* IPDS track, Washington, DC, June 23-26, 2002

[14] Web Services Resource Properties 1.2 (WS-Resource Properties), OASIS Working Draft 04, 10 June 2004

[15] Web Services Resource Lifetime 1.2 (WS-Resource Lifetime), OASIS Working Draft 03, 10 June 2004

[16] Web Services Service Group 1.2 (WS-ServiceGroup), OASIS Working Draft 02, 24 June 2004

[17] Web Services Base Notification 1.2 (WS-BaseNotification), OASIS Working Draft 03, 21 June 2004

[18] Web Services Topics 1.2 (WS-Topics), OASIS Working Draft 01, 22 July 2004

Grid Computing

[19] Tuecke, S., et al., Open Grid Services Infrastructure (OGSI), Version 1.0, Global Grid Forum, June 23, 2003

[20] Baker, M., Buyya, B., and Laforenza,D. "Grids and Grid technologies for wide-area distributed computing", Software Practice and Experience, 2002

Survey References

Grid Computing

[21] The Grid Report, The Commercial Implications of the Convergence of Grid Computing, Web Services, and Self-managing Systems, Bloor Research – North America, August 2002

[22] Juhasz, Z., Andics, A., and Pota, S. "Towards A Robust And Fault-Tolerant Discovery Architecture For Global Computing Grids", in Distributed and Parallel Systems – Cluster and Grid Computing, which contains the Proceedings of the Fourth Austrian-Hungarian Workshop on Distributed and Parallel Systems (DAPSYS 2002), Kluwer Academic Publishers, The Kluwer International Series in Engineering and Computer Science, Vol. 706, Linz, Austria, September 29-October 2, 2002

[23] Iamnitchi, A. and Foster, I. "On Fully Decentralized Resource Discovery in Grid Environments", Proceedings of an IEEE International workshop on Grid Computing, Denver, 2001

[24] Czajkowski, K., Fitzgerald, S., Foster, I., and Kesselman, C. "Grid Information Services for Distributed Resource Sharing", Proceedings of the 10th IEEE International Symposium on High Performance Distributed Computing (HPDC-10), IEEE Press, 2001

[25] Foster, I., et al., "A Distributed Resource Management Architecture that Supports Advanced Reservations and Co-Allocation", Proceedings of the International Workshop on Quality of Service, 1999

[26] Czajkowski, K., et al. "A resource management architecture for metacomputing systems", Proceedings of the Fourth Workshop on Job Scheduling Strategies for Parallel Processing, Spring-Verlag, LCNS 1459, 1998, pp. 62-82

[27] Jarvis, S., Thomas, N., and van Moorse, A., "Open Issues in Grid Performability", International Journal of Simulation, Volume 5, Number 5, pp. 3-12.

[28] Web Services Agreement Specification, Global Grid Forum, September, 2005.

[29] Distributed Resource Management Application API Specification 1.0, Global Grid Forum, June, 2004.

Survey References

Grid Computing (con't)

[30] R. Buyya and Manzur Murshed, "GridSim: a toolkit for the modeling and simulation of distributed resource management and scheduling for Grid computing", Concurrency and Computation: Practice and Experience, Vol. 14, 2002, pp. 1175-1220

[31] W. Bell, et al., "Simulation of Dynamic Grid Replication Strategies in OptorSim", Proceedings of the 3rd Int'l IEEE Workshop on Grid Computing (Grid'2002), Baltimore, 2002

[32] I. Legrand and H. Newman, "The Monarc Toolset For Simulating Large Network-Distributed Processing Systems", Proceedings of the 2000 Winter Simulation Conference, pp. 1794-1801, 2002

[33] Legrand, A., Marchal, L., and Casanova, H. "Scheduling Distributed Applications: The SimGrid Simulation Framework," Proceedings of the third IEEE International Symposium on Cluster Computing and the Grid (CCGrid'03), *Tokyo, Japan*. 2003.

[34] H. Casanova and L. Marchal, "A Network Model for Simulation of Grid Application", INRIA Research Report No. 4596, October 2002

[35] Aurora (a FreeNet Simulator) http://www.doc.ic.ac.uk/~twh1/academic/

[36] Baker, M., Grid Performance Modeling, Measurement, Analysis, and Control, Internal Report, University of Portsmouth, September 2004.

[37] Frey, J., Tannenbaum, T., Livny, M., Foster, I., Tuecke, S., "Condor-G: A Computation Management Agent for Multi-Institutional Grids," Proceedings of the Tenth IEEE International Symposium on High Performance Distributed Computing, San Francisco, CA, USA, August 7-9, 2001, pp. 55-67.

[38] Liu, X., Xia, H., Chien, A., "Validating and Scaling the MicroGrid: A Scientific Instrument for Grid Dynamics", Journal of Grid Computing, Volume 2, Number 2, pp. 141–161.

Grid Computing (con't)

[39] Ernemann, C. Hamscher, V., and Yahyapour, R., "Benefits of Global Grid Computing for Job Scheduling," Proceedings of the Fifth IEEE International Workshop on Grid Computing (GRID 2004), Pittsburgh, PA, USA, November 8, 2004, pp. 374-379.

[40] Ernemann, C. Hamscher, V., and Yahyapour, R., "Economic Scheduling in Grid Computing," in Job Scheduling Strategies for Parallel Processing, Edinburgh, Scotland, July 2002. Springer.

[41] Ernemann, C. Hamscher, V., Schwiegelshohn, U., Streit, A., and Yahyapour, R. "On Advantages of Grid Computing for Parallel Job Scheduling," Proceedings of the Second IEEE/ACM International Symposium. on Cluster Computing and the Grid (CCGRID2002), Berlin, May 2002.

[42] Gomoluch, J., and Schroeder, M., "Market-based Resource Allocation for Grid Computing: A Model and Simulation," Proceedings of the First International Workshop on Middleware for Grid Computing, Rio de Janeiro, Brazil, June 16-20, 2003, pp. 211-218.

[43] In, J., Avery, P., Cavanaugh, R., and Ranka, S., "Policy Based Scheduling for Simple Quality of Service in Grid Computing," Proceedings of the Eighteenth International Parallel and Distributed Processing Symposium (IPDPS'04), Santa Fe, NM USA, April 26-30, 2004, p. 23.

[44] He, X., Sun, X., Von Laszewski, G., "A QoS Guided Scheduling Algorithm for Grid Computing," Journal of Computer Science and Technology, Special Issue on Grid Computing, Volume 18, Number 4, 2003.

[45] Cooper, K. et al., "New Grid Scheduling and Rescheduling Methods in the GrADS Project," Proceedings of the Eighteenth International Parallel and Distributed Processing Symposium (IPDPS'04), Santa Fe, NM USA, April 26-30, 2004, p. 199.

Grid Computing (con't)

[46] Krothapalli, N. and Deshmukh, A. "Distributed Dynamic Allocation in Computational Grids," draft submitted for publication.

[47] Chen, H. and Maheswaran, M., "Distributed Dynamic Scheduling of Composite Tasks on Grid Computing Systems," Proceedings of the Sixteenth International Parallel and Distributed Processing Symposium (IPDPS 2002), Fort Lauderdale, FL USA, April 15-19 2002.

[48] Subramani, V., Kettimuthu, R., Srinivasan, S., and Sadayappan, P., "Distributed Job Scheduling on Computational Grids using Multiple Simultaneous Requests," Proceedings of the Eleventh IEEE International Symposium on High Performance Distributed Computing (HPDC-11 '02), Edinburgh, Scotland, July 24-26 2002, p. 359.

[49] Czajkowski, K., Foster, I., Kesselman, C., Sander, V., and Tuecke, S. "SNAP: A Protocol for Negotiation of Service Level Agreements and Coordinated Resource Management in Distributed Systems," Proceedings of the Eighth Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP'02), Edinburgh, Scotland, July 24 2002, pp. 153-183.

[50] MacLaren, J. et al. "Towards Service Level Agreement Based Scheduling on the Grid," Proceedings of the Workshop on Planning and Scheduling for Web and Grid Services, Whistler, British Columbia, Canada, June 3-7, 2004.

[51] Thain, D. and Livny, M., "The Ethernet Approach to Grid Computing," Proceedings of the Twelfth IEEE Symposium on High Performance Distributed Computing, Seattle, WA, June 2003.

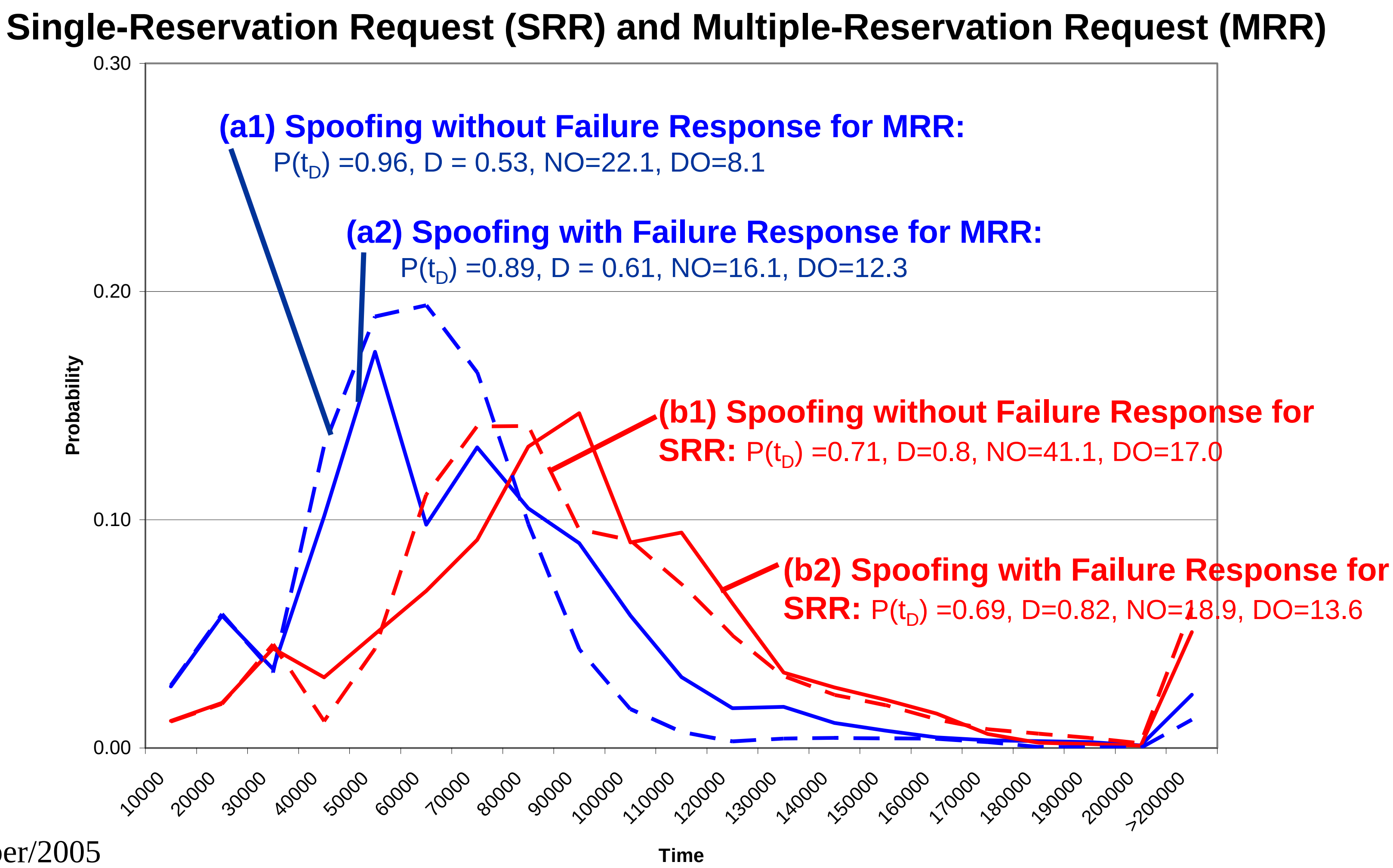
[52] Chang, B. Y., et al., "Trustless Grid Computing in ConCert," Proceedings of the Grid Computing - GRID 2002: Third International Workshop, Baltimore MD, USA, November 18, 2002, pp. 112-125.

Grid Computing (con't)

[51] Sherwani, J. Ali, N., Lotia, N., Hayat, Z. and Buyya, R. "Libra: a computational economy-based job scheduling system for clusters," Software: Practice and Experience, Volume 34, pp. 573-590, 2004.

EXTRA SLIDES

Performance degradation caused by spoofing activity decomposed by failure response and negotiation strategy



Research Group for Reliability and Robustness?

- ***Motivation:*** (*from previous slides*) As grid systems are increasingly commercialized and grow in size, they are likely to be subjected to volatile and uncertain conditions that endanger or severely degrade their effectiveness in everyday use.
- ***Question to be addressed:*** How can we determine that the web-service and grid standards currently being developed will enable large-scale grids to detect and overcome failures to provide a level of reliability and robustness needed for industrial and scientific purposes?
- ***RG Focus/Purpose:***
  - Identify issues related to reliability and robustness in grid computing systems designed in conformance to Web Services and Grid standards
  - Make recommendations, and explore methods, for improving reliability and robustness of standards-based grid systems developed for critical enterprise applications and production systems.