Predicting Global Failure Regimes in Complex Information Systems

Chris Dabrowski, Jim Filliben and Kevin Mills
June 19, 2012 NetONets 2012

- Decrease in probability of transition from Allocating_Minimum state (8) to Allocating_Maximum state (9)
- Increase in Probability of Transition from Allocating_Minimum state (8) to Transferring_Failure_Estimate state (10)

(a) Total Grants (Markov Simulation)
(b) Total Grants (Large Scale Simulation)
Overview of Our Past & Ongoing Research – with application to complex information systems, e.g., Internet, Clouds, Grids

What is the problem?

Why is it hard?

Four Approaches we are investigating:

1. Combine Markov Models, Graph Analysis & Perturbation Analysis
2. Sensitivity Analysis + Correlation Analysis & Clustering
3. Anti-Optimization + Genetic Algorithm
4. Measuring Key System Properties Such as Critical Slowing Down
Past ITL Research: How can we understand the influence of distributed control algorithms on global system behavior and user experience?


For more see: [http://www.nist.gov/itl/antd/emergent_behavior.cfm](http://www.nist.gov/itl/antd/emergent_behavior.cfm)
Ongoing & Planned ITL Research: How can we help to increase the reliability of complex information systems?

Research Goals: (1) develop design-time methods that system engineers can use to detect existence and causes of costly failure regimes prior to system deployment and (2) develop run-time methods that system managers can use to detect onset of costly failure regimes in deployed systems, prior to collapse.

Ongoing: investigating

b. Sensitivity Analysis + Correlation Analysis & Clustering

c. Anti-Optimization + Genetic Algorithm (AO+GA)

Planned: investigate run-time methods based on approaches that may provide early warning signals for critical transitions in large systems (e.g., Scheffer et al., “Early-warning signals for critical transitions”, NATURE, 461, 53-59, 2009).

June 19, 2012 NetONets
What is the Problem?

- **Problem**: Given a complex information system (represented using a simulation model), how can one identify conditions that could cause global system behavior to degenerate, leading to costly system outages?

Why is it Hard? – Reason 1

Determining causality is hard given that only global system behavior is observable. (in a complex system, global behavior cannot always be understood, even if behavior of components is completely understood)

For example, unexpected collapse in the mitigation probability density function of job completion times in a computing grid was unexplainable without more detailed data and analysis.

June 19, 2012 NetONets
Size of the search space!!

\[ y_1, \ldots, y_m = f( x_1|[1,\ldots,k], \ldots, x_n|[1,\ldots,k] ) \]

Model Response Space

Model Parameter Space

For example, the NIST Koala simulator of IaaS Clouds has about \( n = 125 \) parameters with average \( k = 6.6 \) values each, which leads to a model parameter space of \( \sim 10^{100} \) (note that the visible universe has \( \sim 10^{80} \) atoms) and the Koala response space ranges from \( m = 8 \) to \( m = 200 \), depending on the specific responses chosen for analysis (typically \( m \approx 42 \)).
Using simulated failure scenarios in a Markov chain model to predict failures in a Cloud

Example: Markov simulation and perturbation of a minimal s-t cut set of a Markov chain graph:

- Corresponds to software failure scenario involving multiple faults/attacks.
- Simulation identifies threshold beyond which increased failure incidence causes drastic performance collapse

→ **Verified in target system being modeled (i.e., Koala, a large-scale simulation of a Cloud)**

June 19, 2012 NetONets
**Sensitivity Analysis**: Determine which parameters most significantly influence model behavior and what response dimensions the model exhibits. Allows reduction parameter search space and identifies model responses that must be analyzed.

**Correlation Analysis & Cluster**: Determine response dimensions of a model.

---

Use 2-level, orthogonal fractional factorial (OFF) experiment design to identify the most significant parameters of your model.

\[
\begin{align*}
(2^3)^{32} & \rightarrow O(10^{789}) \quad [10^{60} = \text{atoms in visible universe}] \\
(2^3)^{23} & \rightarrow O(10^{221}) \\
(2^3)^{11} & \rightarrow O(10^{105}) \\
2^{11} & \rightarrow 2048 \\
2^{11} & \rightarrow 64 \\
2^{6-1} & \rightarrow 32
\end{align*}
\]

Use experiment design theory to reduce parameter combinations to 256.

Use sensitivity analysis results to identify six most significant parameters.

Select one response from each cluster to represent the dimension; we selected response with largest mean correlation that was not in another cluster.

Use correlation analysis and clustering to identify unique behavior dimensions of your model.


June 19, 2012 NetONets
MULTIDIMENSIONAL ANALYSIS TECHNIQUES

Principal Components Analysis, Clustering, ...

GENETIC ALGORITHM

Recombination & Mutation

Selection based on Anti-Fitness

Growing Collection of Tuples:

{(Generation, Individual, Fitness, Parameter 1 value, ..., Parameter N value)}
{(Generation, Individual, Fitness, Parameter 1 value, ..., Parameter N value)}
...
{(Generation, Individual, Fitness, Parameter 1 value, ..., Parameter N value)}

Anti-Fitness Reports

MODEL SIMULATORS

Parallel Execution of Model Simulators

List of parameters and for each parameter a MIN, MAX and precision.

Model Parameter Specifications

June 19, 2012 NetONets
A simple univariate example predicting power grid blackout in a human engineered system*


June 19, 2012 NetONets
Questions?

Suggestions?

Ideas?

Contact information about studying Complex Information Systems: {cdabrowski, jfilliben, kmills@nist.gov}

Contact information about Information Visualization: sressler@nist.gov

For more information see: http://www.nist.gov/itl/htmld/emerent_behavior.cfm
and/or http://www.nist.gov/itl/cloud/index.cfm

June 19, 2012 NetONets