Complex Systems Collected Images of the Month
March 2008 - June 2010

NIST Special Publication 500-286

Sandy Ressler, Editor
Information Technology Laboratory
Complex Systems Program Manager

June 2010

U.S. Department of Commerce
Gary Locke, Secretary

National Institute of Standards and Technology
Patrick D. Gallagher, Director
Complex Systems Collected
Images of the Month
(March 2008 - June 2010)

This document is a collection of twenty eight Complex Systems Image of the Month (IOM) illustrations. Each month as part of the Complex Systems Program in the Information Technology Laboratory an IOM is produced, posted to the Complex Systems web site and announced to the complex systems study group mailing list. It has proven to be an effective method of disseminating our research. All illustrations have been edited and composed by Sandy Ressler the program manager based on the work of the many staff participants in the Program.
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which, taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

Abstract—We have implemented tools for the visual investigation of network and computing grid simulation data. In this case, we look at the simulated behavior of the Abilene network. The network is represented by backbone, subnet, and leaf routers. These simulations enable us to consider how the characteristics of routers, connectivity, and data flow affect overall network behavior.

Attributes such as size, shape, color and brightness are easily distinguished by the user, and these can be presented in animated form to show time series data. These visualizations help us to look for patterns or properties that may be difficult to discern from the raw data.

Displays such as the one presented here have been implemented within a visualization tool that enables the researcher to interact with the data in a variety of ways. The user can select data items, assign them to a variety of visual attributes, probe the data for quantitative output, subset and zoom in on areas, data ranges, or times of interest, and so on.

Animation available at: math.nist.gov/mcsd/savg/vis/abilene/
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
Abstract—DiVisa is a multi-dimensional visualization tool developed for researchers to understand the behavior of their data. From raw data, the user can interact with the visualization in order to obtain different "points of views" and thus to extract more information from the data. Geometrical forms such as squares, ellipses or lines are associated with data and visual attributes such as position, size, shape, color, stroke are used to represent different dimensions. Indeed, the researcher can easily modify the associations between data items and visual attributes, apply mathematical functions on and between items, subset and zoom in on areas, data ranges, or times of interest, superpose curves with transparency to compare them, and animate the visualization to show time series data. Moreover, the program can read any kind of data (simulation, statistics, text or numeric, etc.), and converters have been implemented to read several data formats without need for reformatting.


One frame from an animated visualization of a simulation of the "Abilene" network. On the left panel, the innermost nodes are the backbone routers, the next ring represents the subnet routers, and the outermost ring shows the leaf routers. On the right panel, the time series curves of the routers (one is currently selected) are displayed. On the right side, different options available: numeric data display (top); superposed time series curves (middle); and diagram view (bottom).

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
In order to control congestion and maximize network utility, a variety of routing protocols can be used to allocate traffic. When multiple paths linking a source and destination are available but only a single path e.g. a shortest path or path of minimum cost is selected we have the Open Shortest Path First (OSPF) protocol. On the other hand, traffic could be allocated uniformly when the paths have the same cost. A spectrum of path specification strategies lies in between these extremes. The accompanying figure represents these possibilities in the case of a single source destination pair with 3 possible connecting paths. If Beta1, Beta2, Beta3 are the fractions (probability) of traffic assigned to paths 1, 2, and 3 respectively, then each allocation can be associated with a point in or on the equilateral triangle with an altitude height of 1. Here, Beta1, Beta2 and Beta3 are the lengths of perpendiculars from points on edges of the triangle.

The entropy associated with an allocation, \( H(\beta_1, \beta_2, \beta_3) = -\beta_1 \log(\beta_1) - \beta_2 \log(\beta_2) - \beta_3 \log(\beta_3) \) provides an important quantitative description of its degree of randomness. OSPF is associated with \( H = 0 \), and traffic is allocated according to one of the vertices of the triangle. This is the least robust allocation. Equal cost multipath allocation, assigned to the center of the triangle has the maximum value of \( H = \log(3) \), and is the most robust but it has reduced network utility. Bands of points with the same value of \( H \) (up to accuracy .001) are given the same color. As \( H \) decreases an oval emerges from the center point and grows until it touches the sides of the triangle at the maximum entropy for a network with 2 paths. Our theoretical studies predict a change in network dynamics for values of \( H \) smaller than this critical value and this is reflected in the figure where we see the ovals break up into discontinuous curves that shrink toward the triangle vertices as \( H \) tends to 0.

Cluster Analysis of System Responses for Congestion Control Algorithms

Abstract—This data is the result of simulations of the internet operating under TCP traffic. Each graph compares 45 responses for each of seven congestion control algorithms given one of 32 specified conditions. The Y axis in each graph represents the joint distance between all responses for a given algorithm or cluster of algorithms. The larger the distance between clusters the larger the difference in responses of the congestion control algorithms to the various input conditions. TCP flows consist of 3 phases of interest: 1) a connection phase; 2) slow start phase and 3) congestion control. If all of the work takes place in the slow start phase, as is true with no congestion, then the responses of the system using different congestion control algorithms will not vary, as evidenced by condition 12. Under increasing congestion the various algorithms may produce different system responses. This cluster analysis technique is used to identify IF and WHEN different responses exist but not WHAT the differences are. Specific differences are determined by further analysis.

A cluster analysis of the differences in one time period from among 45 responses. The X axis represents 7 congestion control algorithms. Each “small” figure represents one of 32 conditions applied to the 7 algorithms. Note that in condition 12 we see an identical response from each algorithm (due to no congestion on the network). Also note that in many of the conditions the largest distance graphed is from algorithm number 3 and the rest of the algorithms. Number 3 is the “FAST” TCP congestion control algorithm which performs markedly different.

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems

The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

Zanxin Xu

Study Emergent (Mis)behavior in Self-Organized Networks

Abstract—Route Flap Damping (RFD) is a self-adaptive mechanism deployed in the receiver-side to monitor and suppress persistently unstable routes. RFD reduces router processor load caused by instabilities and is intended to prevent sustained routing oscillations.

An emergent behavior is 1) a property that emerges due to interactions among the components of the system, 2) cannot be predicted by analysis of the components of the system.

Congestion, phase transition and synchronization are three types of emergent behaviors in self-organized application systems. Unexpected and unwanted system behavior is called emergent misbehavior.

A simulation of border-gateway protocol (BGP) routers in a 40 x 40 grid topology. Ten BGP routers are randomly selected to go down and up. The entire network becomes unusable and largely unreachable for approximately one full hour. The network-wide unreachability phenomenon is an emergent misbehavior. This emergent misbehavior arises from simple route flap damping (RFD). Interestingly RFD was recommended to retard excessive updates caused when flakey components go up and down often over short periods. Our simulations show that using RFD can lead to prolonged unreachability in the network when a small subset of BGP routers fail for a brief period of time. Thus, this is a case where a mechanism (RFD) put in for a good reason has unintended consequences.

The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

Each column is colored coded as follows:

(1) Blue means there is no statistically significant difference between the outlier and the others (i.e., the Grubbs statistic is below 2.08).
(2) Green means there is a statistically significant difference and the outlier is above the average.
(3) Red means there is a statistically significant difference and the outlier is below the average.
(4) Black – appears on Image 2 and Image 3 - means that there is no outlier - all have the same value.

From the first image, we see that algorithm 3 (FAST) is the positive outlier in 30 of the 32 conditions and is statistically significant in 21 of these conditions. The probability that one algorithm would be an outlier by chance alone in this many conditions is well below 0.01; thus, the effect shown on the image is fairly certain to be real.
Bayesian Belief Nets applied to Complex Systems

This is a graphical representation of a Bayesian network made up of five robots whose function is to perform one of two actions. They could either perform a seismic test to help assess the likelihood of oil being present or they could drill for oil. The machines are located in a large field where the a priori probability of oil is uniform throughout. It is likely that neighboring positions would have similar ground conditions and thus that the presence of oil at a particular location would increase the probability of oil at neighboring locations. This can be represented by the conditional probability tables of the Oil nodes. This is the origin of “cooperation” of the units of this complex system. An observation in the form of a test result at a single location is propagated through the system, updating via Bayes Theorem the marginal probabilities of oil at all of the locations. This has an effect on the choice of optimal action for many if not all of the units and thus is at the heart of the collective behavior of the system.

Using Markov Chain Analysis to Study Dynamic Behavior in Large-Scale Grid Systems

A piece-wise homogeneous Discrete Time Markov chain can be used to model performance of a large-scale computing grid and predict how performance changes in alternative scenarios. Markov chain simulation provides a rapid, scalable representation of system dynamics and can be generated at substantially lower computing costs than detailed simulation models.

Increasing the probability for this transition results in fewer tasks completing, they spend more time negotiating (as illustrated below).

A Markov chain model can be perturbed by altering selected transition probabilities to predict the effect of changed behaviors, failures, and other events on overall system performance. An example of perturbation analysis is shown in second graph above.

The variations illustrate individual simulation runs testing different probability transitions of network states. The variations illustrate results of individual Markov chain simulation runs that test different values for probabilities of transition between states of grid system tasks. Testing this large number of variations is impractical using the large-scale network simulation (shown via the red triangle portion of the graph).


A piece-wise homogeneous Discrete Time Markov chain can be used to model performance of a large-scale computing grid and predict how performance changes in alternative scenarios. Markov chain simulation provides a rapid, scalable representation of system dynamics and can be generated at substantially lower computing costs than detailed simulation models.
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

The key to faithful mathematical models of TCP throughput is a good model for buffer overflow probability. The graphs above, generated from NS2 simulation data, illustrate how spikes in buffer content on a very short time-scale (about half a round-trip time) correspond to multiple overlapping batch packet arrivals. Simply put when large numbers of simultaneous flows are active the TCP buffers fill up resulting in packet losses and a decrease in throughput. The batch packet arrivals are mostly ignored in current TCP models and may be one of the main sources of model inaccuracy.
We examine the effectiveness of creating network models based on the s-metric. Through a series of computational experiments, we compare how well a set of common structural network metrics are preserved between instances of the autonomous system Internet topology and a series of random models with identical degree sequences and similar s values. We demonstrate that creating models based on the s metric can produce moderate improvement in structural characteristics over strictly using degree distribution. The more interconnected a network is, the more robust it is to failure.

We tested how well four structural metrics of the AS topology (a topology similar to the backbone of the Internet) were preserved:
1. Diameter
2. Number of biconnected components
3. Minimum vertex cover size
4. Number of spanning trees

We examine the effectiveness of creating network models based on the s-metric. Through a series of computational experiments, we compare how well a set of common structural network metrics are preserved between instances of the autonomous system Internet topology and a series of random models with identical degree sequences and similar s values. We demonstrate that creating models based on the s metric can produce moderate improvement in structural characteristics over strictly using degree distribution. The more interconnected a network is, the more robust it is to failure.

\[ s(G) = 70 \quad s(G) = 79 \]

note how the hub nodes (in red) are more interconnected in the graph on the right with the higher s Metric value.

The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

FAST is one of several proposed replacements for the standard TCP congestion-avoidance algorithm. The general aim of such replacement algorithms is to increase the ability of TCP to exploit higher transmission speeds becoming available within the Internet. Using simulation, under congested conditions FAST was found to exhibit a higher retransmission rate (including connection-request packets) that reduced user throughput on flows transiting congested areas and that increased the number of flows pending in the connecting state. As demonstrated above, this behavior appears to arise from rapid oscillations in congestion-window size when FAST flows transit routers with insufficient buffers to accommodate the flow volume. Practical implications of this behavior include: (1) flows take longer to connect; (2) flows take longer to complete; (3) user throughput lower for flows transiting congested areas; (4) fewer flows complete. For example, over a 25-minute period FAST completed from $10^5$ to $10^7$ fewer flows (depending on conditions) than other congestion-control algorithms.

One of the active areas in network science has been the analysis of graphs denoting the structure of the World Wide Web. These graphs denote the organic nature of information flow and ideas across hyperlinks, connecting one page to another. It is the result of a cooperative and emergent construction, not the work of a single design, that make such graphs interesting and instrumental in developing the basic theory in the network science (e.g. scale-free networks, node clustering, and small-world characteristics).

The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems


This figure shows a subgraph (1237 nodes & 9172 edges) of a particular domain (the computer science department at George Washington University), with each node being an individual web page (or other document comprising a unit of information). To create such data, we have developed our own web crawlers that record individual time stamps, and employ cryptographic hashing functions (developed at NSA and NIST) to ensure the unique identity of each pages’ contents. This resolves common aliasing problems (where the contents of two different URLs are the same), ensures that quality graph data is maintained, and allows for validation and verification of each page. These graphs are then prepared for immersive visualization on the NIST RAVE environment (http://math.nist.gov/mcsd/highlights/rave.html), using a three-dimensional projection system, where researchers can walk around and examine these complex structures interactively. Graph layout was performed using algorithms from Yifan Hu of AT&T Labs.
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
Extending Markov Chain Analysis for Application to Additional Problems

In the January, 2009 Image of the Month, we showed that Markov chain simulation provides a rapid, scalable representation of system dynamics in a computing grid and can be used to predict system behavior at substantially lower computing costs than detailed simulation models. Expanding on this, we show how Markov chain analysis can be used to model system operation in a new domain: congestion control algorithms for network data flows.

A piece-wise homogeneous Discrete Time Markov chain can be used to simulate flow completion in a network.

Markov chain simulation provides an accurate approximation of large-scale simulation as evidenced by the curves produced by the respective simulations. However, Markov chain analysis is more than two orders of magnitude faster.
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
In our study of TCP congestion control with allocation of paths joining a source and destination, the entropy of the route distribution is a key parameter. For a simple network with three possible paths joining a source and destination, let $a_k$ be the fraction (non-negative) of traffic assigned to route $k$, $k=1,2,3$. We must have $a_1 + a_2 + a_3 = 1$. Every such point $(a_1,a_2,a_3)$ can be represented as a point in the equilateral triangle with unit altitude. The figure shows the surface over the triangle defined by the equation: 

$$H(a_1,a_2,a_3) = -a_1 \log(a_1) - a_2 \log(a_2) - a_3 \log(a_3),$$

where the height of a point on the surface above the triangle is the value of $H$ for the point in the triangle representing $(a_1,a_2,a_3)$. $H$ is the entropy function. The top of the surface is the maximum value of the entropy, i.e. $\log 3$ which corresponds to the point $(1/3,1/3,1/3)$ whose image in the triangle is the pink dot in the center.

The surface illustrates the changes in the level curves of $H$ (i.e. points in the triangle with a constant entropy value) as $H$ decreases from $\log 3$ to 0. The level curves are closed ovals for $H< \log 3$ and grow larger until $H=\log 2$. The pink circle on the surface is the intersection of the surface with the plane of height $\log 2$ above the triangle. The corresponding level curve is an oval that is tangent to boundaries of the triangle. If $H>\log 2$ the level curves change. For such values the level curve is piecewise continuous and includes part of the boundary of the triangle. Points on the boundary correspond to distributions where one of the $a_i=0$, therefore they can only be realized by 2 routes rather than 3. These figures highlight the distributions realized by 3 routes and this accounts for the gaps in the surface. The vertices of the triangle (in red) correspond to the distribution that assigns a value 1 to one of the routes and 0 to the other two. The entropy of such a distribution is $H=0$. 

The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior. Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
This figure shows a temporal model of a subset of the autonomous system (AS) topology of the Internet. The red edges denote the original “core” graph from a November 2007 instance of the AS topology. The black edges are the additional connections added to the network model to represent a future time period.

Kevin Mills
James Filliben

A Glimpse at Internet Congestion-Control Analysis

This figure shows that (under the conditions simulated) changing from fewer to more buffers has a larger effect when network speed is high and propagation delay is long. We can assign the average rank for each condition to the vertex of a cube, where each vertex represents a specific combination of settings for propagation delay, network speed and buffer size. This makes intuitive sense because more packets could potentially be inside the network when speed and propagation delay increases, thus, a higher proportion of the increased buffers would likely be occupied.


This is a small part of an extensive study that describes a coherent set of modeling and analysis methods to investigate the behavior of large distributed systems. The methods are applied to compare several proposed Internet congestion-control mechanisms operating under a wide range of conditions. The study provides insights and recommendations regarding various congestion-control algorithms. The study also evaluates the modeling and analysis methods adopted.

The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
Christopher Dabrowski
Fern Hunt

**complex systems**

February

**Improving the Efficiency of Markov Chain Analysis**

In this new approach, we use graph theory concepts to identify states and state transitions in Markov chain graphs where perturbation of transition probabilities produces drastic changes in system behavior. This approach exhibits as good or better success ratio as the earlier perturbation algorithm reported in [Jan 09 IOM], but with an additional reduction in computing costs of 1-2 orders of magnitude.

Graph is effectively disconnected at this point by raising probability of staying in the Discovering state to 1

Graph is effectively disconnected at this point by lowering probability of transition of Negotiating → Monitoring to 0


The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: [www.itl.nist.gov/ITLPrograms/ComplexSystems](http://www.itl.nist.gov/ITLPrograms/ComplexSystems)

In this simple example, the Markov chain can be represented as a directed graph with two paths (shown in red) from the Initial State to the absorbing state, Tasks Completed. If individual states, such as Discovering or Monitoring (circled) or individual state transitions corresponding to edges in a graph, such as Negotiating to Monitoring, are removed, then both paths to the absorbing states are cut. In graph theory, a set of edges in a graph, which if removed, would disconnect all paths between two vertices (or points), is referred to as an s-t cut set.

States and state transitions whose removal disconnects all paths between the Initial state and Tasks Completed can predict where perturbation is most likely to drastically change system performance, as figures 1 & 2 show.

In this new approach, we use graph theory concepts to identify states and state transitions in Markov chain graphs where perturbation of transition probabilities produces drastic changes in system behavior. This approach exhibits as good or better success ratio as the earlier perturbation algorithm reported in [Jan 09 IOM], but with an additional reduction in computing costs of 1-2 orders of magnitude.


In this simple example, the Markov chain can be represented as a directed graph with two paths (shown in red) from the Initial State to the absorbing state, Tasks Completed. If individual states, such as Discovering or Monitoring (circled) or individual state transitions corresponding to edges in a graph, such as Negotiating to Monitoring, are removed, then both paths to the absorbing states are cut. In graph theory, a set of edges in a graph, which if removed, would disconnect all paths between two vertices (or points), is referred to as an s-t cut set.

States and state transitions whose removal disconnects all paths between the Initial state and Tasks Completed can predict where perturbation is most likely to drastically change system performance, as figures 1 & 2 show.

In this new approach, we use graph theory concepts to identify states and state transitions in Markov chain graphs where perturbation of transition probabilities produces drastic changes in system behavior. This approach exhibits as good or better success ratio as the earlier perturbation algorithm reported in [Jan 09 IOM], but with an additional reduction in computing costs of 1-2 orders of magnitude.


In this simple example, the Markov chain can be represented as a directed graph with two paths (shown in red) from the Initial State to the absorbing state, Tasks Completed. If individual states, such as Discovering or Monitoring (circled) or individual state transitions corresponding to edges in a graph, such as Negotiating to Monitoring, are removed, then both paths to the absorbing states are cut. In graph theory, a set of edges in a graph, which if removed, would disconnect all paths between two vertices (or points), is referred to as an s-t cut set.
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: [www.itl.nist.gov/ITLPrograms/ComplexSystems](http://www.itl.nist.gov/ITLPrograms/ComplexSystems)
Matching Observed Lengths to Predicted Structure

We are interested in generating a library of probable matchings of observed alpha helix lengths to 1D structure. The idea is to provide a first attempt at understanding the 3D structure.

Measuring distance between orders

We used Kendall-tau distance to compare orders. As shown our heuristic method often produces better results to the actual order than an optimal order.

Extending a base order to a library

Starting from a base order π, we create a random permutation σ with a probability proportional \((1-p)^{d_k(\tau,\nu)}\) for a parameter \(p\).

Testing the method

We generated 200 random permutations and measured distance to actual protein order. The results show that every time either the exact or a close order was produced.

Because of the complexity in determining the 3D structure of a protein, the use of partial information determined from experimental techniques can greatly reduce the overall computational expense. We examined the problem of matching observed lengths of alpha helices to their predicted location on a protein’s amino acid sequence. This potentially can be a first step towards determining the 3D structure of the protein.

We showed that the effort in finding optimal potential solutions does not seem to be worth the computational expense. In particular, we showed evidence that, because of the uncertainty of the helix prediction, the optimal coverings can be relatively distant from the actual ordering on the protein. Instead, we introduced a simple greedy heuristic for estimating the order. Using this heuristic as a starting point, we chose random orders around it using the BubbleSearch method of Lesh and Mitzenmacher. When compared to the actual orderings, this method was able to either find the correct ordering or an ordering that was very close. Thus, we believe that our method is a fast and efficient algorithm for determining a set of potential placements of helix lengths onto a protein sequence.

The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interconnected, interacting entities which, taken together, exhibit macroscopic behavior that is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems
The Complex Systems Program is part of the National Institute of Standards and Technology’s Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: www.itl.nist.gov/ITLPrograms/ComplexSystems

An overview of the types of mathematical tools and statistical methodologies used in the Innovations in Measurement Science project “Measurement Science in Complex Information Science”. See the NIST Special Publication 500–282 “Study of

For more information see: "http://www.nist.gov/itl/antd/emergent_behavior.cfm"
