Using Sensitivity Analysis to Identify Significant Parameters in a Network Simulation

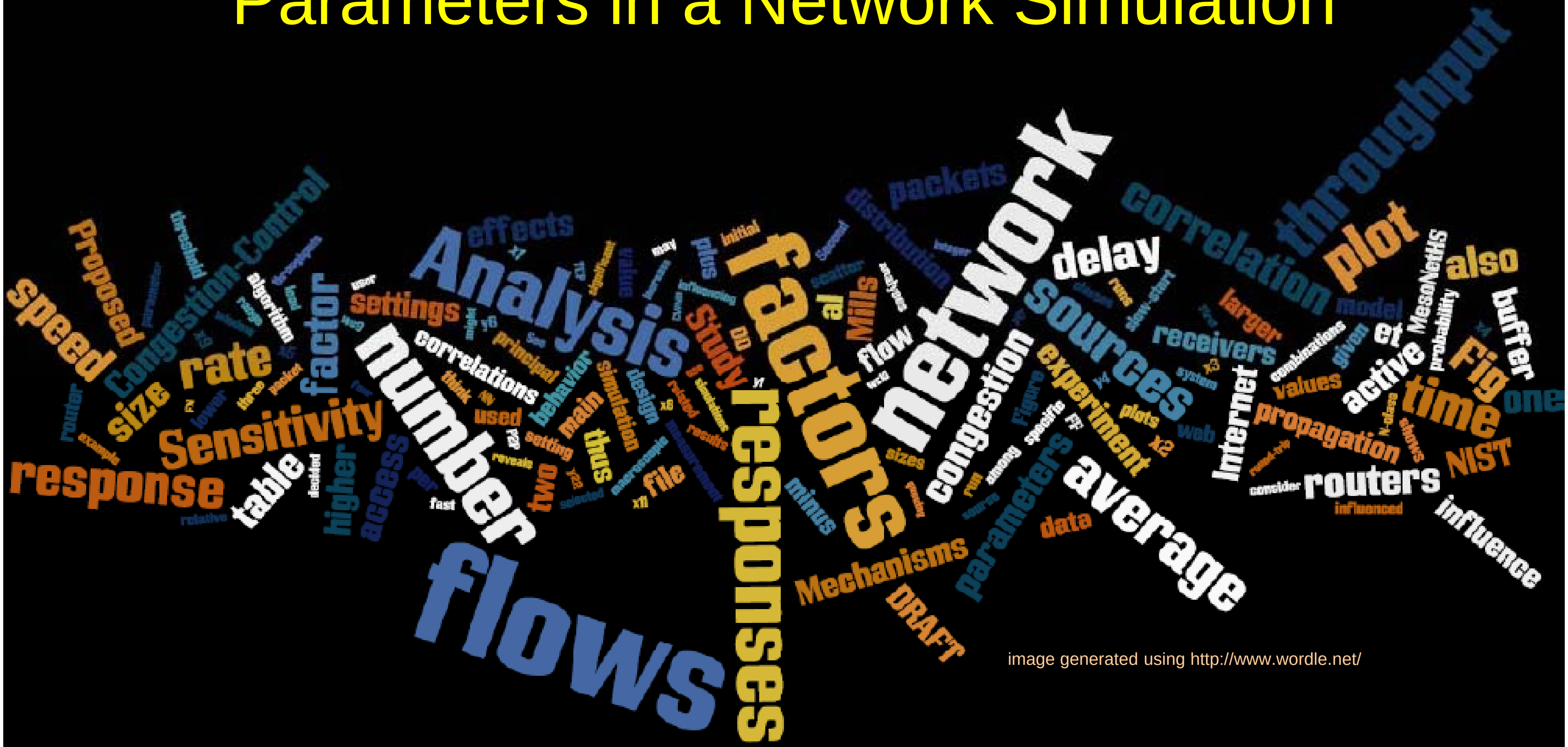


image generated using http://www.wordle.net/

Kevin Mills & Jim Filliben, NIST

Winter Simulation Conference – Dec 6, 2010

Outline

- Goal – Problem – Solution
- Scale Reduction: Theory and Practice
- Overview of the 20 Parameters in *MesoNet*, a TCP\IP Network Simulator
- 2-Level Per Parameter Experimental Design
  - Theory
  - Application to *MesoNet*
- Selected Analysis Techniques
  - Main Effects Analysis
  - Two Factor Interaction Analysis
  - Tabular Summary Analysis
- Relative Importance of *MesoNet* Parameters
- Findings and Conclusions

Goal – Problem – Solution

- **Goal** – compare proposed Internet congestion control algorithms under a wide range of controlled, repeatable conditions, as simulated by selecting combinations of parameter values for *MesoNet*, a mesoscopic network model

- **Problem** – how to determine key parameters influencing behavior in *MesoNet*, a 20-parameter network model?

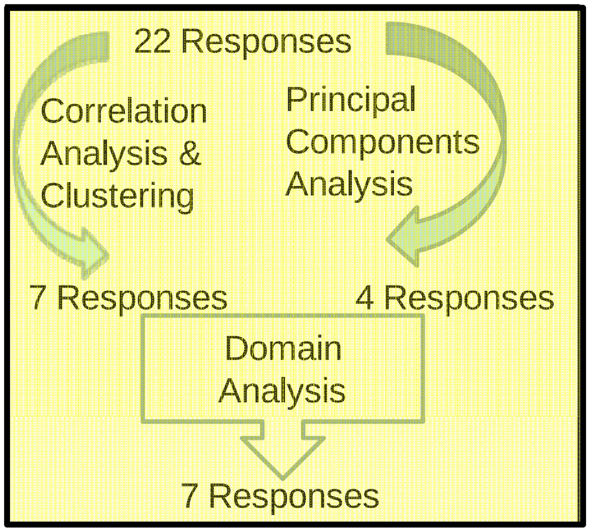
- **Solution** – apply 2-level-per-factor orthogonal fractional factorial (OFF) experimental design and related data analysis techniques to identify the relative importance of model parameters

Scale Reduction: Theory & Practice

Simulating large, fast networks across many conditions and congestion control algorithms requires scale reduction in both model parameters & responses

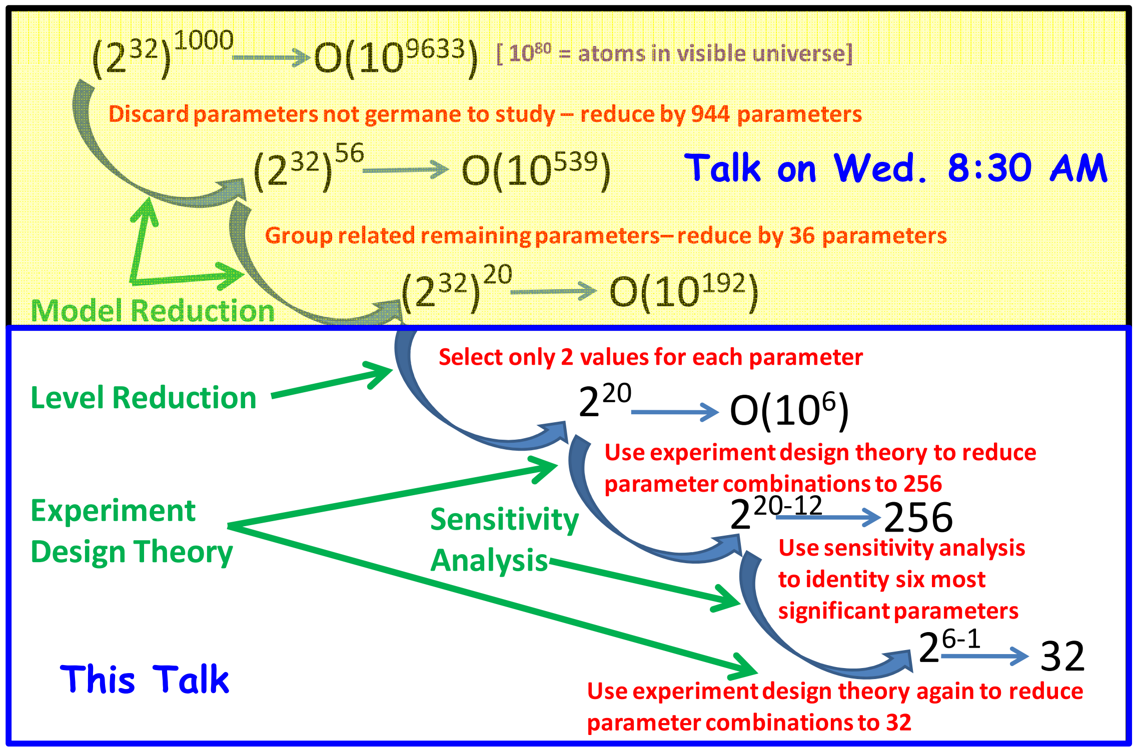
$$\underbrace{y_1, \ldots, y_z}_{\text{Response State-Space}} = f(\underbrace{x_{1|[1,\ldots,\ell]} \ldots, x_{p|[1,\ldots,\ell]}}_{\text{Stimulus State-Space}})$$

Multidimensional Response Reduction



Next Talk

Parameter Reduction



2-Level Per Parameter Orthogonal Fractional Factorial Experimental Design Theory

Sample & Population

$$Y = f(X1, X2, X3, \ldots, Xk)$$

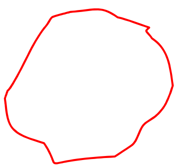
?

Key: What factors are in (my) population?

What scope (robustness) do I want my conclusions to be valid over?

Sample {…}

n =

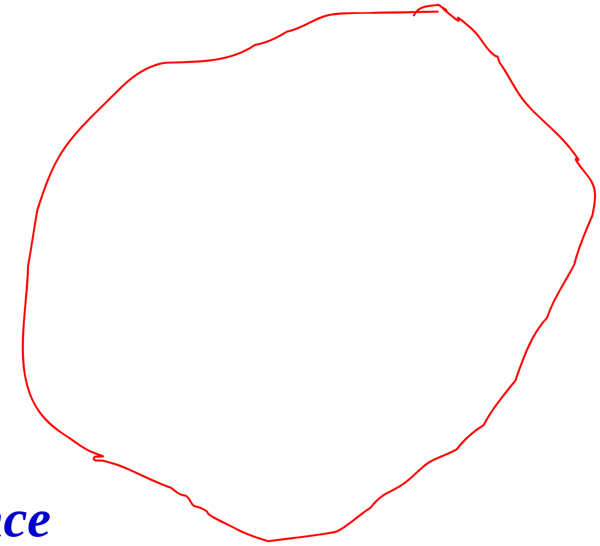


Representative

1. *random sampling*
2. *blocking*
3. *proportional sampling*
4. *orthogonal fractionation*

Population {…}

N =



Summarization

Inference

What is a 2-Level Per Parameter Design?

Each experimental parameter, *p*, is assigned only 2 of its possible values

What is a 2-Level Full Factorial Design?

An experiment is conducted for each of the 2^p parameter combinations

What is a 2-Level Fractional Factorial (FF) Design?

An experiment is conducted for a 2^{p-m} subset of parameter combinations

What is a 2-Level Orthogonal FF (OFF) Design?

The choice of the 2^{p-m} subset of parameter combinations for experiments is made in a fashion that achieves balance and orthogonality, minimizing confounding of interactions between main effects and also between main effects and 2-term interactions and minimizing the variance in the estimation of effects

Why 2 Levels Per Factor?

Pros

- Requires relatively few runs per factor
- Facilitates interpretation of response data
- Identifies promising directions for future experiments, and may be augmented with thorough local exploration
- Forms basis for 2-level fractional factorial designs
- Fits naturally into a sequential strategy, which supports the scientific method

Cons

- Limited exploration of parameter values
- Assumes monotonicity in range between chosen values

Why Orthogonal Fractional Factorial Design?

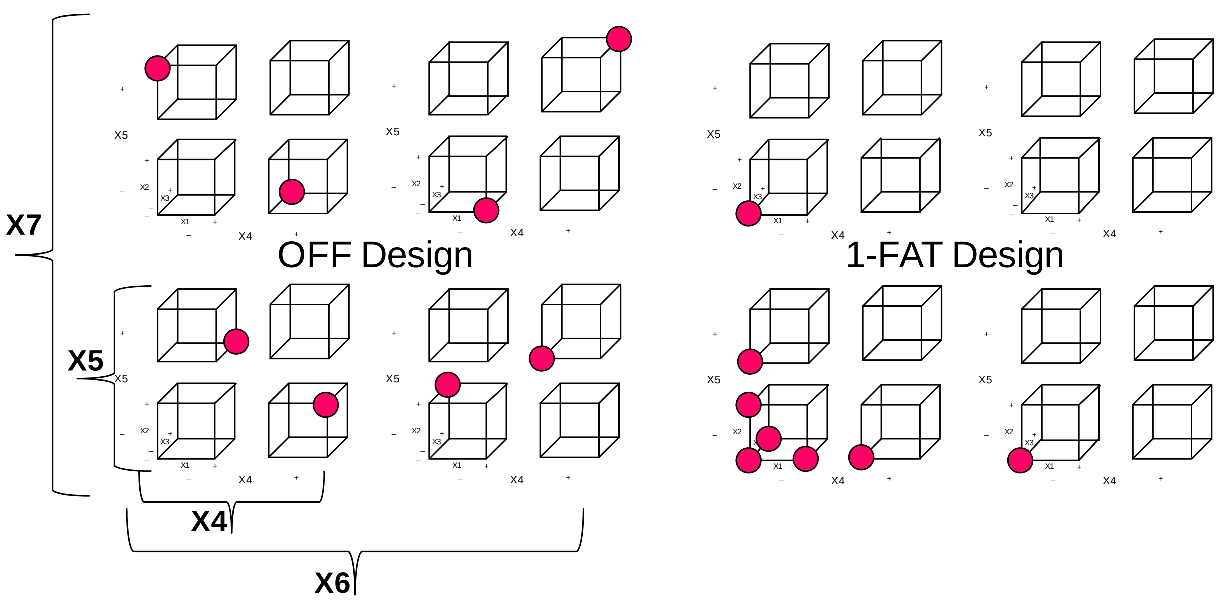
2-Level Design for *MesoNet* requires 2^{20} = 1 048 576 runs

At 28 processor hours per run and with 48 available processors, these runs would require about 612 000 hours (70 years)

Adopting a 2^{20-12} OFF experimental design would reduce the resource requirement to only 256 runs, which could be completed in about 150 hours (1 week)

Cost: misses 2^{20} - 2^{8} parameter combinations

OFF Benefit #1: Superior Coverage & Robustness when compared with1-Factor-at-a-Time Designs (k = 7)



What is the minimum number of required runs?

Minimally strive for a resolution IV design, i.e., a design where there is no confounding among parameters and between parameters and 2-parameter interactions

Requires sufficient runs, n, to resolve a leading constant, the parameters and 2-parameter interactions: $n = 1 + p + C(p, 2)$

MesoNet example – parameters, $p = 20$

Minimum runs $n = 1 + 20 + C(20, 2) = 1 + 20 + 190 = 211$

Given 2-levels per factor, we can choose the first power of 2 above 211

$n = 256 = 2^{20-12}$ – this is a resolution IV design $n = 2^{p-r}$, where the reduction factor is r

Specifying Parameter Combinations

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1
3	-1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	1	-1
4	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	-1
5	-1	-1	1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	1	1	1	1	1	1
6	1	-1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	1
7	-1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	-1	-1	-1	-1	-1	1
8	1	1	1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
9	-1	-1	-1	1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	-1	-1	1
10	1	-1	-1	1	-1	-1	-1	-1	1	-1	1	1	1	1	1	-1	-1	-1	-1	1
11	-1	1	-1	1	-1	-1	-1	-1	1	-1	1	1	1	1	-1	1	1	1	1	1
12	1	1	-1	1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	1	1	1	1	1
13	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	-1	1	1	1	1	-1
14	1	-1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	-1	1	1	1	1	-1
15	-1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	-1	-1	-1	-1	-1
16	1	1	1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1
17	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	-1	1
18	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1	1	1	-1	1	-1	-1	-1	1
19	-1	1	-1	-1	1	-1	-1	-1	1	1	-1	1	1	1	1	-1	1	1	1	1
20	1	1	-1	-1	1	-1	-1	-1	-1	-1	1	-1	-1	-1	1	-1	1	1	1	1
21	-1	-1	1	-1	1	-1	-1	-1	1	-1	1	-1	-1	-1	1	-1	1	1	1	-1
22	1	-1	1	-1	1	-1	-1	-1	-1	1	-1	1	1	1	1	-1	1	1	1	-1
23	-1	1	1	-1	1	-1	-1	-1	-1	1	-1	1	1	1	-1	1	-1	-1	-1	-1
24	1	1	1	-1	1	-1	-1	-1	1	-1	1	-1	-1	-1	-1	1	-1	-1	-1	-1
25	-1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	-1
26	1	-1	-1	1	1	-1	-1	-1	1	-1	-1	1	1	1	1	1	-1	-1	-1	-1
27	-1	1	-1	1	1	-1	-1	-1	1	-1	-1	1	1	1	-1	-1	1	1	1	-1
28	1	1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	-1	1	1	1	-1
29	-1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	-1	-1	1	1	1	1
30	1	-1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	-1	-1	1	1	1	1
31	-1	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	-1	-1	-1	1
32	1	1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	-1	-1	-1	1
⋮																				
256																				

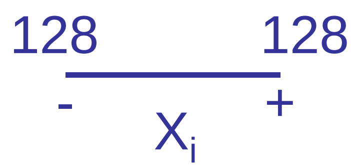
2^{20-12} Experiment Design Template

Design Properties: Balance & Orthogonality

(p = 20, n = 256)

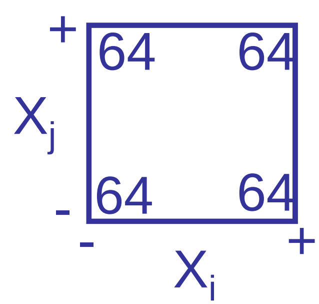
Balance

All 20:

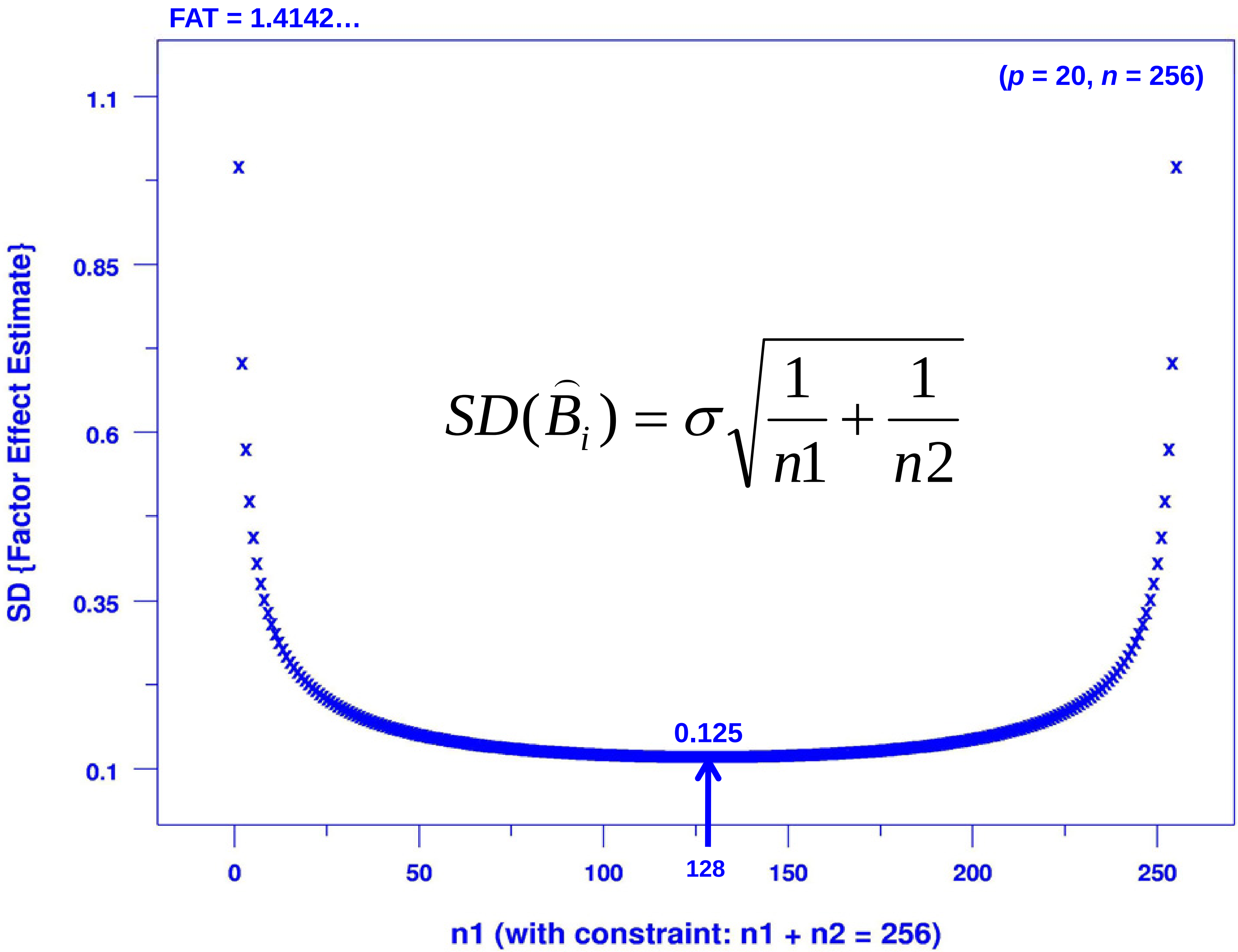


Orthogonality

All $\binom{20}{2}$:



OFF Design Benefit #2: Minimizes Variation in Effect Estimates



$$SD(\widehat{B}_i) = \sigma\sqrt{\frac{1}{n1} + \frac{1}{n2}}$$

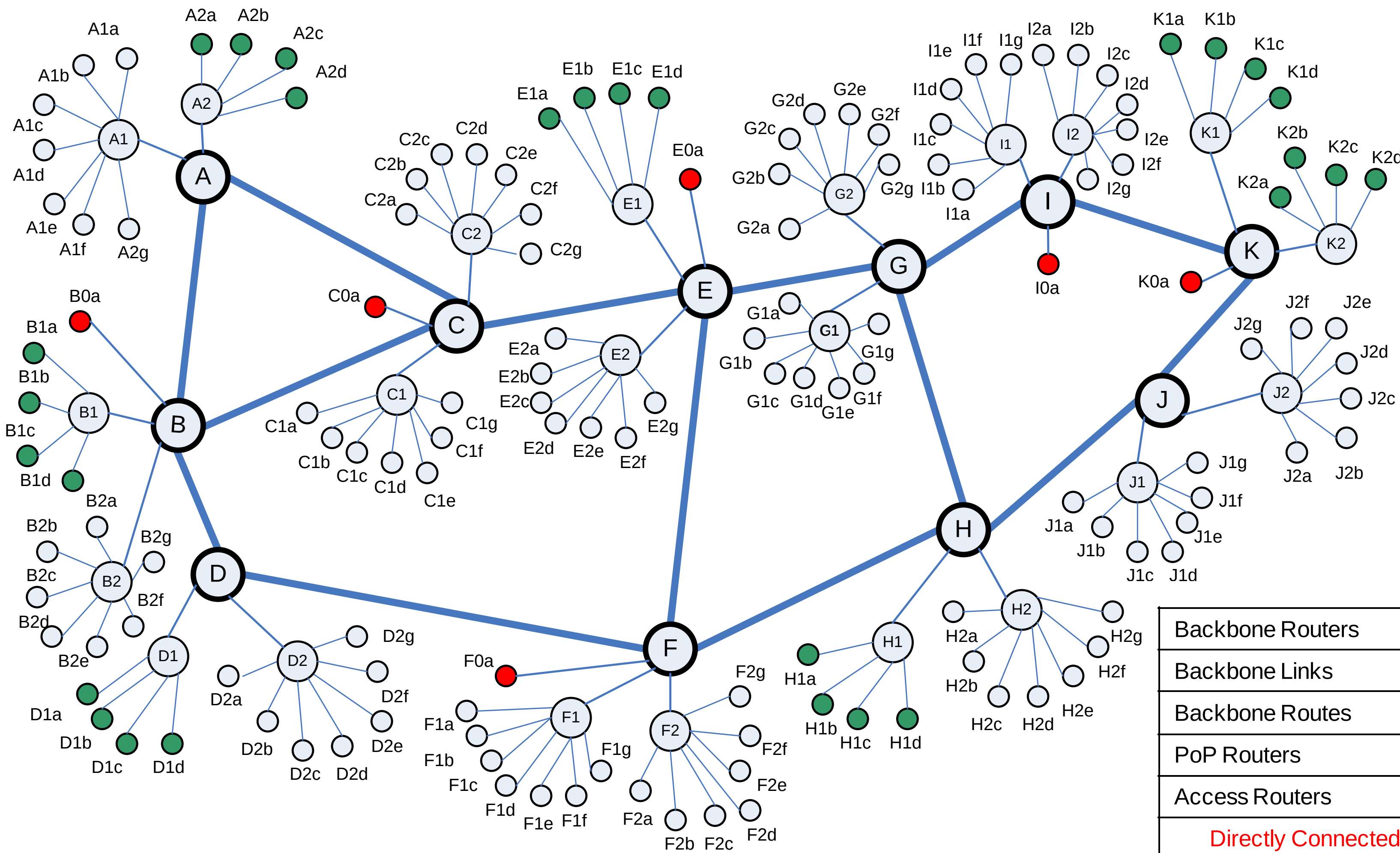
2-Level Per Factor OFF Design Applied to *MesoNet* Sensitivity Analysis

Overview of the 20 Parameters* in *MesoNet*, a TCP\IP Network Simulator

x1	Network Speed	Network Parameters
x2	Propagation Delay	
x3	Buffer Provisioning	
x4	Topology	
x5	Web Browsing File Sizes	User Behavior
x6	Larger File Download Probability & Sizes	
x7	User Think Time	
x8	User Patience	
x9	Spatiotemporal Congestion on Very Fast Paths	
x10	Number, Location and Start Time for Long-Lived Flows	
x11	Speed of Interfaces Connecting Sources & Receivers to Network	Sources & Receivers
x12	Number of Sources & Receivers	
x13	Distribution of Sources	
x14	Distribution of Receivers	
x15	Probability of Source using a specific Congestion Control Algorithm	Protocols
x16	Initial Size of Congestion Window (*cwnd*)	
x17	Initial Slow Start Threshold (*sst*)	
x18	Measurement Interval Size	Simulation & Measurement Control
x19	Simulation Duration	
x20	Startup Pattern for Sources	

*For more details, attend related talk on Wed. 8:30 AM

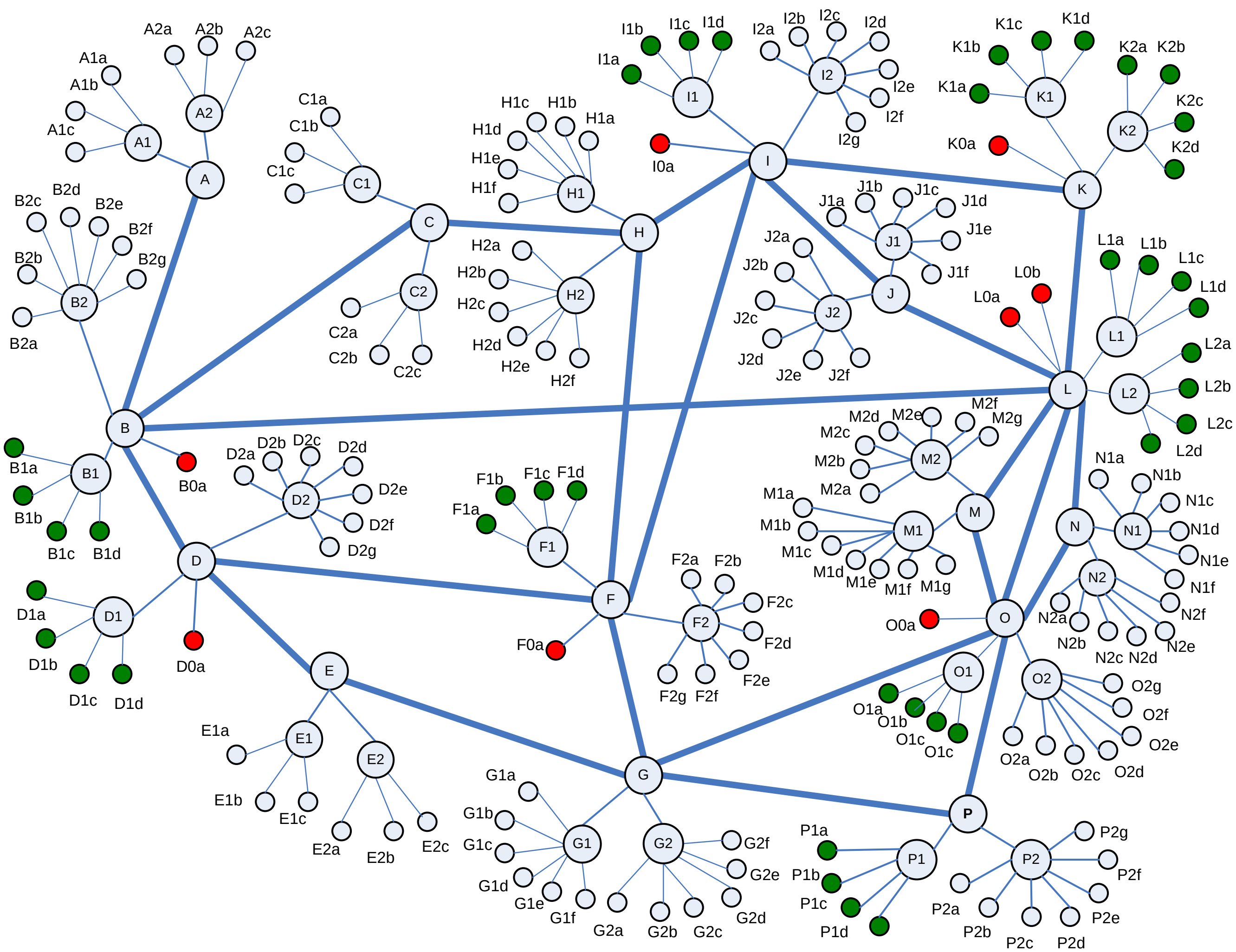
Abilene-based Topology: (-) Level



Backbone Routers	11
Backbone Links	14
Backbone Routes	110
PoP Routers	22
Access Routers	139
Directly Connected	6
Fast	28
Typical	105

Routes are shortest-path based on propagation delay

Commercial ISP-based Topology: (+) Level



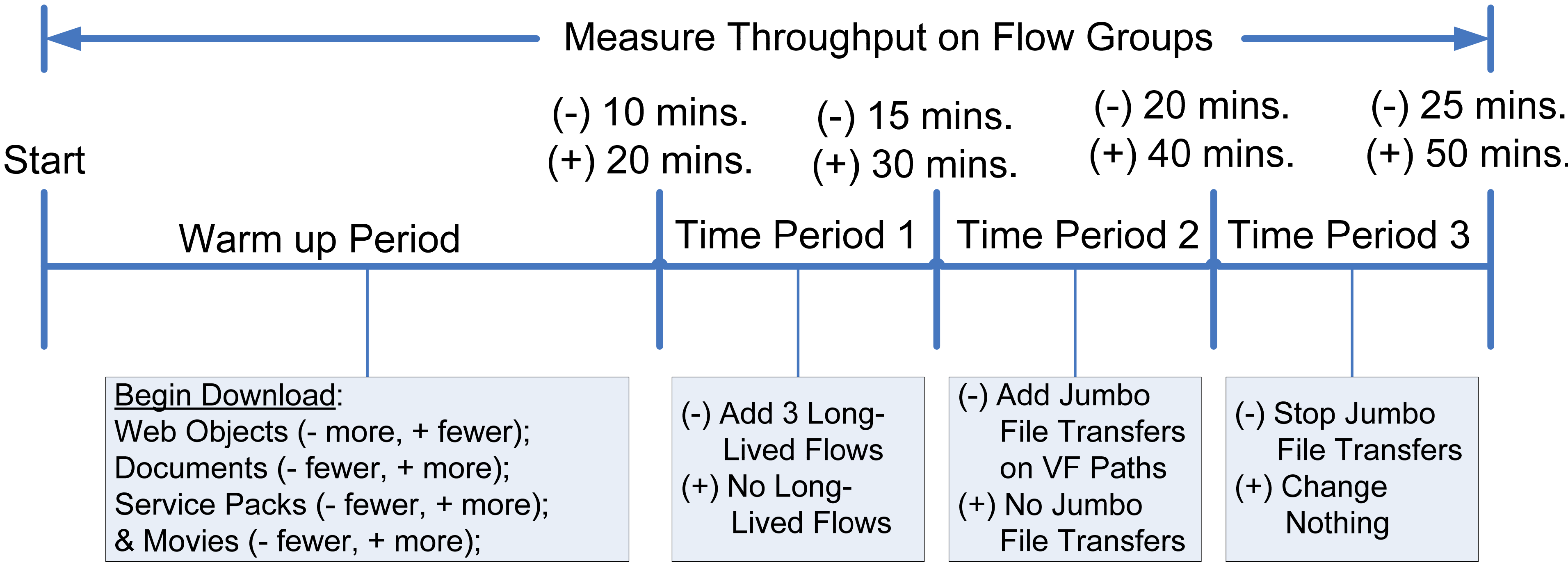
Backbone Routers	16
Backbone Links	24
Backbone Routes	240
PoP Routers	32
Access Routers	170
Directly Connected	8
Fast	40
Typical	122

Routes are shortest-path based on traffic engineering goals

2 Levels Per Factor Used in Sensitivity Analysis

Category	ID	Name	Minus (-1) Level	Plus (+1) Level
Network Configuration	X1	Network Speed	800 packets/ms	1600 packets/ms
	X2	Propagation Delay	1	2
	X3	Buffer Provisioning	*RTT* x *C*/sqrt(*n*)	*RTT* x *C*
	X4	Topology	Abilene - SPF prop. delay	ISP - SPF assigned costs
User Behavior	X5	Web Object Size (α = 1.5)	75 packets	150 packets
	X6	Proportion/Size of Larger Files (*Fx* = 10 *Sx* = 1000 *Mx* = 10000)	*Fp* = 0.02 *Sp* = 0.002 *Mp* = 0.0002	*Fp* = 0.04 *Sp* = 0.004 *Mp* = 0.0004
	X7	User Think Time	2 s	5 s
	X8	User Patience	0.0 (Infinite)	1.0 (Finite)
	X9	Selected Spatiotemporal Congestion	4^{th} Time Period	None
	X10	Long-lived Flows	3 Start in 3^{rd} Time Period	None
Sources & Receivers	X11	Probability of Fast Interface	0.2	0.8
	X12	Number of Sources & Receivers	2	3
	X13	Distribution of Sources	Web Centric	Peer-to-Peer Centric
	X14	Distribution of Receivers	Web Centric	Peer-to-Peer Centric
Protocols	X15	Probability of Algorithms	*TCP* = 0.8 *CTCP* = 0.2	*TCP* = 0.2 *CTCP* = 0.8
	X16	Initial Congestion Window Size	2 packets	8 packets
	X17	Initial Slow Start Threshold	43 packets	1 073 741 823 packets
Simulation & Measurement Control	X18	Measurement Interval Size	200 ms	1 s
	X19	Simulation Duration	25 minutes	50 minutes
	X20	Source Startup Pattern	Exponential (mean = X7)	50 % start early

Traffic Scenario(s)



102 Response Variables

18 Macroscopic Response Variables

Macroscopic Responses		
Category	**Identity**	**Definition**
Flow State	Y1	Average # sources connecting
	Y2	Average # sources sending
	Y3	% sending flows in initial slow start
	Y4	% sending flows in standard congestion avoidance
	Y5	% sending flows in alternate congestion avoidance
Congestion	Y6	Retransmission rate
	Y7	Average congestion window size (packets)
	Y8	Aggregate # connection failures
Delay	Y9	Average round-trip time (ms)
	Y10	Average queuing delay (ms)
Work	Y11	Average # flows completed per second
	Y12	Average # packets output per second
Long-Lived Flows	Y13	Average throughput on long-lived flow #1
	Y14	Average throughput on long-lived flow #2
	Y15	Average throughput on long-lived flow #3
Flows by Path Class	Y16	Average throughput on flows transiting VF paths
	Y17	Average throughput on flows transiting F paths
	Y18	Average throughput on flows transiting T paths

Averaged over each of three time periods (3 x 18 = 54 responses)

Throughput in each of 24 Flow Groups

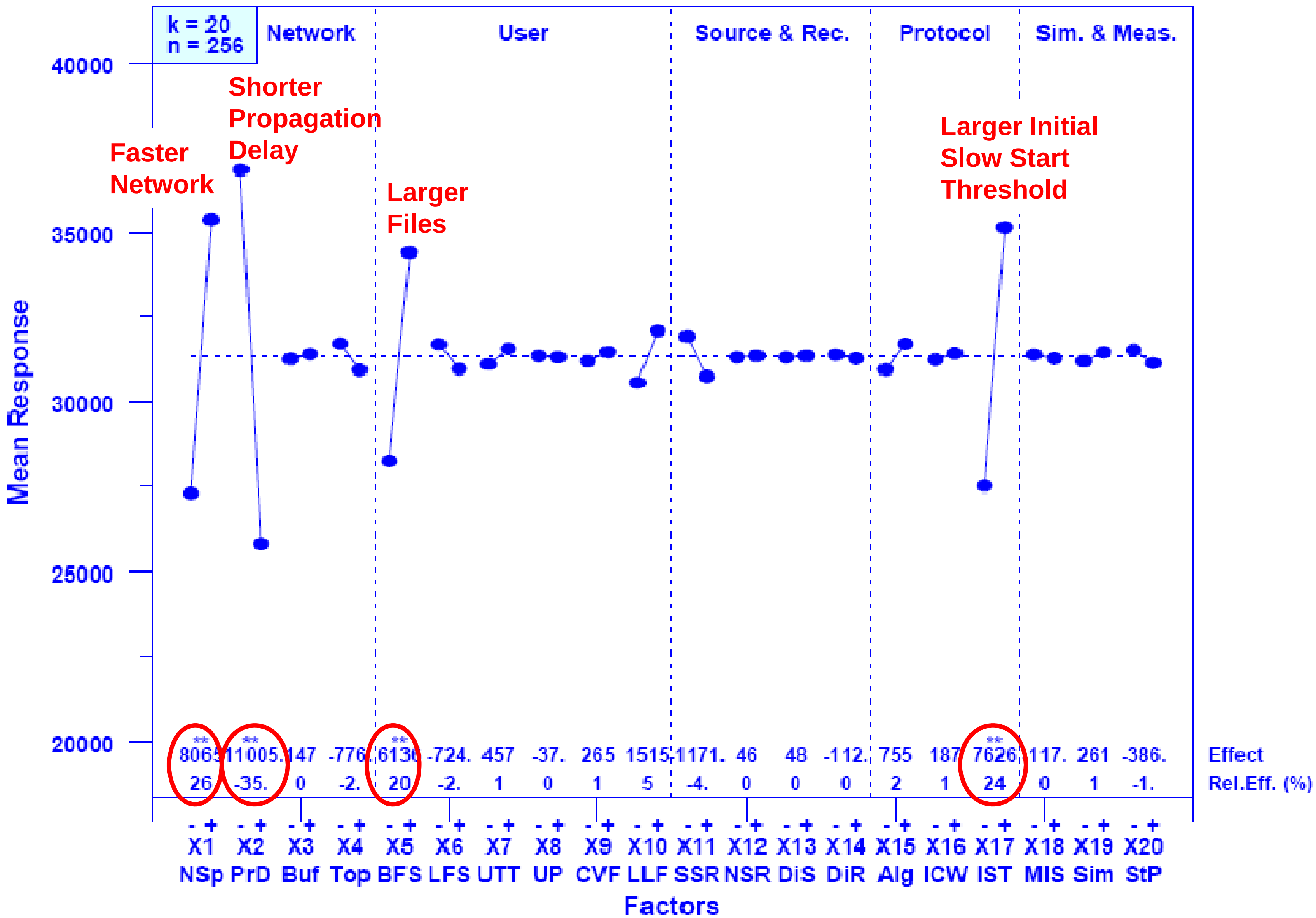
Flow Groups for Throughput Averages			
Number	**File Size**	**Path Class**	**Max. Rate**
1	Movie	VF	F
2		VF	N
3		F	F
4		F	N
5		T	F
6		T	N
7	Software Service Pack	VF	F
8		VF	N
9		F	F
10		F	N
11		T	F
12		T	N
13	Document	VF	F
14		VF	N
15		F	F
16		F	N
17		T	F
18		T	N
19	Web Object	VF	F
20		VF	N
21		F	F
22		F	N
23		T	F
24		T	N

Average Computed Separately for TCP Flows and CTCP Flows (2 x 24 = 48 responses)

Selected Analysis Techniques

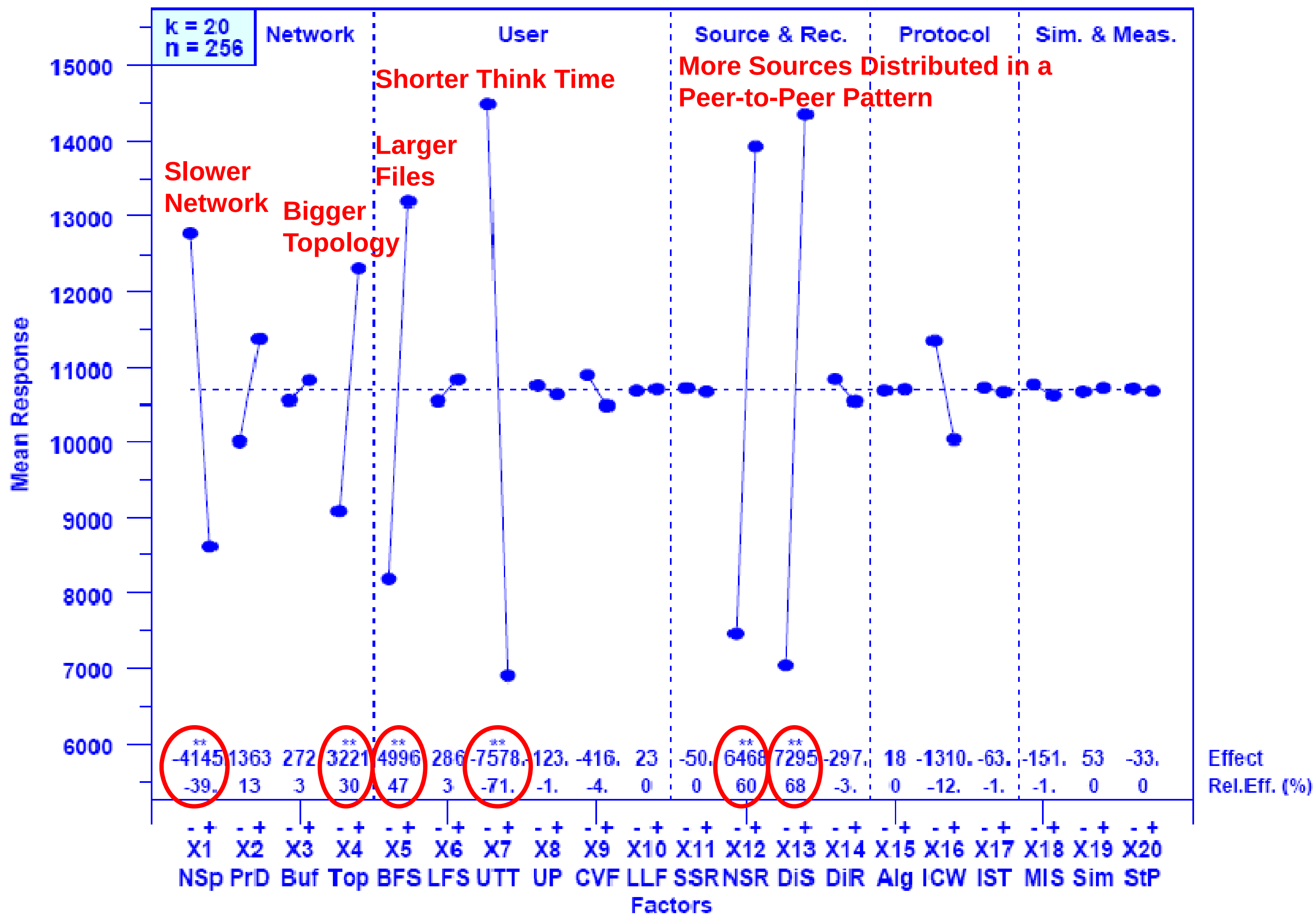
1. Main Effects Analysis
2. Two Factor Interaction Analysis
3. Tabular Summary Analysis

Sample: 1 of 102 Main Effects Analyses



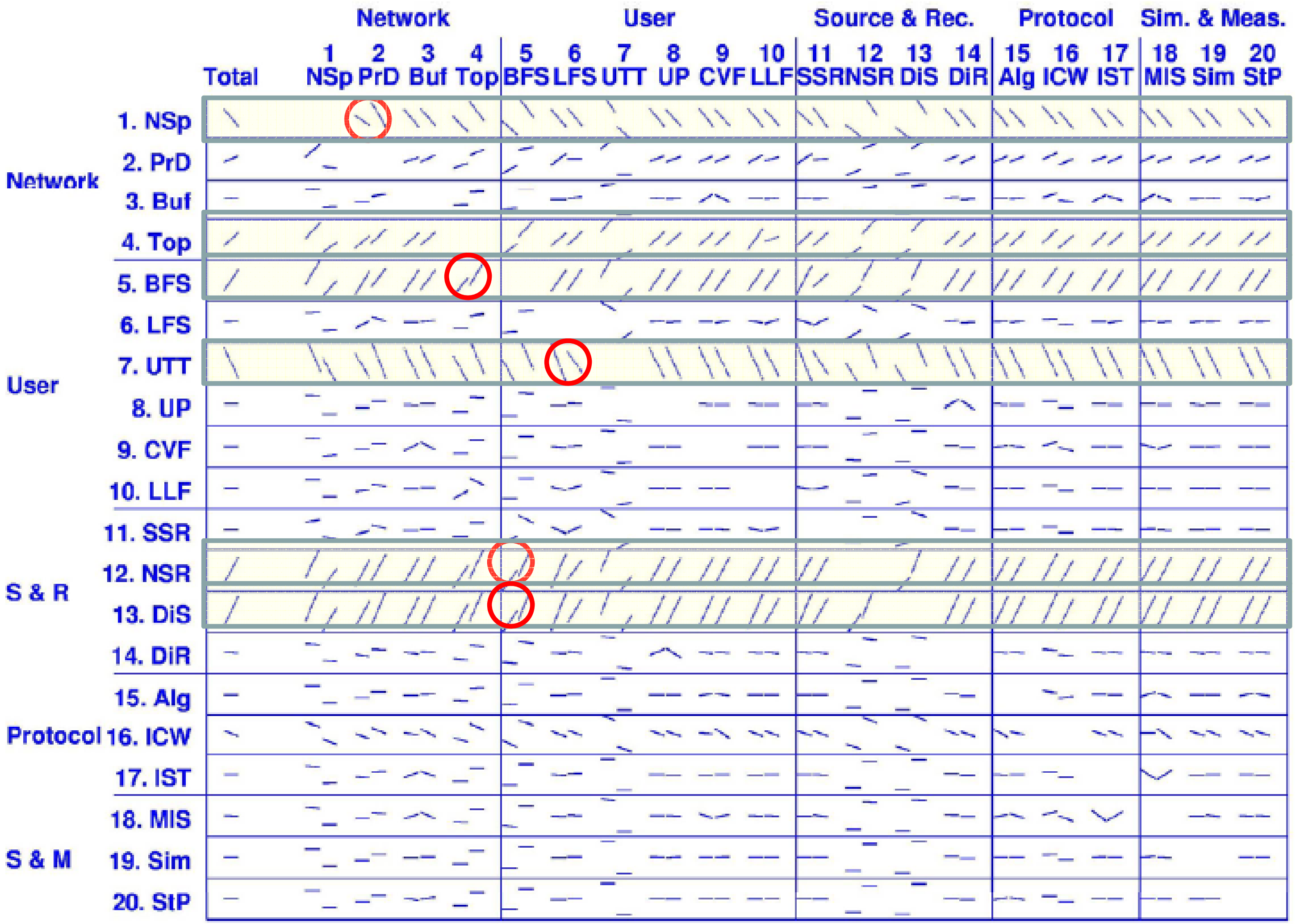
Throughput (pps) for Movies transferred over Very Fast Paths with Fast Interfaces using CTCP

Another Sample: 1 of 102 Main Effects Analyses



Y2 – Average Number of Sending Sources in Time Period #2

Sample: 1 of 102 Two Factor Interaction Analyses



Two Factor Interaction Plot for Y2 – Avg. Number of Sending Sources in Time Period #2
(not much in the way of significant 2 factor interactions)

Sample: 1 of 5 Tabular Summary Analyses

Metric Class	Y#	Name	Network				User Behavior						Source/Receiver				Protocol			Sim. Control & Meas.		
			X1 NSp	X2 PrD	X3 Buf	X4 Top	X5 FS	X6 LFS	X7 ThT	X8 UP	X9 CVF	X10 LLF	X11 SSR	X12 NSR	X13 DiS	X14 DiR	X15 CCA	X16 ICW	X17 IST	X18 MIS	X19 DUR	X20 StP
Flows	Y1	# Connecting	+**		+**	+*			-**					+**	+**			+**				
	Y2	# Active	+**			+**	+**		-**					+**	+**							
	Y3	% ISS	+**	+**	+**		-**		+**					-**	-**				+**			
	Y4	% NCA	-**	-**	-**		+**		-**					+**	+**		-*					
	Y5	% ACA	+**		+**				+**					-**	-**		+**	+*	-**			
Congestion	Y6	Retrans. Rate	-**	-**	-**		+**		-**					+**	+**			+**				
	Y7	*cwnd* Size	+*																			
	Y8	# conn. fails	-**	-**	-**		+**							+**	+**			+**				
Delay	Y9	SRTT	-**	+**	+**									+*	+*							
	Y10	Queue Delay	-**	+**	+**				-*					+**	+**							
Aggregate TP	Y11	Flows/sec	+**	-*		+**	-**		-**					+**	+**							
	Y12	Packets/sec	+**		+**	+**	+**		-**		-**			+**								
Long-Lived Flow TP	Y13	LLF 1	+**		+*	+*					+**	-**										
	Y14	LLF 2	+*			+**					+**	-**										
	Y15	LLF 3	+**								+**	-*				-*						
Other Flow TP	Y16	VF Paths	+**	-**		-**			+*		+**			-**		-*			+**	-*		
	Y17	F Paths	+**	-**		+*			+**		+**			-**	+**	-**		+**		-*		
	Y18	N Paths	+**	-**		-**			+**					-**	-**			+**				

Significant Influence of each Factor on each Macroscopic Response in Time Period #2

Another Sample: 1 of 5 Tabular Summary Analyses

			Network				User Behavior						Source/Receiver				Protocol			Sim. Control & Meas.		
File Type	Path Class	Connection Speed	X1 NSp	X2 PrD	X3 Buf	X4 Top	X5 FS	X6 LFS	X7 ThT	X8 UP	X9 CVF	X10 LLF	X11 SSR	X12 NSR	X13 DiS	X14 DiR	X15 CCA	X16 ICW	X17 IST	X18 MIS	X19 DUR	X20 StP
Movies	VF	Fast	+**	-**			+**												+**			
	VF	Normal	+**	-**							+**			-*					+**			
	F	Fast	+**		+**	-**	-**		+**					-*	+**	-*						
	F	Normal	+**		+**	+**	-**		+**					-**	+**	-*						
	T	Fast	+**		+**		-**		+**					-**	-**							
	T	Normal	+**		+**		-**		+**					-**	-**							
Service Packs	VF	Fast	+**	-**			+**												+**			
	VF	Normal		-**		-*	+**												+**			
	F	Fast	+**	-**	+**	+**			+*						+**	-*			+**			
	F	Normal	+**	-**	+**	+**			+**					-**	+**	-*			+**			
	T	Fast	+**		+**		-**		+**					-*	-**							
	T	Normal	+**		+**		-**		+**					-*	-**							
Documents	VF	Fast		-**		+*												+*	+**			
	VF	Normal		-**			+**											+**	+**			
	F	Fast	+**	-**		+**	+*		+*					-*	+**				+**			
	F	Normal	+**	-**		+**	+**		+**					-*	+**				+**			
	T	Fast	+**	-**	+**				+**					-**	-**				+**			
	T	Normal	+**	-**	+**				+**					-**	-**				+**			
Web Objects	VF	Fast		-**		+**	+**											+**				
	VF	Normal		-**		+*	+**											+**				
	F	Fast	+**	-**	-*	+**	+**		+*						+**			+**				
	F	Normal	+**	-**	-*	+**	+**		+**						+**			+**				
	T	Fast	+**	-**					+**					-**	-**			+**				
	T	Normal	+**	-**					+**					-**	-**			+**				

Significant Influence of Each of 20 Factors on Throughput for Each of 24 Flow Groups when using CTCP

Relative Importance of *MesoNet* Parameters

Summary of Influence of Each Factor on All Responses

Protocol	T-test Statistic	Network				User Behavior						Source/Receiver				Protocol			Sim. Control & Meas.		
		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
		NSp	PrD	Buf	Top	FS	LFS	ThT	UP	CVF	LLF	SSR	NSR	DiS	DiR	CCA	ICW	IST	MIS	DUR	StP
Time Period #1	>0.99	17	9	10	8	8	0	11	0	0	3	0	12	11	2	1	7	2	1	0	0
	>0.95≤0.99	1	1	3	2	2	0	2	0	0	0	0	3	2	1	0	2	1	1	0	0
	Total	18	10	13	10	10	0	13	0	0	3	0	15	13	3	1	9	3	2	0	0
Time Period #2	>0.99	16	9	9	6	7	0	10	0	6	2	0	13	11	1	1	5	3	0	0	0
	>0.95≤0.99	2	1	1	2	0	0	2	0	0	1	0	1	1	2	1	1	0	2	0	0
	Total	18	10	10	8	7	0	12	0	6	3	0	14	12	3	2	6	3	2	0	0
Time Period #3	>0.99	17	9	11	6	9	0	12	0	4	3	0	12	11	2	1	5	3	1	0	0
	>0.95≤0.99	1	2	0	3	1	0	1	0	1	0	0	3	2	0	0	0	0	2	0	0
	Total	18	11	11	9	10	0	13	0	5	3	0	15	13	2	1	5	3	3	0	0
TCP	>0.99	19	16	12	8	11	0	10	0	1	0	0	4	16	0	0	8	16	1	0	0
	>0.95≤0.99	0	2	3	3	5	0	4	0	0	0	0	2	0	0	0	0	0	0	0	0
	Total	19	18	15	11	16	0	14	0	1	0	0	6	16	0	0	8	16	1	0	0
CTCP	>0.99	19	18	10	9	15	0	13	0	1	0	0	8	16	0	0	7	12	0	0	0
	>0.95≤0.99	0	0	2	3	1	0	3	0	0	0	0	6	0	4	0	1	0	0	0	0
	Total	19	18	12	12	16	0	16	0	1	0	0	14	16	4	0	8	12	0	0	0
Total	>0.99	88	61	52	37	50	0	56	0	12	8	0	49	65	5	3	32	36	3	0	0
	>0.95≤0.99	4	6	9	13	9	0	12	0	1	1	0	15	5	7	1	4	1	5	0	0
	Total	92	67	61	50	59	0	68	0	13	9	0	64	70	12	4	36	37	8	0	0
% of responses influenced		90%	66%	60%	49%	58%		67%		13%	9%		63%	69%	12%	4%	35%	36%	8%		

Significant Influence of Each of 20 Factors on Each of 18 Macroscopic Responses

Findings: 7 main factors drive *MesoNet* Response

- Capacity (network speed)
- Demand (number, distribution and activity of sources)
- Physics (propagation delay)
- Buffer sizing
- File sizes

Conclusions

- 2-Level-per-Factor Orthogonal Fractional Factorial (OFF) experimental designs can reveal significant information about simulation models
- *MesoNet* simulation appears to be driven by the same key factors that influence behavior in real networks
- Appears feasible to compare proposed Internet congestion control algorithms while varying only 7 *MesoNet* parameters

Entire Study Report NIST Special Publication 500-282:
<u>Study of Proposed Internet Congestion Control Mechanisms</u>
Available online at http://www.nist.gov/itl/antd/Congestion_Control_Study.cfm