An Efficient Sensitivity Analysis Method for Mesoscopic Network Models

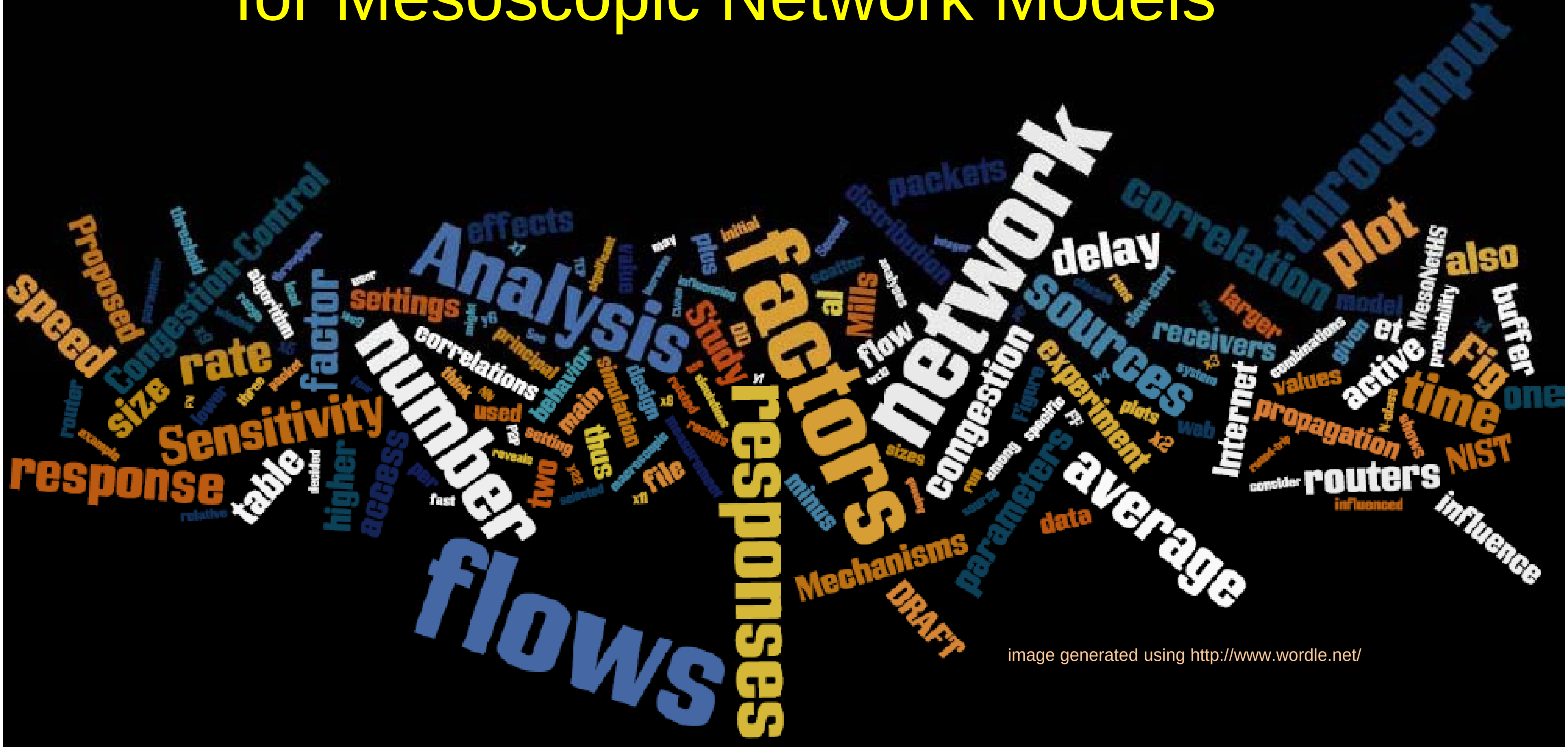


image generated using http://www.wordle.net/

Kevin Mills & Jim Filliben, NIST

CxS Study Group – Feb 2, 2010

Outline

- Goal – Problem – Solution
- Scale Reduction: Theory and Practice
- Overview of the 20 *MesoNet* Parameters
- 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
- Conclusions
- Future Work

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

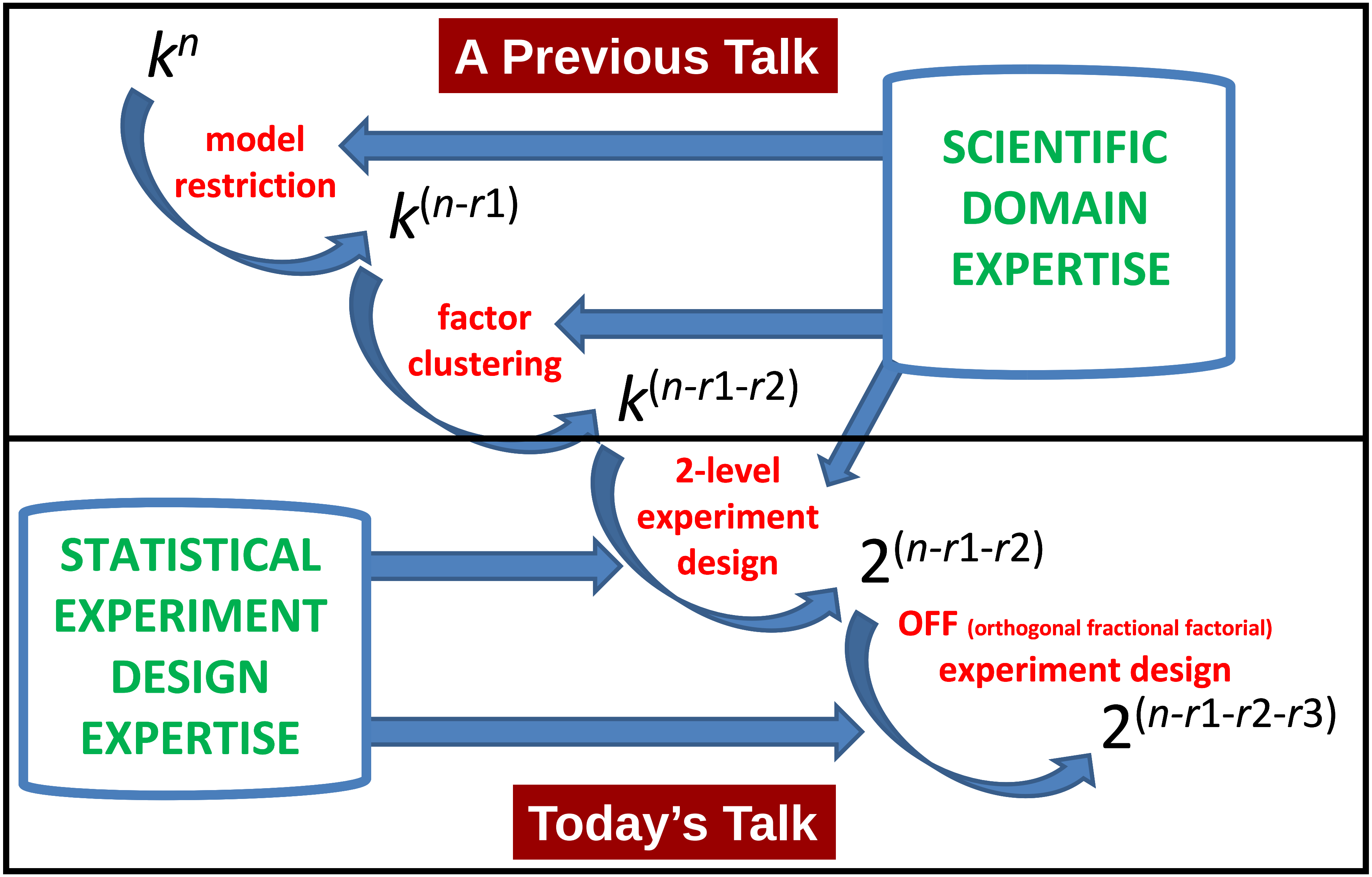
The Function ☺ of a Simulation Model

$$y_1, \ldots, y_m = f(\, x_{1|[1,\ldots,k]}, \ldots, x_{n|[1,\ldots,k]} \,)$$

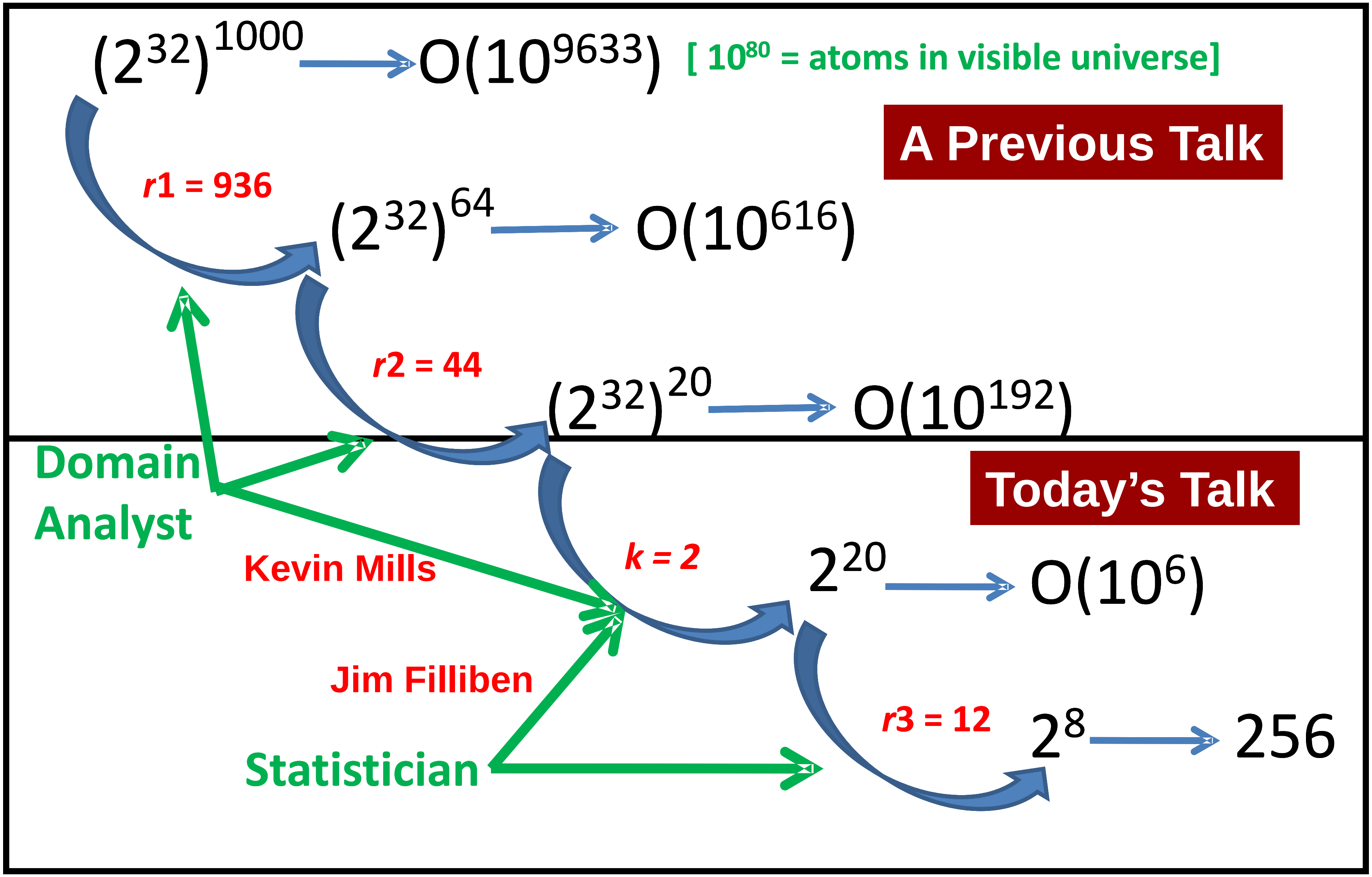
Model Response Space[+] **Model Parameter Space**

[+] Determining which responses to examine is an interesting problem in its own right. Though not addressed in this presentation, we used correlation and principal components analyses to reduce the response space.

Theory – Scale Reduction in Two Parts



Practice – Scale Reduction in Two Parts



Brief Review of *MesoNet* Parameters

MesoNet – a TCP/IP network model using only 20 parameters

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**	

2-Level Per Parameter Orthogonal Fractional Factorial Experimental Design Theory

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 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 linear behavior in range between chosen values

Why Orthogonal Fractional Factorial Design?

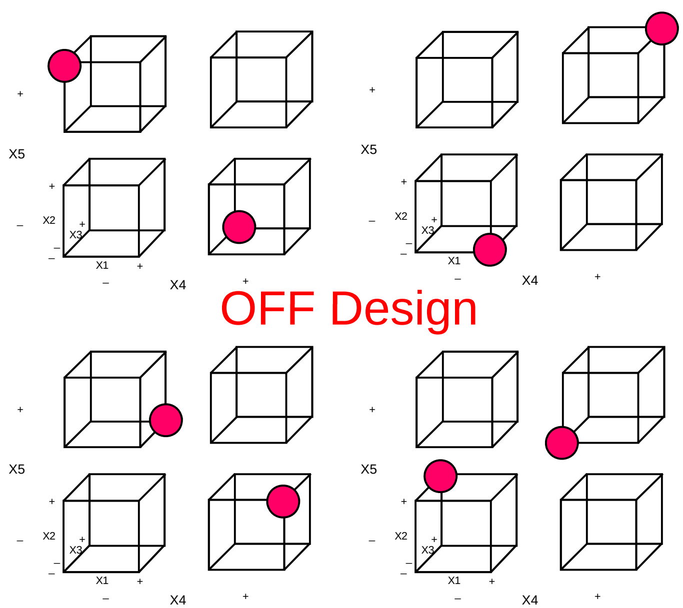
2-Level Design for *MesoNet* requires 2^{20} = 1 048 576 runs

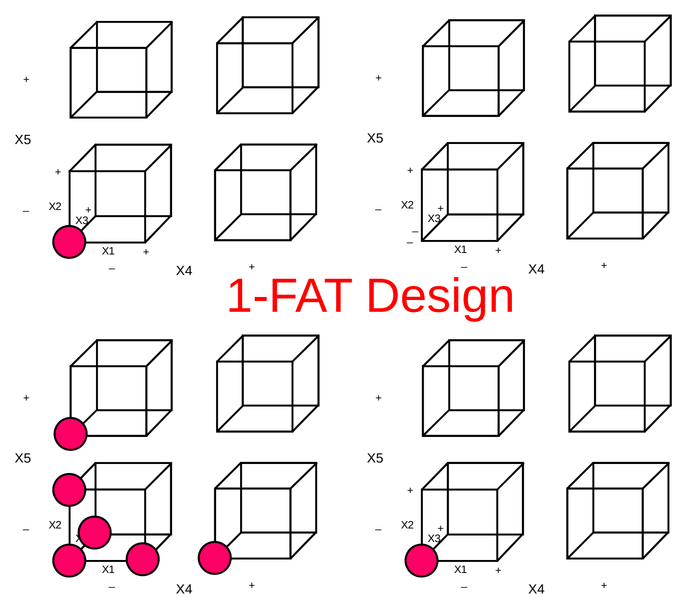
At 28 processor hours per run and with 48 available processors, these runs would require about 612 000 hours (70 years)

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^{12} parameter combinations

OFF Benefit #1: Superior Coverage & Robustness when compared with1-Factor-at-a-Time Designs





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 and where any confounding among specific pairs of 2-parameter interactions is known

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

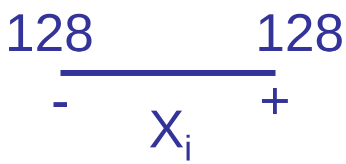
•
•
•
256

2^{20-12} Experiment Design

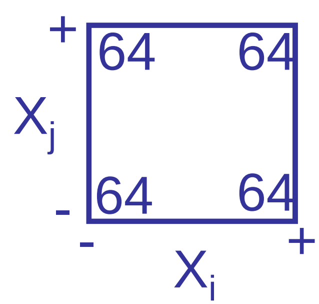
Design Properties: Balance & Orthogonality

(p = 20, n = 256)

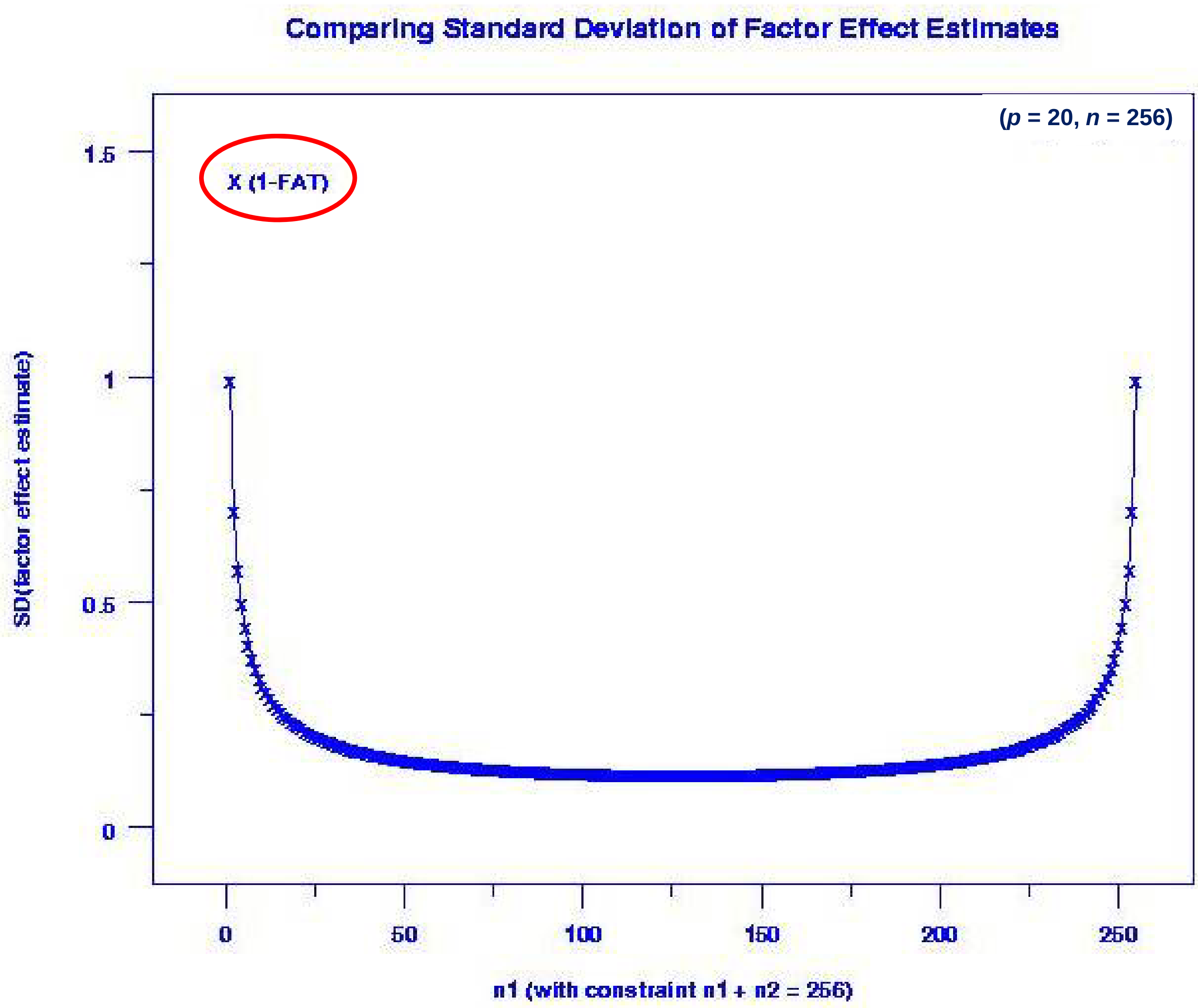
Balance All 20:



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



OFF Design Benefit #2: Minimizes Variation in Effect Estimates

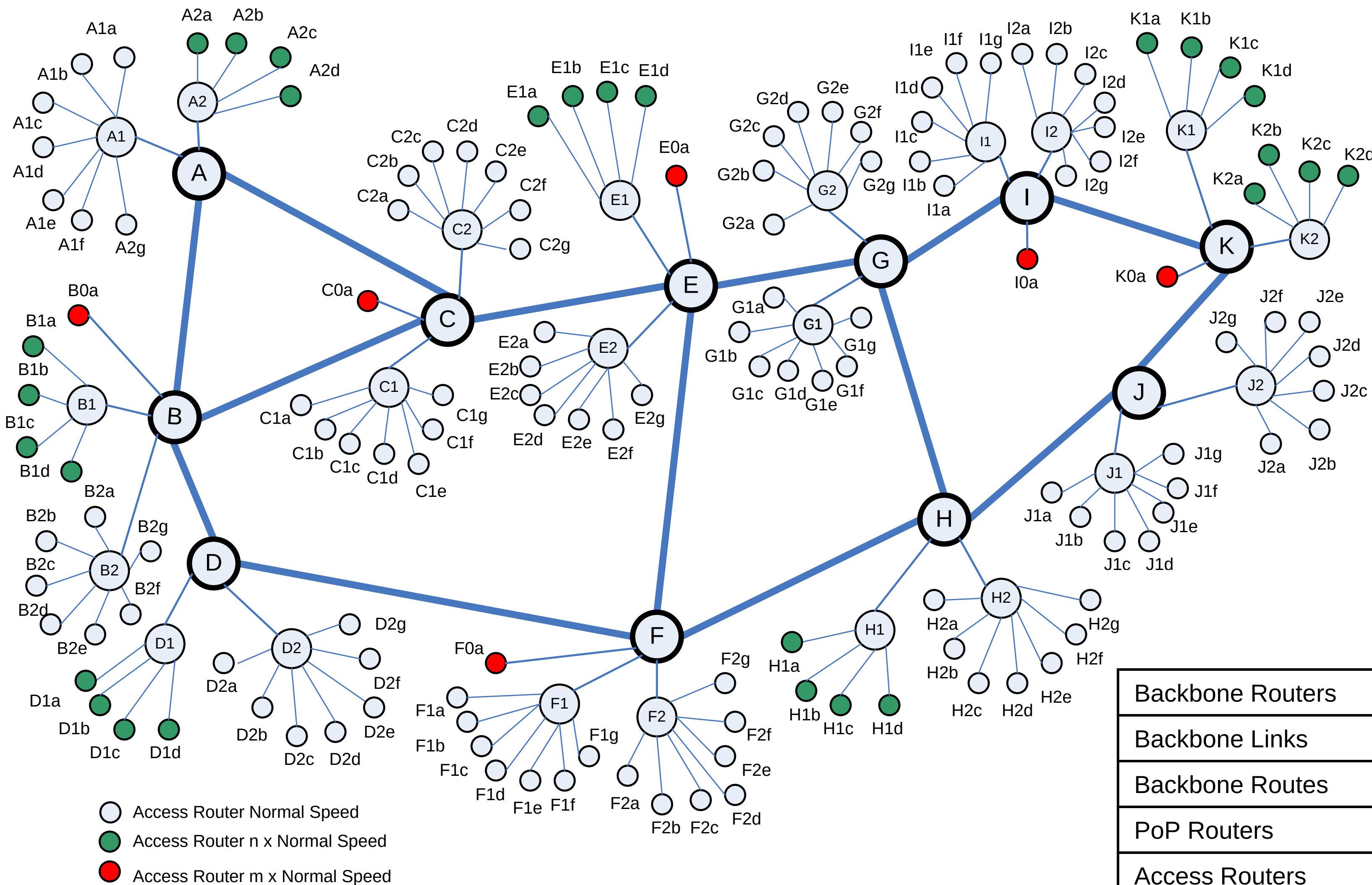


2-Level Per Factor OFF Design Applied to *MesoNet* Sensitivity Analysis

2 Levels Per Factor Used in Sensitivity Analysis

Factor	Parameter Definition	MINUS (-) LEVEL	PLUS (+) LEVEL
x1	Network Speed	*BBspeedup* = 2 *R1* = 800 packets/ms	*BBspeedup* = 2 *R1* = 1600 packets/ms
x2	Propagation Delay	ΔX = 1	ΔX = 2
x3	Buffer Provisioning	*RTT* x *C* /sqrt(*n*)	*RTT* x *C*
x4	Topology	Abilene - SPF propagation delay	ISP - SPF traffic engineering goals
x5	Web Browsing File Sizes	λon = 75 α = 1.5	λon = 150 α = 1.5
x6	Larger File Download Probability & Sizes	*Fx* = 10 *Sx* = 1000 *Mx* = 10000	*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 seconds	5 seconds
x8	User Patience	NONE REACTIVE *RFp* = 0.0	ALL REACTIVE *RFp* = 1.0
x9	Spatiotemporal Congestion on Very Fast Paths	4th Time Period	NONE
		Jon = 0.6 *Joff* = 0.8 *Jx* = 100	*Jon* = 1.0 *Joff* = 1.0 *Jx* = 100
x10	Number, Location and Start Time for Long-Lived Flows	3 Start 3rd Time Period with	NONE
		distances: short, medium, long	
x11	Speed of Interfaces Connecting Sources & Receivers to Network	FastHostProb = 0.2	FastHostProb = 0.8
x12	Number of Sources & Receivers	ΔU = 2	ΔU = 3
x13	Distribution of Sources	WEB *pNs* = 0.1 *pNsf* = 0.6 *pNsd* = 0.3	P2P *pNs* = 0.34 *pNsf* = 0.33 *pNsd* = 0.33
x14	Distribution of Receivers	WEB *pNr* = 0.6 *PNrf* = 0.2 *pNfd* = 0.2	P2P *pNr* = 0.34 *PNrf* = 0.33 *pNfd* = 0.33
x15	Probability of Source using a specific Congestion Control Algorithm	*prTCP* = 0.8 *prCTCP* = 0.2	*prTCP* = 0.2 *prCTCP* = 0.8
x16	Initial Size of Congestion Window (*cwnd*)	2 packets	8 packets
x17	Initial Slow Start Threshold (*sst*)	43 packets	1 073 741 823 packets
x18	Measurement Interval Size	200 ms	1 second
x19	Simulation Duration	25 minutes	50 minutes
x20	Startup Pattern for Sources	*prOn1st* = 0.0 *prOn2nd* = 0.0	*prOn1st* = 0.25 *prOn2nd* = 0.08
		prOn3rd = 0.0 *prRest* = 1.0	*prOn3rd* = 0.17 *prRest* = 0.50

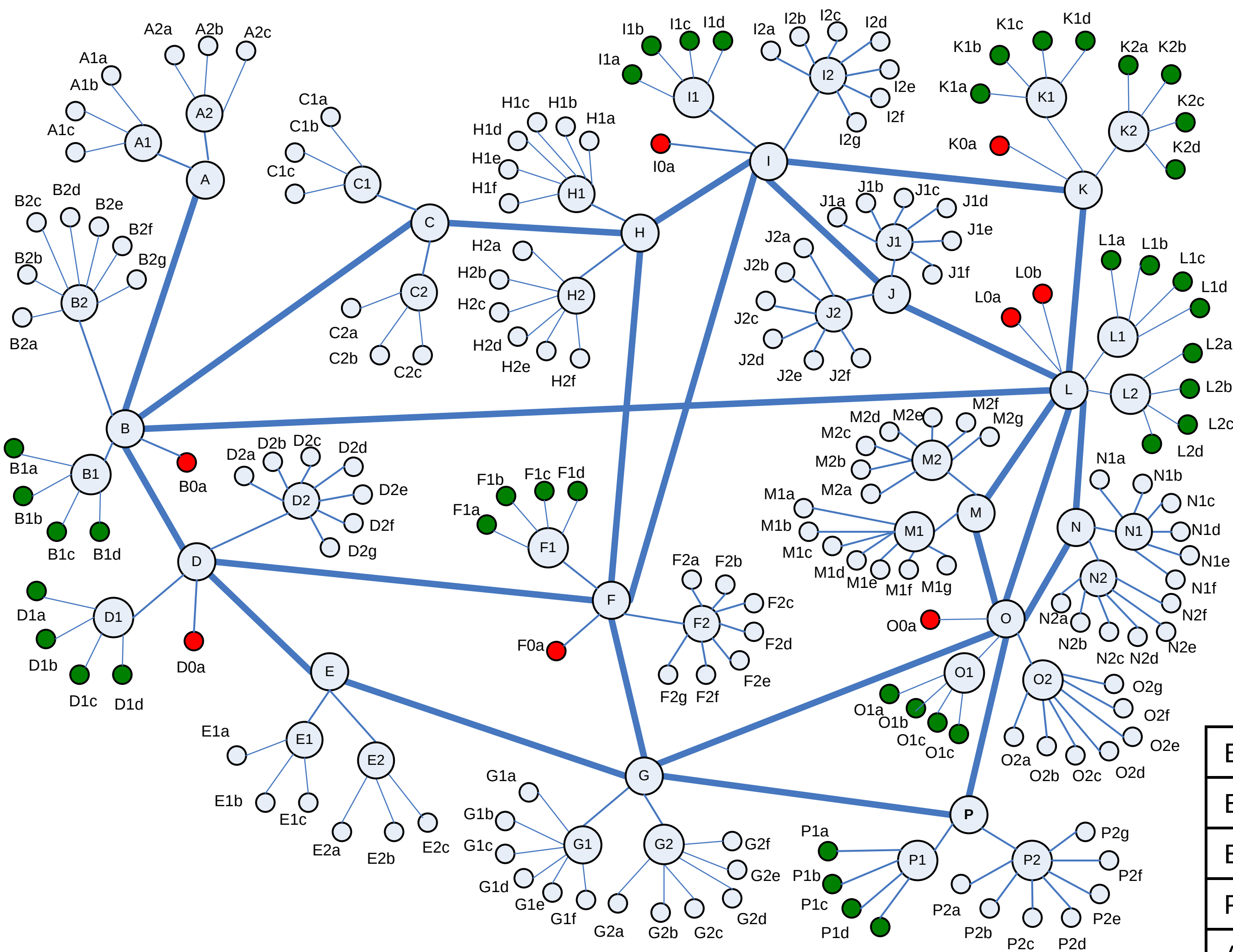
Abilene-based Topology: (-) Level



Routes are shortest-path based on propagation delay

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

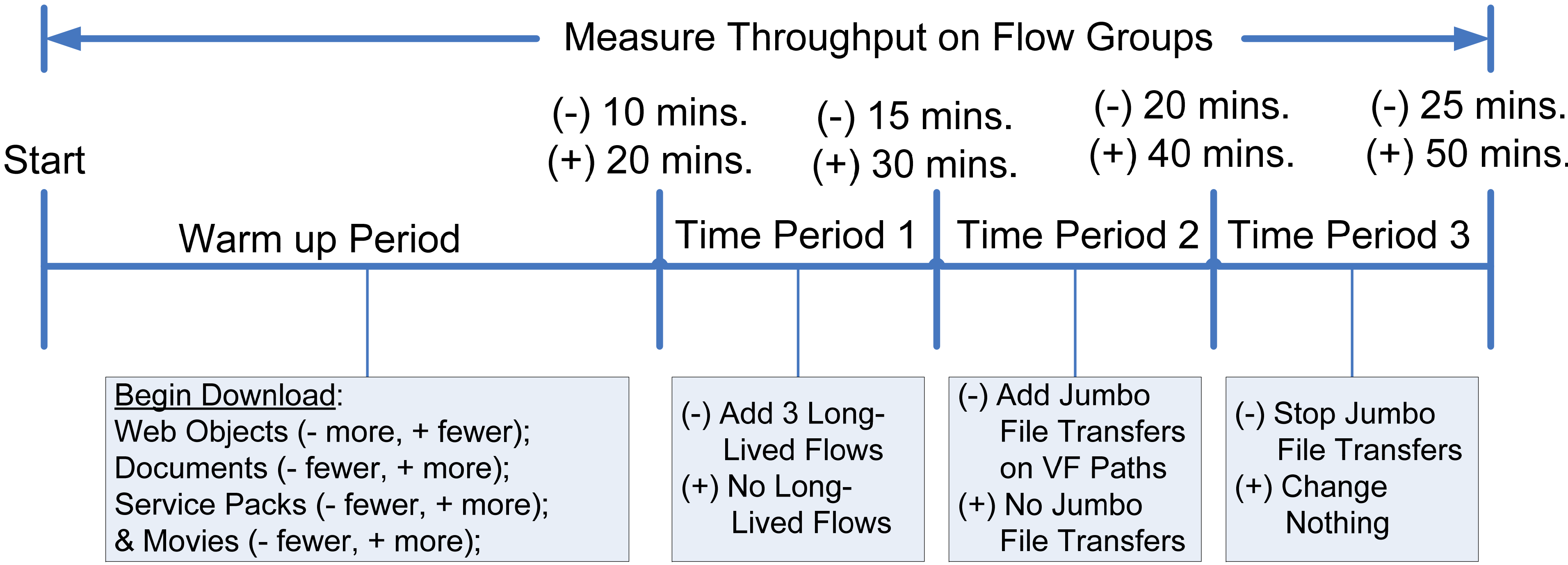
Commercial ISP-based Topology: (+) Level



Routes are shortest-path based on traffic engineering goals

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

Traffic Scenario(s)



18 Macroscopic Response Variables

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

Network-wide Flow State	y1	Average number of sources connecting
	y2	Average number of sources sending
	y3	Proportion of sending flows in initial slow-start
	y4	Proportion of sending flows in normal congestion avoidance
	y5	Proportion of sending flows in alternate congestion avoidance
Network-wide Congestion	y6	Retransmission Rate
	y7	Average Congestion Window size
	y8	Aggregate Connection Failures
Network Delay	y9	Average Round-Trip Time
	y10	Average Queuing Delay
Network Throughput	y11	Average number of flows completed per second
	y12	Average number of flows output per second
Throughput on 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
Throughput for Flows on each Path Class	y16	Average throughput for flows transiting Very Fast (VF) Paths
	y17	Average throughput for flows transiting Fast (F) Paths
	y18	Average throughput for flows transiting Typical (T) Paths

Average Throughput in each of 24 Flow Groups

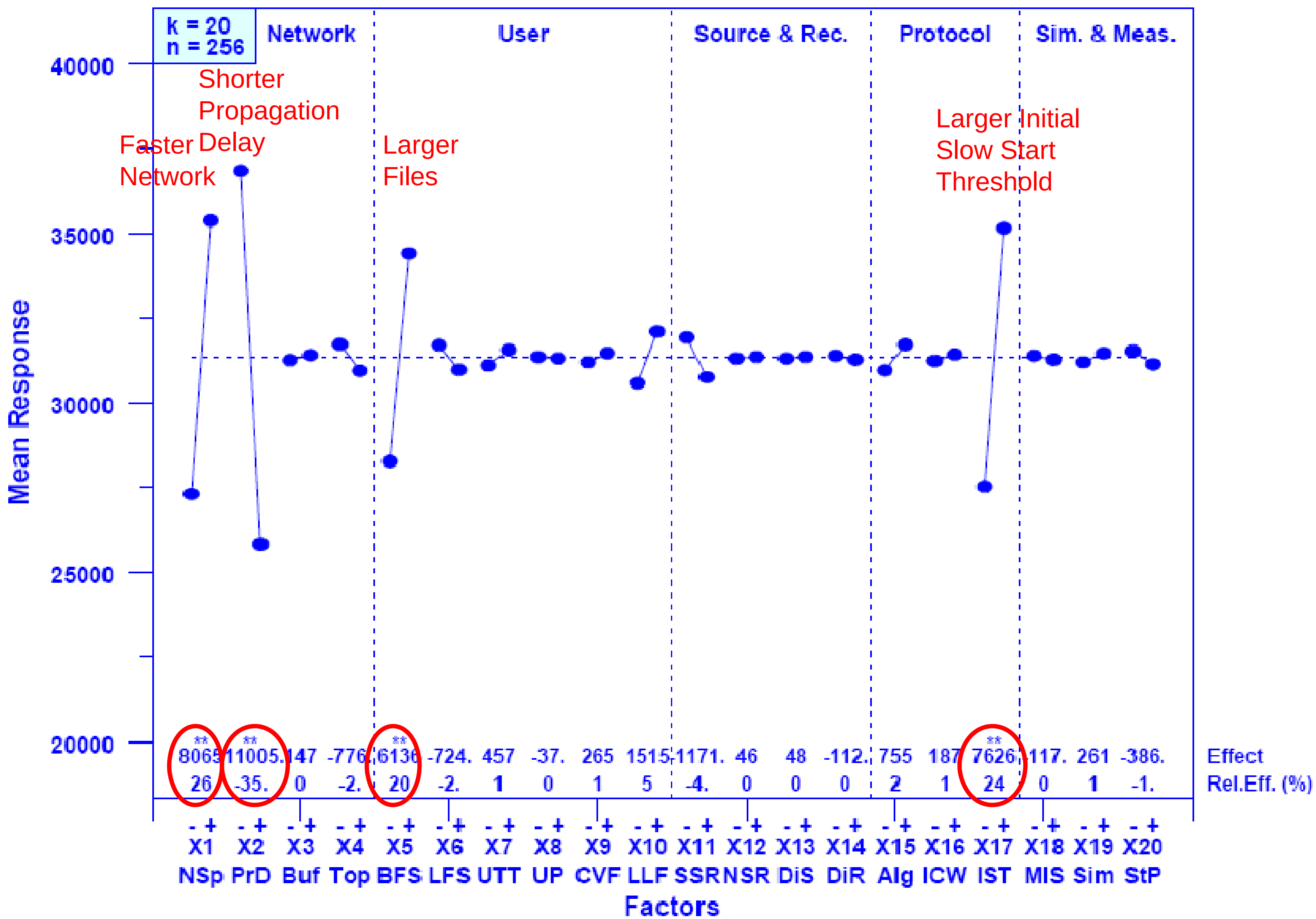
Average Computed Separately for TCP Flows and CTCP Flows (2 x 24 = 48 responses)

File Type	Path Class	Connection Speed
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

Selected Analysis Techniques

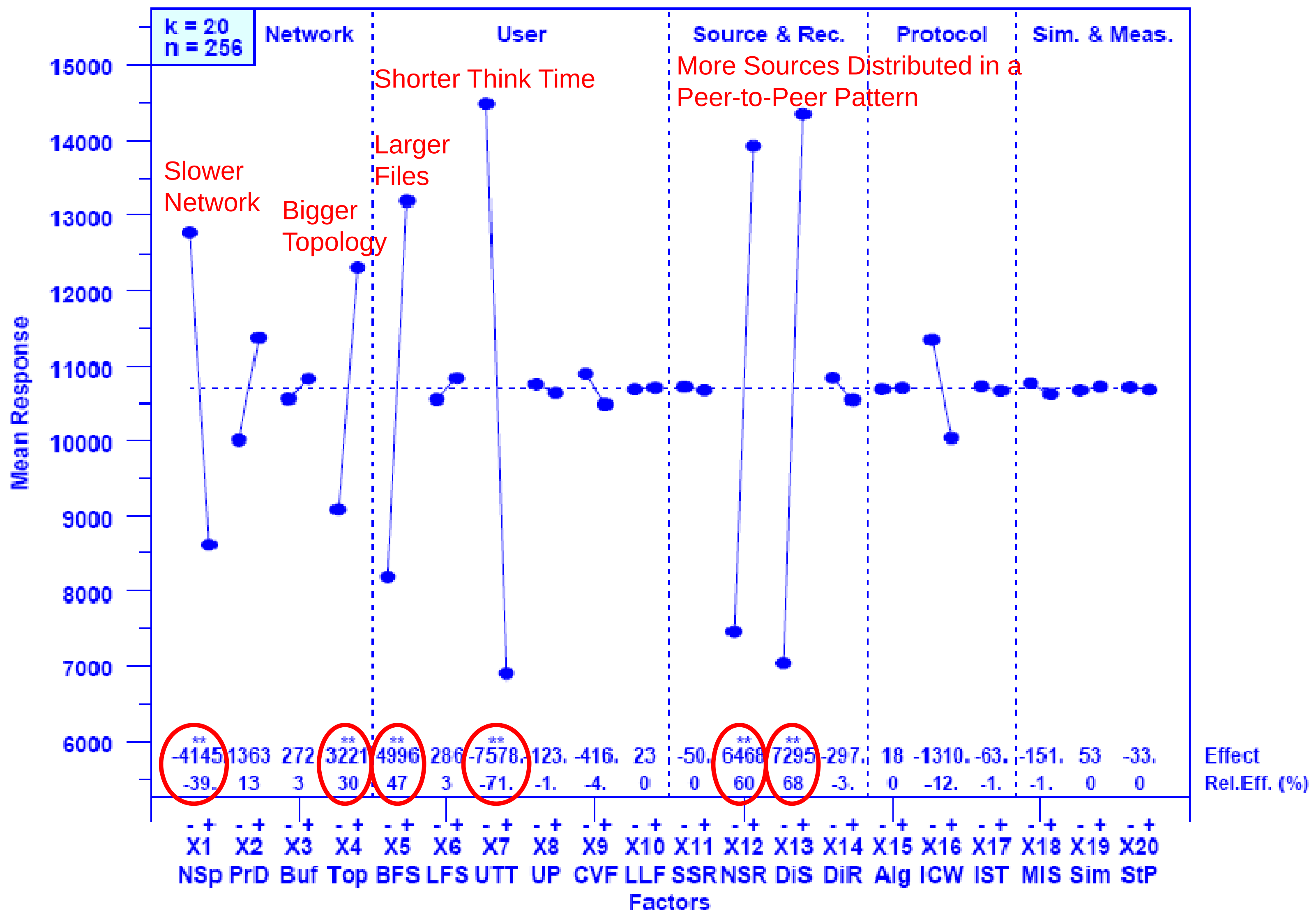
1. Main Effects Analysis
2. Two Factor Interaction Analysis
3. Tabular Summary Analysis

Sample Main Effects Analysis



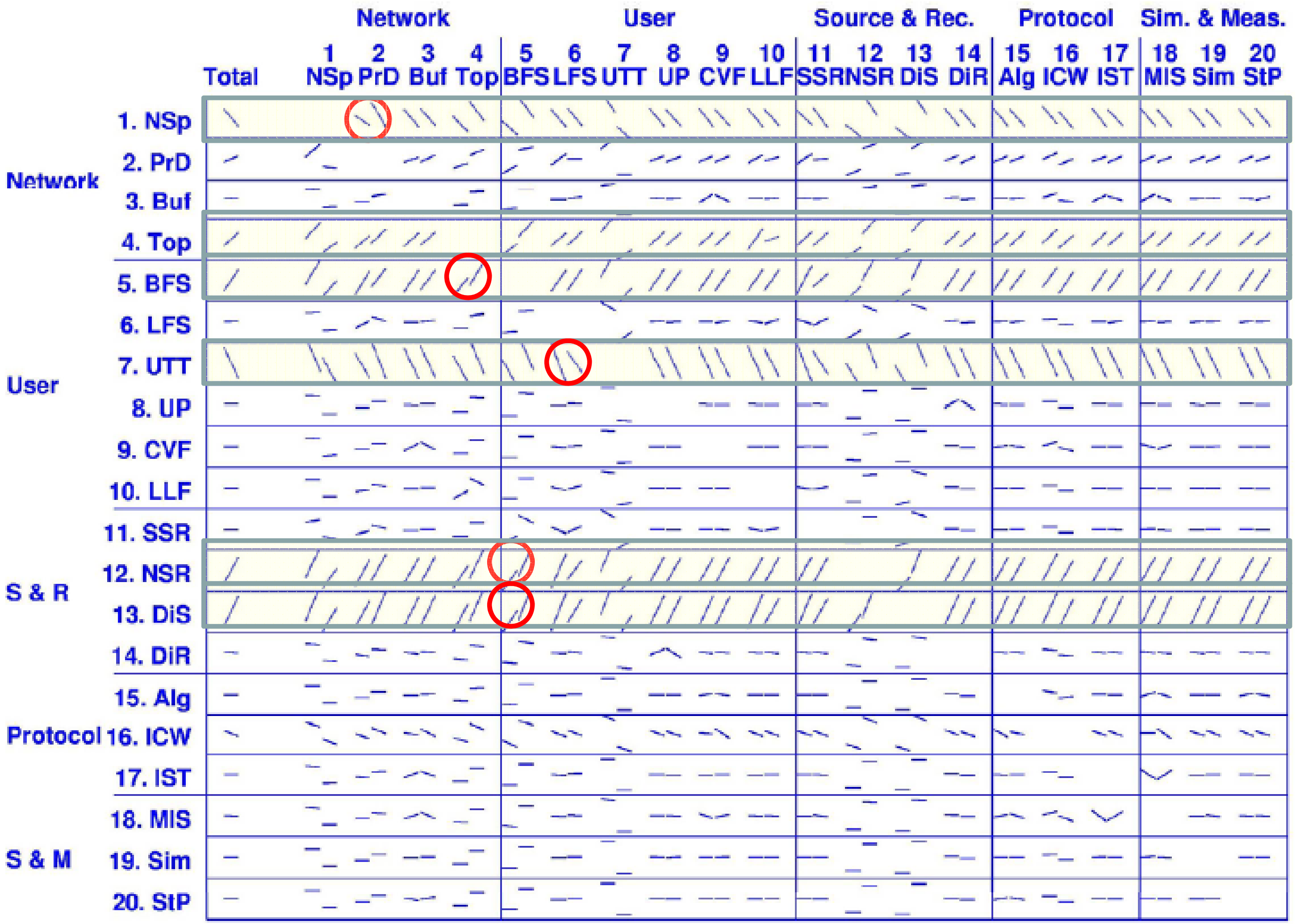
Throughput (pps) for Movies transferred over Very Fast Paths with Fast Interfaces using CTCP

Another Sample Main Effects Analysis



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

Sample Two Factor Interaction Analysis



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 Tabular Summary Analysis

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 Tabular Summary Analysis

File Type	Path Class	Connection Speed	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
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

What main factors drives *MesoNet* Response?

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

Conclusions

- 2-Level-per-Factor Orthogonal Fractional Factorial (OFF) experimental designs can reveal significant information about mesoscopic 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 6 *MesoNet* parameters

Future Work

- Apply insights from *MesoNet* sensitivity analysis to compare proposed Internet congestion control algorithms [*future presentation*]

JOINT WORK BETWEEN CxS and CNS Programs

- Develop a reduced scale simulation model for cloud computing IaaS (infrastructure-as-a-service) [*studying literature, code and deployments*]
- Conduct sensitivity analysis of IaaS model
- Compare propose IaaS resource allocation algorithms [*studying literature and code*]