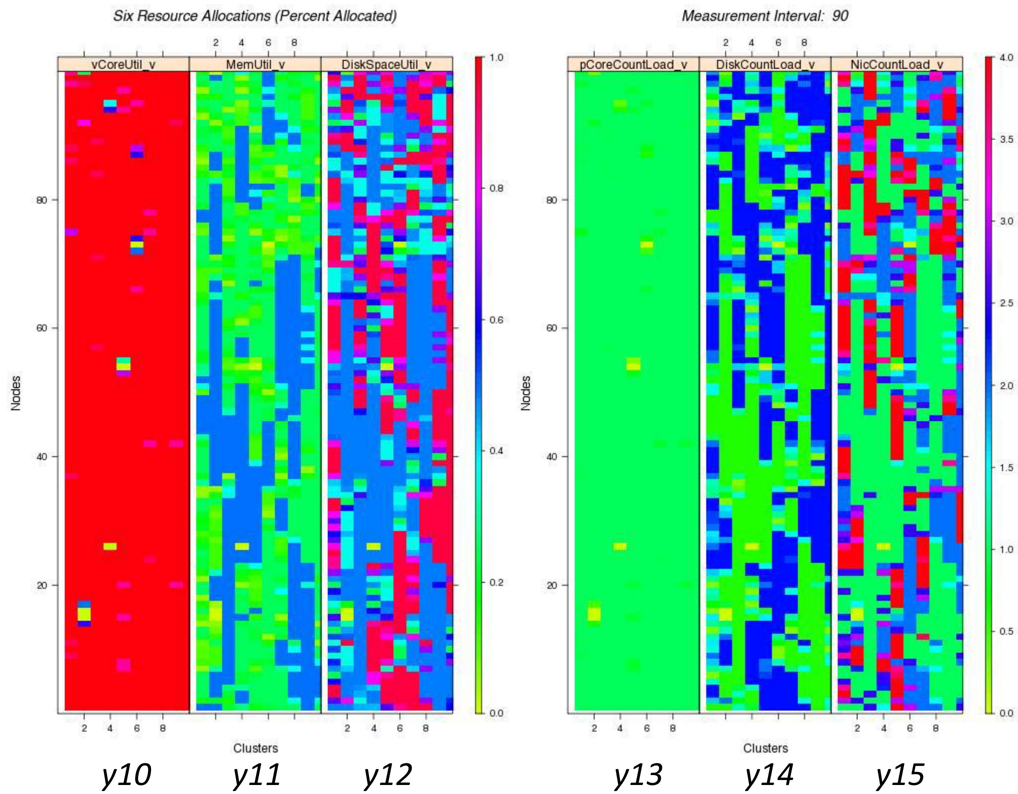
Feb. 14, 2012 NIST Presentation to LSN

Comparing VM-Placement Algorithms for On-Demand Clouds

Kevin Mills, Jim Filliben and Chris Dabrowski



Koala Information Visualizations by Sandy Ressler

(see http://math.nist.gov/~SRessler/cloudviz.html for animations and more)

NIST 2011 Papers Related to Clouds as Complex Systems

1. C. Dabrowski and K. Mills, "VM Leakage and Orphan Control in Open-Source Clouds", *Proceedings of IEEE CloudCom 2011*, Nov. 29-Dec. 1, Athens, Greece, pp. 554-559.

2. **K. Mills, J. Filliben and C. Dabrowski, "Comparing VM-Placement Algorithms for On-Demand Clouds",** ***Proceedings of IEEE CloudCom 2011*****, Nov. 29-Dec. 1, Athens, Greece, pp. 91-98.**

3. C. Dabrowski and K. Mills, "Extended Version of VM Leakage and Orphan Control in Open-Source Clouds", NIST Publication 909325; an abbreviated version of this paper was published in the *Proceedings of IEEE CloudCom 2011*, Nov. 29-Dec. 1, Athens, Greece.

4. C. Dabrowski and F. Hunt, "Identifying Failure Scenarios in Complex Systems by Perturbing Markov Chain Models", *Proceedings of ASME 2011 Conference on Pressure Vessels & Piping*, Baltimore, MD, July 17-22, 2011.

5. K. Mills, J. Filliben and C. Dabrowski, "An Efficient Sensitivity Analysis Method for Large Cloud Simulations", *Proceedings of the 4th International Cloud Computing Conference*, IEEE, Washington, D.C., July 5-9, 2011.

Synopsis

- Define a general and objective method for comparing possible VM-placement algorithms through simulation of large, on-demand infrastructure clouds.

- Demonstrate the method to compare 18 selected algorithms.

- Generate some insights regarding two-level (cluster then node) VM-placement algorithms.

- Provide evidence showing that, on average, alternative algorithms yield small quantitative differences in many model responses, but also show that selection of algorithm for choosing a cluster can lead to very large difference in provider revenue, when aggregated over time.

- Introduce Ongoing Work

We base our study on the ***Koala*** infrastructure cloud simulator.

We Developed a 4-Step Method* to Compare Resource Allocation Algorithms in Large Distributed Systems

Develop Reduced Parameter Model

Identify Most Influential Model Parameters & Significant Behaviors

Select Parameter Combinations under which to Compare Algorithms

Apply Techniques for Analyzing Multidimensional Result Datasets

THIS TALK

*Previously, we applied this method to compare congestion-control algorithms proposed for the Internet:

K. Mills, J. Filliben, D. Cho, E. Schwartz and D. Genin, Study of Proposed Internet Congestion Control Mechanisms, NIST Special Publication 500-282, May 2010, 534 pages.

First two steps,
as applied to On-Demand Clouds, reported at IEEE Cloud 2011:

K. Mills, J. Filliben and C. Dabrowski, "An Efficient Sensitivity Analysis Method for Large Cloud Simulations", *Proceedings of IEEE Cloud 2011*, July 5-9, Washington, D.C., where we introduced our Koala cloud simulator and identified six significant parameters that influence eight behavioral dimensions .

Outline

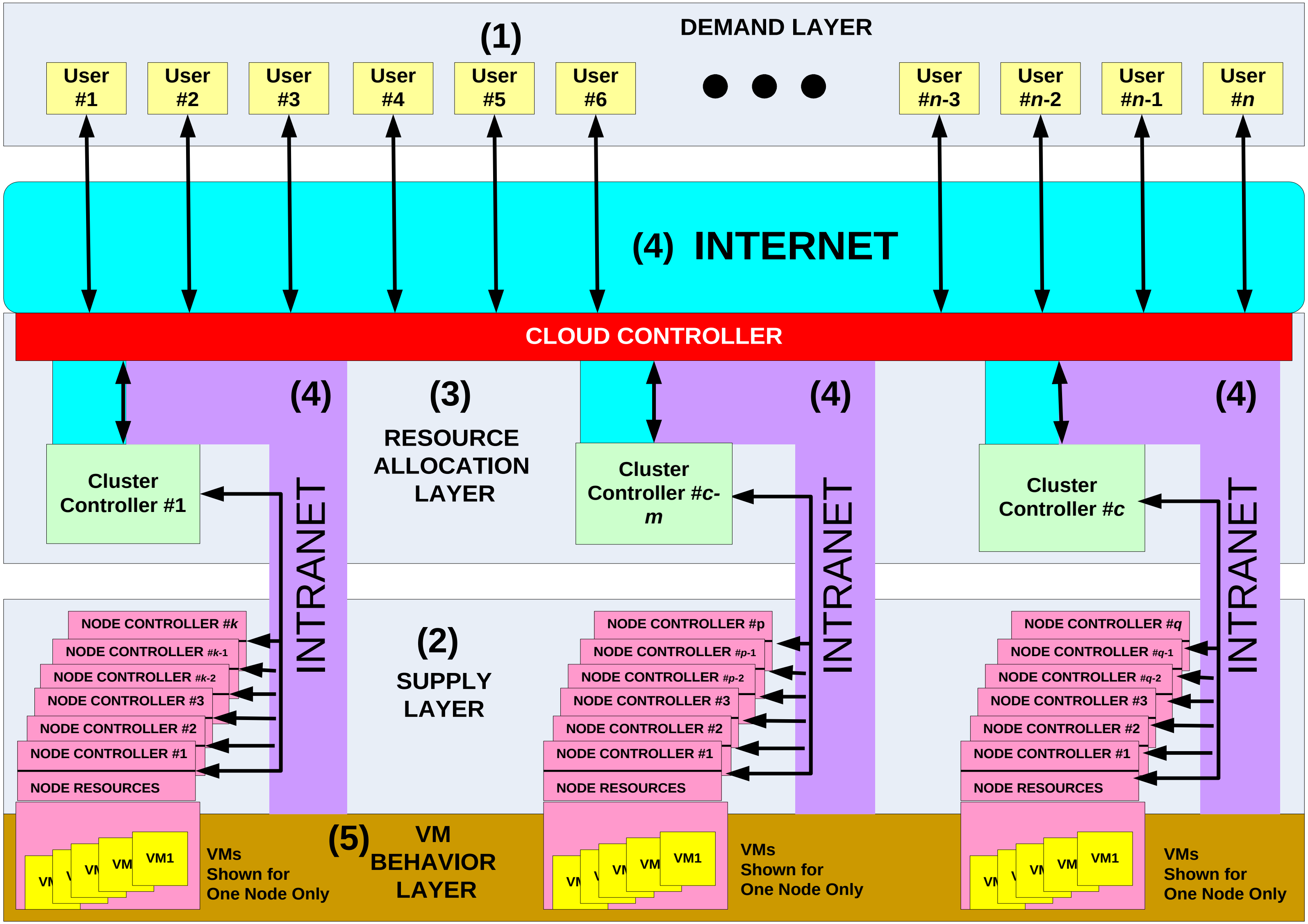
- Overview of ***Koala*** Infrastructure Cloud Simulator – 5 slides
- Experiment Design – 3 slides
- Analysis Method & Results – 3 slides
- Findings – 2 slides
- Ongoing Work – 1 slide

Outline

- Overview of *Koala* Infrastructure Cloud Simulator – 5 slides
- Experiment Design – 3 slides
- Analysis Method & Results – 3 slides
- Findings – 3 slides
- Ongoing Work – 1 slide

Overview of *Koala* Infrastructure Cloud Simulator

Schematic of *Koala* IaaS Cloud Computing Model



Virtual Machine (VM) Types* Simulated in *Koala*

VM Types are offered by the Cloud provider and requested by Cloud users

VM Type	Virtual Cores		Virtual Block Devices		# Virtual Network Interfaces	Memory (GB)	Instruct. Arch.
	#	Speed (GHz)	#	Size (GB) of Each			
M1 small	1	1.7	1	160	1	2	32-bit
M1 large	2	2	2	420	2	8	64-bit
M1 xlarge	4	2	4	420	2	16	64-bit
C1 medium	2	2.4	1	340	1	2	32-bit
C1 xlarge	8	2.4	4	420	2	8	64-bit
M2 xlarge	8	3	1	840	2	32	64-bit
M4 xlarge	8	3	2	850	2	64	64-bit

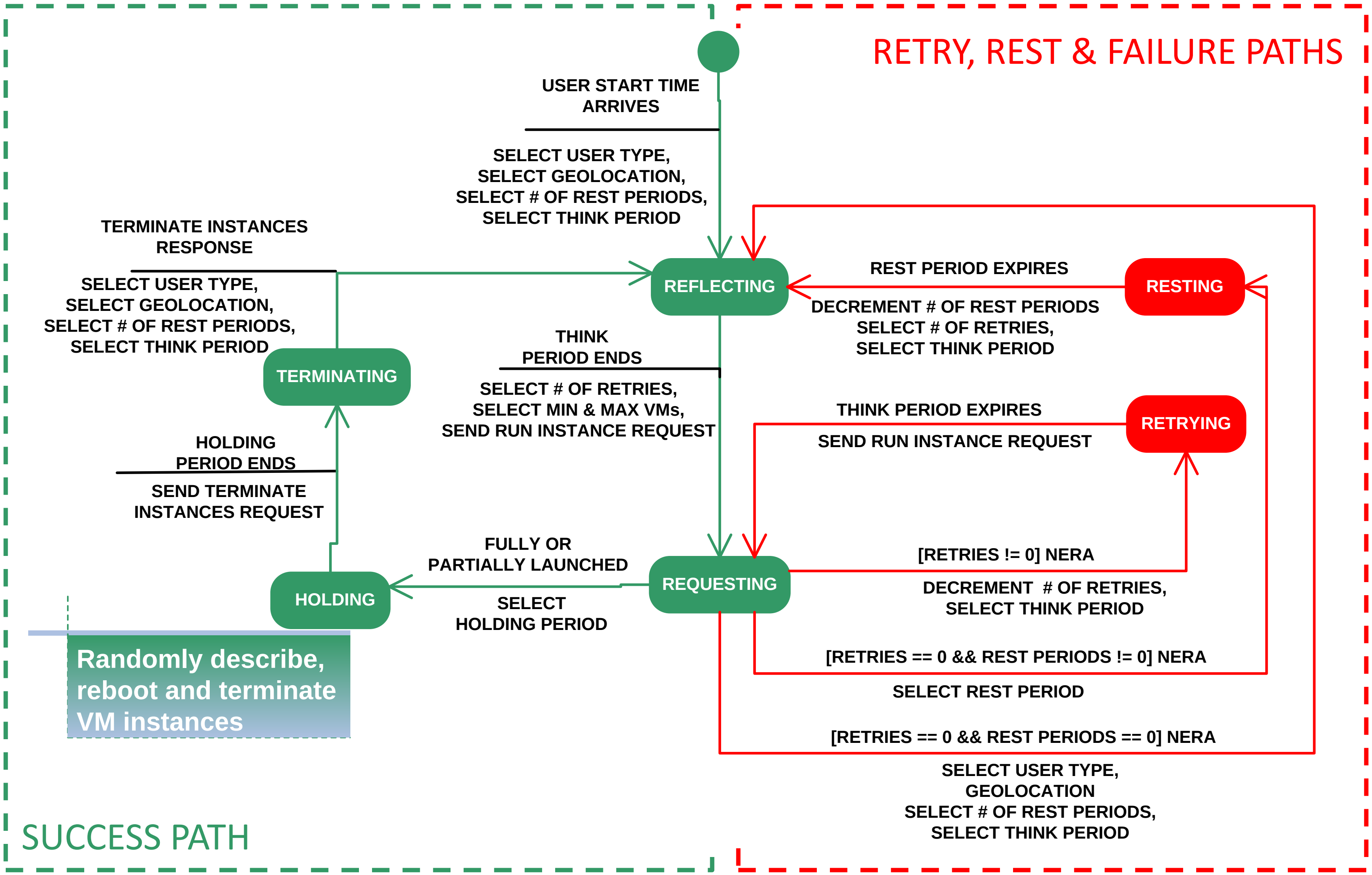
*Inspired by Amazon Elastic Compute Cloud VM Types

Description of User Types Simulated in *Koala*

We created different classes of demand, such as processing users (PU), distributed simulation users (MS), peer-to-peer users (PS), Web service users (WS) and data search users (DS)

User Type	VM Type(s)	Max-Min VMs	Max-Max VMs	User Type	VM Type(s)	Max-Min VMs	Max-Max VMs
PU1	M1 small	10	100	PS1	C1 medium	3	10
				PS2		10	50
PU3		100	500	PS3		50	100
PU5		500	1000	WS1	M1 large M2 xlarge C1 xlarge	1	3
PU2	M1 large	10	100	WS2	M1 large M2 xlarge C1 xlarge	3	9
PU4		100	500	WS3	M1 large M2 xlarge C1 xlarge	9	12
PU6		500	1000	DS1	M4 xlarge	10	100
MS1	M1 xlarge	10	100	DS2		100	500
MS3		100	500	DS3		500	1000

Finite-State Machine of Simulated User Behavior in *Koala*



Description of Selected Platform Types Simulated in *Koala*

We created 22 platform classes, inspired by
a visit to an Amazon EC2 data center – only four platform types were
used in these experiments

Platform Type	Physical Cores		Memory (GB)	# Physical Disks by Size				# Network Interfaces	Instruct. Arch.
	#	Speed (GHz)		250 GB	500 GB	750 GB	1000 GB		
C8	2	2.4	32	0	3	0	0	1	64-bit
C14	4	3	64	0	4	0	3	2	64-bit
C18	8	3	128	0	0	4	3	4	64-bit
C22	16	3	256	0	0	0	7	4	64-bit

Outline

- Overview of ***Koala*** Infrastructure Cloud Simulator – 5 slides
- Experiment Design – 3 slides
- Analysis Method & Results – 3 slides
- Findings – 3 slides
- Ongoing Work – 1 slide

Experiment Design

VM-Placement Algorithms Simulated in *Koala*

We compared 18 VM-Placement Algorithms that require two levels: (1) choosing a cluster and (2) placing VMs on nodes within that cluster.

Criteria for Choosing a Cluster		Heuristics for Choosing Nodes	
Identifier	**Criterion Name**	**Identifier**	**Heuristic Name**
LLF	**Least-Full First**	**FF**	**First Fit**
		LF	**Least-Full First**
PAL	**Percent Allocated**	**MF**	**Most-Full First**
		NF	**Next Fit**
RAN	**Random**	**RA**	**Random**
		TP	**Tag & Pack**

3 x 6 = 18

Sensitivity Analysis of *Koala* Revealed 6 Influential Parameters

Sensitivity Analysis also Guided our Choice of Two Values for Each Parameter

We compared the 18 Algorithms under 2^{6-1} = 32 conditions, chosen using Orthogonal Fractional Factorial (OFF) experiment design theory

Layer	Parameter	Parameter Name	Plus (1) Level	Minus (-1) Level
Demand Layer	*x1*	Number of users	2500	250
	x2	Probability of a user's type	PU1 = 0.20 PU2 = 0.20 PU3 = 0.10 PU4 = 0.10 MS1 = 0.10 MS3 = 0.01 PS1 = 0.10 PS2 = 0.01 WS1 = 0.15 WS2 = 0.07 WS3 = 0.03 DS1 = 0.10 DS2 = 0.01	PU1 = 1/6 PU2 = 1/6 MS1 = 1/6 PS1 = 1/6 WS1 = 1/6 DS1 = 1/6
	x3	Average (& shape) of user's holding time	8 hours (a = 1.2)	4 hours (a = 1.2)
Supply Layer	*x4*	Number of clusters	20	10
	x5	Number of nodes per cluster	1000	100
	x6	Probability of a node's platform configuration type	C22 = 1.0	C8 = 0.25 C14 = 0.25 C18 = 0.25 C22 = 0.25

Response Variables used for Experiment

We selected 42 variables that we wanted to explore, though Sensitivity Analysis indicated ***Koala*** exhibited only 8 Behavioral Dimensions:

Category	ID	Response Name	Definition
User	*y1*	User Request Rate	(Requests by All Users / # User Cycles)
	y2	NERA Rate	(NERAs / Requests by All Users)
	y3	Full Grant Rate	(Full Grants / (Full Grants + Partial Grants))
	y4	User Arrival Rate	(# User Cycles / Simulated Hours)
	y5	User Give-up Rate	(# Users that Gave Up / # User Cycles)
	y6	Grant Latency	Weighted Avg. Delay in Granting VMs to Users that Got VMs
	y40	User Success Rate	((Full Grants + Partial Grants)/# User Cycles)
	y41	Avg. Fraction VMs Obtained	(Allocated VMs/Requested VMs)
	y42	Avg. *RunInstance* Response Time	Weighted avg. for successful allocations
Cloud	*y7*	Reallocation Rate	(# Times Alternate Cluster Chosen / Requests Granted)
	y8	Full Grant Proportion	(Avg. Fraction Clusters Offering Full Grants)
	y9	NERA Proportion	(Avg. Fraction Clusters Reporting NERA)
	y10	vCore Utilization	(Avg. Fraction of Virtual Cores Used in Cloud)
	y11	Memory Utilization	(Avg. Fraction of Memory in Use in Cloud)
	y12	Disk Space Utilization	(Avg. Fraction of Disk Space in Use in Cloud)
	y13	pCore Load	(Avg. Virtual Cores Allocated / Physical Cores in Cloud)
	y14	Disk Count Load	(Avg. Virtual Disks Allocated / Physical Disks in Cloud)
	y15	NIC Count Load	(Avg. Virtual NICs Allocated / Physical NICs in Cloud)
Cluster	*y16*	vCore Utilization Variance	Avg. Variance in vCore Utilization across Clusters
	y17	Memory Utilization Variance	Avg. Variance in Memory Utilization across Clusters
	y18	Disk Space Utilization Variance	Avg. Variance in Disk Space Utilization across Clusters
	y19	pCore Load Variance	Avg. Variance in pCore Load across Clusters
	y20	Disk Count Variance	Avg. Variance in Disk Count Load across Clusters
	y21	NIC Count Variance	Avg. Variance in NIC Count Load across Clusters
	y22	Node Reallocation Rate	(# Times Alternate Node Chosen / VMs Allocated)
	y23	Cluster NERA Rate	(# NERAs / # Responses Avg. across Clusters)
	y24	Cluster Full-Grant Rate	(# Full Grants / # Responses Avg. across Clusters)
	y25	Allocation Rate	(Times Cluster chosen / Cluster offered Avg. across Clusters)
	y26	Standard Deviation-NERA	Stand. Dev. in Avg. NERA Rate across Clusters
	y27	Standard Deviation-Full-Grant	Stand. Dev. in Avg. Full-Grant Rate across Clusters
	y28	Standard Deviation-Allocation Rate	Stand. Dev. in Allocation Rate across Clusters
VMs	*y29*	Current Instances	Avg. # VM Instances Extant in Cloud
	y30	M1small Instances	Fraction of Current Instances that are M1 small VMs
	y31	M1large Instances	Fraction of Current Instances that are M1 large VMs
	y32	M1xlarge Instances	Fraction of Current Instances that are M1 xlarge VMs
	y33	C1medium Instances	Fraction of Current Instances that are C1 medium VMs
	y34	C1xlarge Instances	Fraction of Current Instances that are C1 xlarge VMs
	y35	M2xlarge Instances	Fraction of Current Instances that are M2 xlarge VMs
	y36	M4xlarge Instances	Fraction of Current Instances that are M4 xlarge VMs
Internet/ Intranet	*y37*	WS Message Rate	Avg. # WS Messages Send Per Simulated Hour
	y38	Intra-Site Messages	(# WS Messages Sent with Sites / # WS Messages Sent)
Revenue	*y39*	Aggregate Revenue in $/Hour	Calculated from *y29* through *y36* & VM prices

- *y3* – cloud-wide demand/supply
- *y4* – user arrival rate
- *y7* – reallocation rate
- *y15* – cloud-wide resource usage
- *y21* – variance in cluster load
- *y28* – variance in cluster choice
- *y29* – number of VMs
- *y31* – mix of VM types

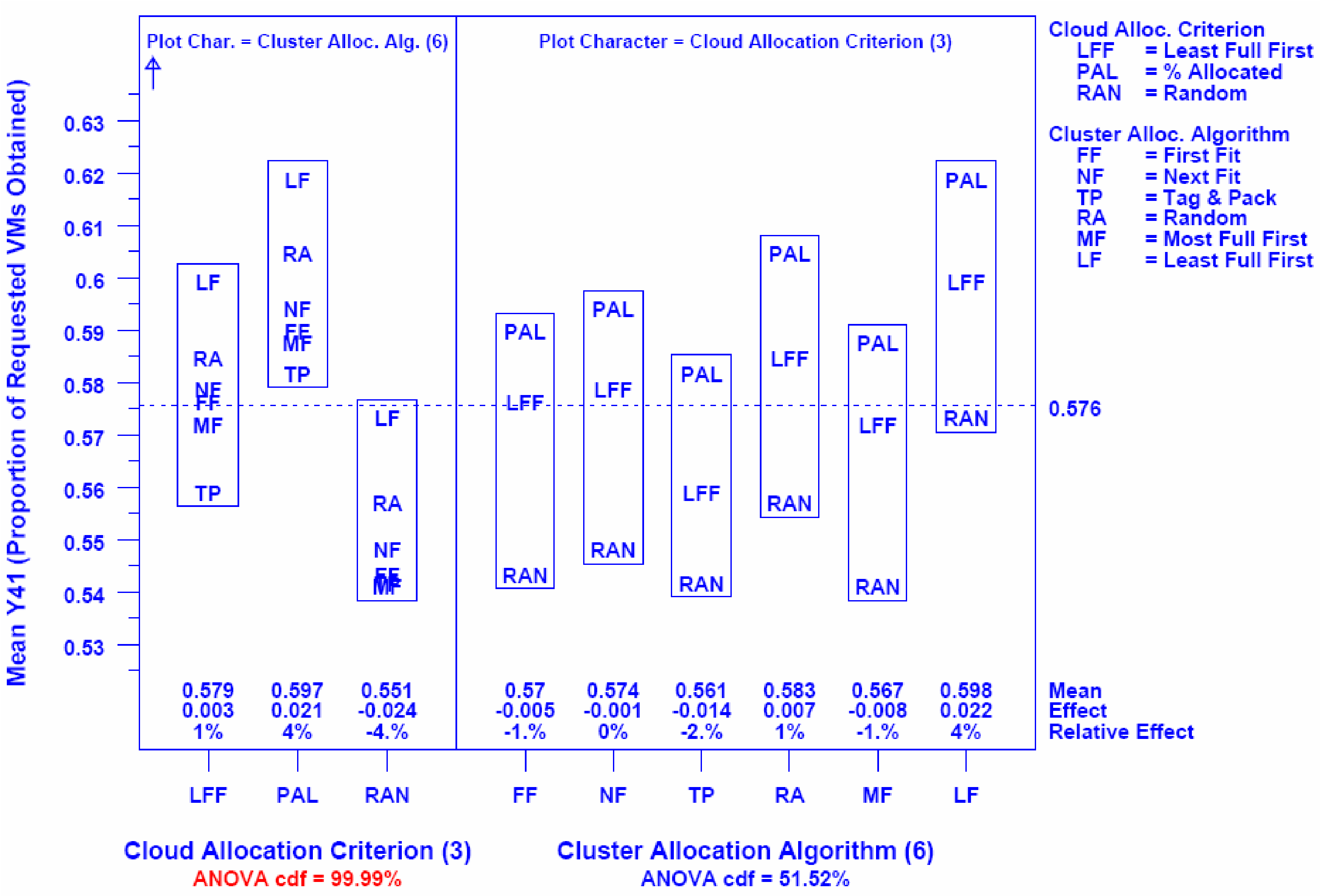
Outline

- Overview of *Koala* Infrastructure Cloud Simulator – 5 slides
- Experiment Design – 3 slides
- Analysis Method & Results – 3 slides
- Findings – 3 slides
- Ongoing Work – 1 slide

Analysis Method & Results

Used ANOVA (Analysis of Variance) to Compare Each Algorithm Level

$$F = \frac{f_2}{f_1} \cdot \frac{\sum_{i=1}^{3}\sum_{j=1}^{6}\sum_{k=1}^{32}(x_{ijk} - \bar{\bar{x}})^2}{\sum_{i=1}^{3}\sum_{j=1}^{6}\sum_{k=1}^{32}(x_{ijk} - \overline{x_k})^2}$$



Summary of 84 ANOVA Tests: 42 Responses x 2 Algorithm Levels

Category	ID	Response Name	ANOVA Cdf Cloud Crit (3)	ANOVA Cdf Cluster Alg (6)
User	*y1*	User Request Rate	99.96	62.19
	y2	NERA Rate	100	22.33
	y3	Full Grant Rate	100	2.75
	y4	User Arrival Rate	99.87	77.15
	y5	User Give-up Rate	94.63	98.6
	y6	Grant Latency	98.01	96.11
	y40	User Success Rate	95.86	98.02
	y41	Avg. Fraction VMs Obtained	99.99	51.52
	y42	Avg. *RunInstance* Response Time	37.35	97.49
Cloud	*y7*	Reallocation Rate	99.99	9.5
	y8	Full Grant Proportion	100	0.02
	y9	NERA Proportion	100	0.4
	y10	vCore Utilization	67.85	99.81
	y11	Memory Utilization	98.97	91.47
	y12	Disk Space Utilization	97.29	96.27
	y13	pCore Load	67.85	99.81
	y14	Disk Count Load	96.76	97.56
	y15	NIC Count Load	99.78	79.49
Cluster	*y16*	vCore Utilization Variance	100	1.28
	y17	Memory Utilization Variance	100	0.09
	y18	Disk Space Utilization Variance	100	0.14
	y19	pCore Load Variance	100	1.28
	y20	Disk Count Variance	100	0.42
	y21	NIC Count Variance	100	1.02
	y22	Node Reallocation Rate	100	6.09
	y23	Cluster NERA Rate	100	0.19
	y24	Cluster Full-Grant Rate	100	0.06
	y25	Allocation Rate	99.88	77.64
	y26	Standard Deviation-NERA	63.92	61.08
	y27	Standard Deviation-Full-Grant	99.73	30.95
	y28	Standard Deviation-Allocation Rate	100	0.02
VMs	*y29*	Current Instances	99.98	50.54
	y30	M1small Instances	99.99	35.85
	y31	M1large Instances	60.58	99.02
	y32	M1xlarge Instances	99.83	77.1
	y33	C1medium Instances	99.97	27.57
	y34	C1xlarge Instances	82.1	99.89
	y35	M2xlarge Instances	74.62	99.97
	y36	M4xlarge Instances	99.95	66.03
Internet/ Intranet	*y37*	WS Message Rate	99.7	83.74
	y38	Intra-Site Messages	89	99.05
Revenue	*y39*	Aggregate Revenue in $/Hour	99.99	44.51

Means for Each Response Under Each Value of Each Algorithm Level

Category	ID	LLF	PAL	RAN
User	y1	7.461	8.386	7.696
	y2	0.444	0.506	0.450
	y3	0.624	0.574	0.514
	y4	37324	35878	37170
	y5	0.066	0.074	0.067
	y6	9044	10488	9526
	y40	0.925	0.915	0.923
	y41	0.579	0.597	0.551
	y42	0.278	0.277	0.278
Cloud	y7	0.000052	0.000084	0.000057
	y8	0.438	0.332	0.389
	y9	0.481	0.587	0.537
	y10	0.774	0.791	0.783
	y11	0.188	0.197	0.199
	y12	0.413	0.428	0.418
	y13	0.774	0.791	0.783
	y14	0.964	0.997	0.948
	y15	1.591	1.645	1.554
Cluster	y16	0.0017	0.019	0.0071
	y17	0.0009	0.0034	0.0015
	y18	0.0022	0.0086	0.0038
	y19	0.0017	0.019	0.0071
	y20	0.018	0.052	0.024
	y21	0.045	0.127	0.052
	y22	0.00015	0.00015	0.00008
	y23	0.507	0.606	0.562
	y24	0.421	0.323	0.375
	y25	0.19	0.232	0.232
	y26	0.01	0.01	0.011
	y27	0.008	0.011	0.015
	y28	0.034	0.058	0.02
VMs	y29	21808	22139	20365
	y30	0.355	0.354	0.333
	y31	0.308	0.311	0.307
	y32	0.138	0.142	0.151
	y33	0.057	0.053	0.052
	y34	0.025	0.022	0.025
	y35	0.026	0.023	0.026
	y36	0.091	0.096	0.106
Internet/ Intranet	y37	60867	62677	60841
	y38	0.977	0.977	0.977
Revenue	y39	11322	11706	11624

Category	ID	FF	LF	MF	NF	TP	RA
User	y1	7.643	8.450	7.692	7.710	7.871	7.718
	y2	0.460	0.493	0.458	0.462	0.455	0.470
	y3	0.566	0.593	0.563	0.57	0.555	0.577
	y4	37138	35624	37188	36938	37051	36807
	y5	0.065	0.080	0.065	0.067	0.067	0.069
	y6	10130	8636	10439	9643	10420	8848
	y40	0.925	0.908	0.925	0.923	0.922	0.921
	y41	0.57	0.598	0.567	0.574	0.561	0.583
	y42	0.278	0.276	0.278	0.279	0.277	0.278
Cloud	y7	0.000063	0.000064	0.000068	0.000073	0.000055	0.000063
	y8	0.387	0.387	0.378	0.389	0.385	0.39
	y9	0.529	0.55	0.536	0.528	0.536	0.532
	y10	0.789	0.761	0.812	0.786	0.764	0.78
	y11	0.198	0.188	0.204	0.196	0.191	0.193
	y12	0.419	0.428	0.424	0.421	0.402	0.424
	y13	0.789	0.761	0.812	0.786	0.764	0.78
	y14	0.958	1.013	0.958	0.97	0.928	0.99
	y15	1.58	1.639	1.597	1.592	1.542	1.631
Cluster	y16	0.0085	0.008	0.0127	0.0097	0.008	0.008
	y17	0.0019	0.0020	0.0022	0.0019	0.0019	0.0017
	y18	0.0045	0.0054	0.0053	0.0050	0.0046	0.0045
	y19	0.0085	0.0089	0.0127	0.0097	0.0080	0.0080
	y20	0.029	0.036	0.032	0.032	0.029	0.029
	y21	0.067	0.088	0.080	0.074	0.065	0.073
	y22	0.00013	0.00012	0.00013	0.00014	0.00011	0.00012
	y23	0.555	0.569	0.562	0.552	0.558	0.553
	y24	0.373	0.375	0.364	0.376	0.373	0.378
	y25	0.228	0.192	0.237	0.216	0.232	0.201
	y26	0.011	0.009	0.013	0.010	0.010	0.009
	y27	0.012	0.010	0.015	0.011	0.012	0.010
	y28	0.037	0.040	0.037	0.037	0.035	0.038
VMs	y29	21237	22244	21020	21409	20824	21888
	y30	0.344	0.356	0.342	0.348	0.341	0.352
	y31	0.306	0.315	0.304	0.305	0.311	0.312
	y32	0.144	0.149	0.145	0.147	0.135	0.142
	y33	0.054	0.053	0.053	0.053	0.056	0.054
	y34	0.025	0.018	0.026	0.024	0.027	0.022
	y35	0.027	0.019	0.028	0.026	0.029	0.023
	y36	0.100	0.090	0.103	0.097	0.101	0.095
Internet/ Intranet	y37	61018	63016	61223	61156	60571	61785
	y38	0.977	0.977	0.977	0.977	0.976	0.977
Revenue	y39	11603	11529	11683	11587	11362	11541

Outline

- Overview of ***Koala*** Infrastructure Cloud Simulator – 5 slides
- Experiment Design – 3 slides
- Analysis Method & Results – 3 slides
- Findings – 3 slides
- Ongoing Work – 1 slide

Findings

Choice of Cluster has Larger Influence on System Behavior than Choice of Node

- Cluster choice caused significant differences in 79% of responses, covering 100% of the eight behavioral dimensions *Koala* exhibits
- Node selection influenced only 29% of responses, covering only one of the eight behavioral dimensions *Koala* exhibits
- Percent-Allocation (PAL) cluster choice generates an average of $384/hour more revenue for the cloud provider, which, when aggregated over a year, reaches about $3.4M more than Least-Full First (LFF)
- On the other hand, PAL has an overall harmful effect on the general population of users, who receive more negative responses and must retry more, incurring on average 20 minutes more waiting time to obtain VMs
- PAL serves fewer users but gives each served user a larger proportion of their requested VMs, and also increases variance in resource loads and utilizations

Choice of Node Influences Only a Few Responses

- Least-Full First (LF) and Tag-and-Pack (TP) lead to lower cloud-wide virtual core utilization because these heuristics more often choose empty nodes
- On the other hand, LF tends to squeeze out some larger VM types – by tagging nodes TP avoids this behavior
- LF and Random (RA) lead to lower grant latencies, because these heuristics allow successful users to acquire VMs with one fewer retries, on average

Outline

- Overview of ***Koala*** Infrastructure Cloud Simulator – 5 slides
- Experiment Design – 3 slides
- Analysis Method & Results – 3 slides
- Findings – 3 slides
- Ongoing Work – 1 slide

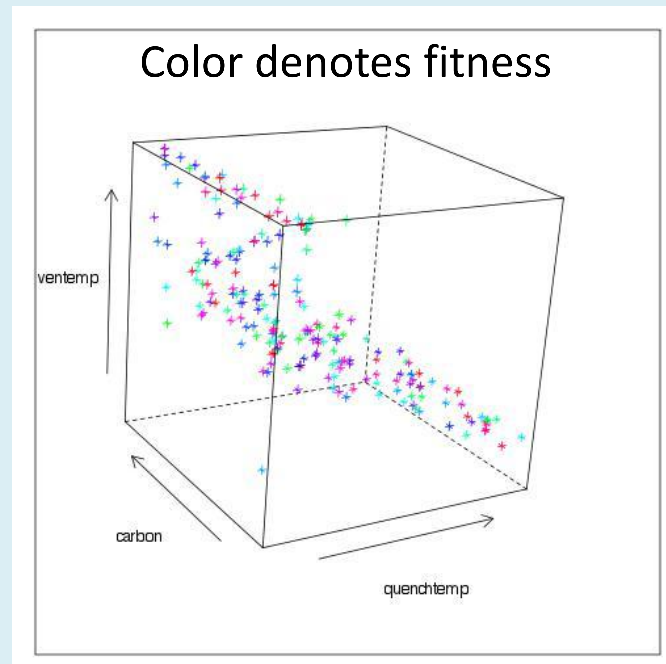
Ongoing Work

Ongoing Work – Related to Cloud Reliability

1. ***Characterizing the Effects of Stressful Conditions*** (e.g., asymmetries, dynamics and failures) on IaaS Clouds

	Parameter Summary
x1	Cluster Distribution around the Internet (same site or unique sites)
x2	Platform Types per Cluster (fixed or random probabilities)
x3	Node Failure (supply nodes fail more or less frequently)
x4	Absolute Cluster Size Variation (fewer larger clusters or more smaller clusters)
x5	Relative Cluster Size Variation (uniform clusters or some large and some small)
x6	Cloud Reconfiguration (cloud adds or subtracts clusters or not)
x7	Cluster Reconfiguration (clusters add or subtract nodes or not)
x8	Variability in Inter-site Communication Delays (very long delays vs. typical delays)
x9	Variability in Intra-site Communication Delays (very long delays vs. typical delays)
x10	Failure of Node Components (VCPUs, Memory and Disks fail and recover more or less frequently)
x11	Starting Load (100% or 50%)
x12	Time Varying User Type Probability Map (switching user type maps vs. fixed user type map)
x13	User VM Demand Changes (users grow or shrink number of VMs during holding time or do not)
x14	Probability Bogus User Request (high or low probability of user generating invalid request)
x15	probability Node NERA (high or low probability that a node reneges on accepting a VM)
x16	probability Inter-Site Message Loss (high or low probability of message loss on the Internet)
x17	probability Intra-Site Message Loss (high or low probability of message loss on Intranets)
x18	Cluster Communication Cut Function (high or low probability of cuts in communication with clusters)

2. Applying Anti-Optimization and Directed Search (e.g., Genetic Algorithms) to ***Predict Catastrophic Failure Scenarios*** in IaaS Clouds



Questions?

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

Contact information about Cloud Information Visualization:
sressler@nist.gov

Contact information about NIST Cloud Computing Program:
dawn.leaf@nist.gov

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