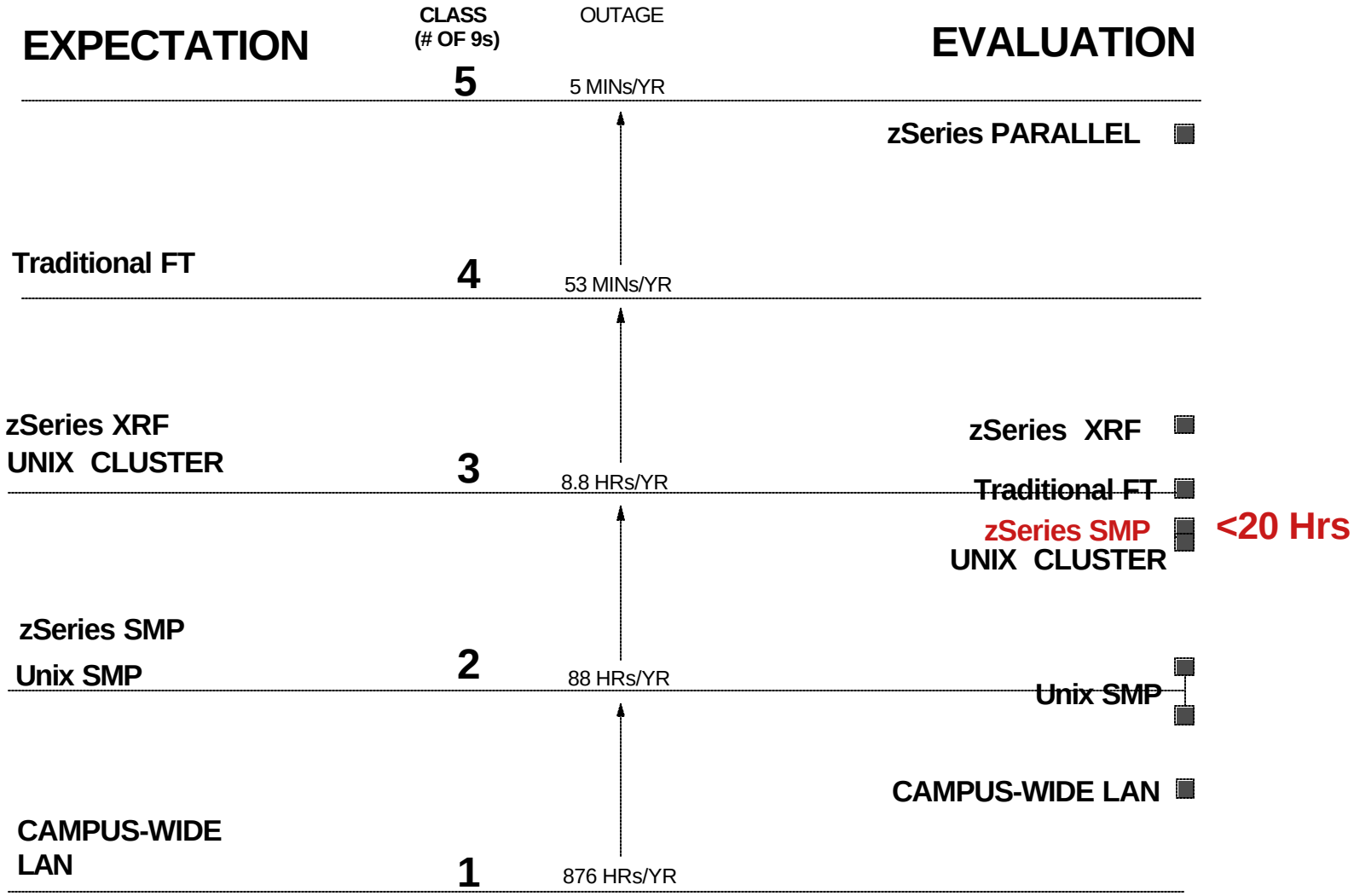
Two 45 minute talks in 20 minutes

- Server Lessons & Opportunities
- Autonomic Computing Challenge

Lisa Spainhower

AVAILABILITY BY SYSTEM TYPE



Downtime in a poorly-managed zSeries LPAR

	Impact events	Events	
ATTRIBUTION	# OUTAGES	# OUTAGES	IMPACT RATIO
Control Center	70	24	2.9
Environment	18	5	3.6
Hardware	10	1	**10**
Software	118*	52	2.3
Total	216	82	2.6

ATTRIBUTION	OUTAGE (min)	OUTAGE (min)	IMPACT RATIO
Control Center	5202	**1949**	2.7
Environment	1275	454	2.8
Hardware	875	88	10
Software	6209	**3062**	2.0
Total	13561	5553 **226 Hrs**	2.4

*TM-56%, Apps-16%, DBA-14%, OS- 6%, other- 8%

Downtime in a poorly-managed zSeries LPAR

Total Outage per log: 226 hours
Per one outage/event: 93 hours

✓# 1 Contributor is software: product & process

1818 process
453 product
791 uncertain
3062 total (51 hours)

Assume all CC outages are process (1949 min.; 32 hours)

✓68%-82% of all unscheduled outages are process
(63-76 hours)

✓Technology - HW/SW - 10-24%
(9-22 hours)

Aggregated UNIX server data

Downtime Cause %	UNIX Standalone	UNIX Cluster
Hardware	**42**	**46**
Software	34	36
Other	24	18

Data from a very large well-managed Unix customer

% Unplanned downtime	
HW	**43.8**
OS	7.8
App	7.3
Com link	18.8
DB	2.0
Environment	0.7
Supplier	0
Op tools	1.8
Process	6.2
Org/structure	0.8
Human error	5.0
Other	5.8

% Planned downtime	
HW maint	26.5
OS install	1.6
App Release	34
Com link	2.5
DB admin, BU	31.5
Dis Rec Test	0
Pwr Test	0.2
Other	3.7

zSeries Hardware Fault Tolerant Design Results

- Mean Time To Repair Action = 10 months

  3Q01 full fleet average

  MTTF < MTTRA

    Fails for which there is no associated repair include:

      Cache line sparing

      Memory chip sparing

      Dynamic CPU sparing

- Mean Time to Application Loss = 30 years

  82% of repairs are online

  15% of repairs are deferred/scheduled

  3% of repairs cause crash of customer application

Lessons from the 90s

- Management discipline is critical to HA
- Fault tolerant servers make a difference
- Clusters are difficult to implement

Challenges for the 00s

- Increased importance of firmware
- Circuit failure mechanisms
- State encapsulation
- On-the-fly change
- Dynamic resource allocation
- Configuration validation

What is Autonomic Computing?

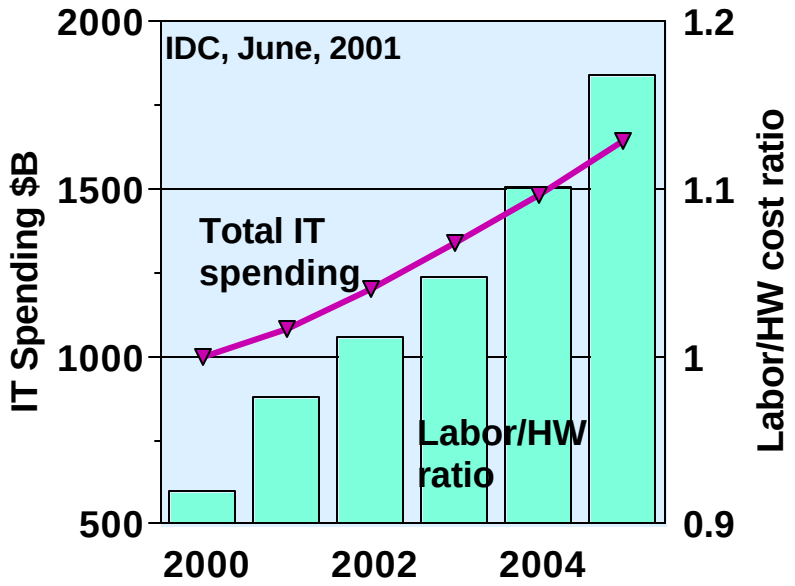
What is Project eLiza?

Anolis sagrei - 2,117 darwins

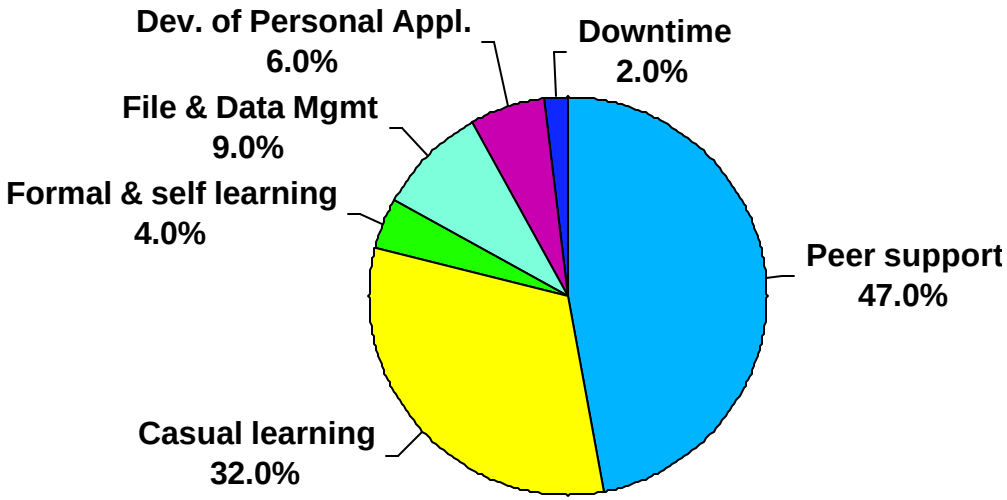
People Costs are Growing

Labor related costs are escalating, and in some cases are dominant

Worldwide I/T Spending



Indirect TCO Factors

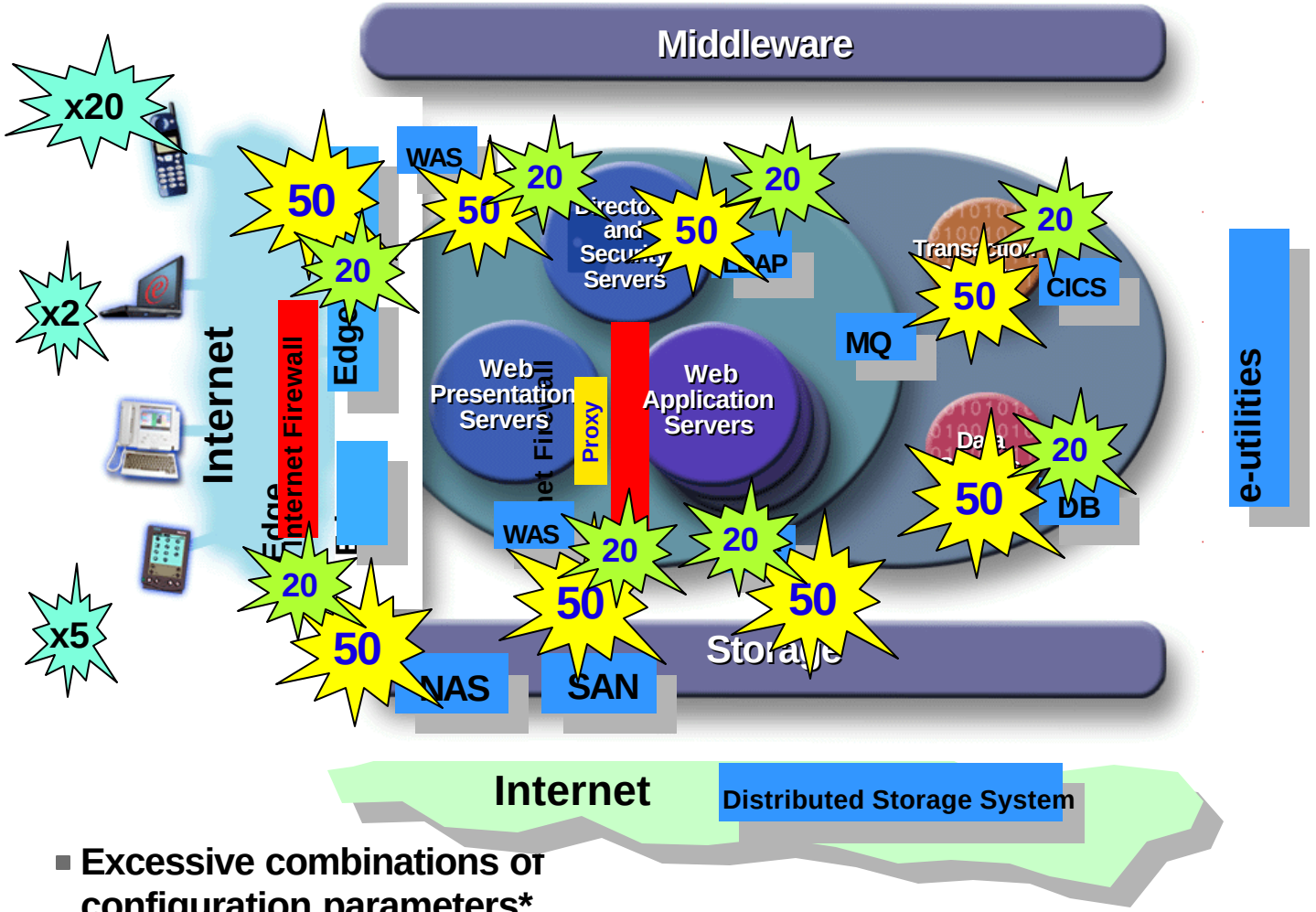


- **In storage segment, labor cost is already dominant.**
  - **labor/HW cost ratio approaching 3X***

- **Indirect costs may contribute > 60% of overall TCO** (Gartner, 3Q99)

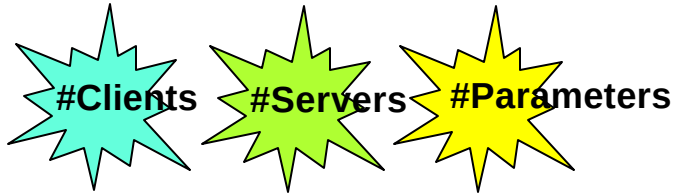
* based on $120K/person, storage HW @ $120K/TB with 4 year life and 2001 ITCentrix survey result of 0.83 person-year/TB

The System Complexity Challenge

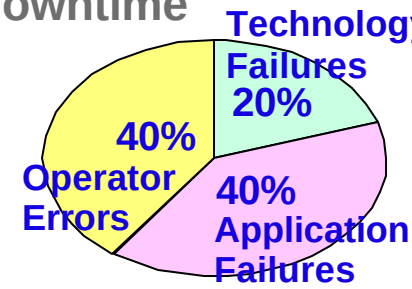


- **Excessive combinations of configuration parameters***
- **Unknown dependencies**
- **HW & SW version control issues**
- **Increased security exposure**

* not all combinations need to be independent



Complexity is a major contributor to unplanned application downtime

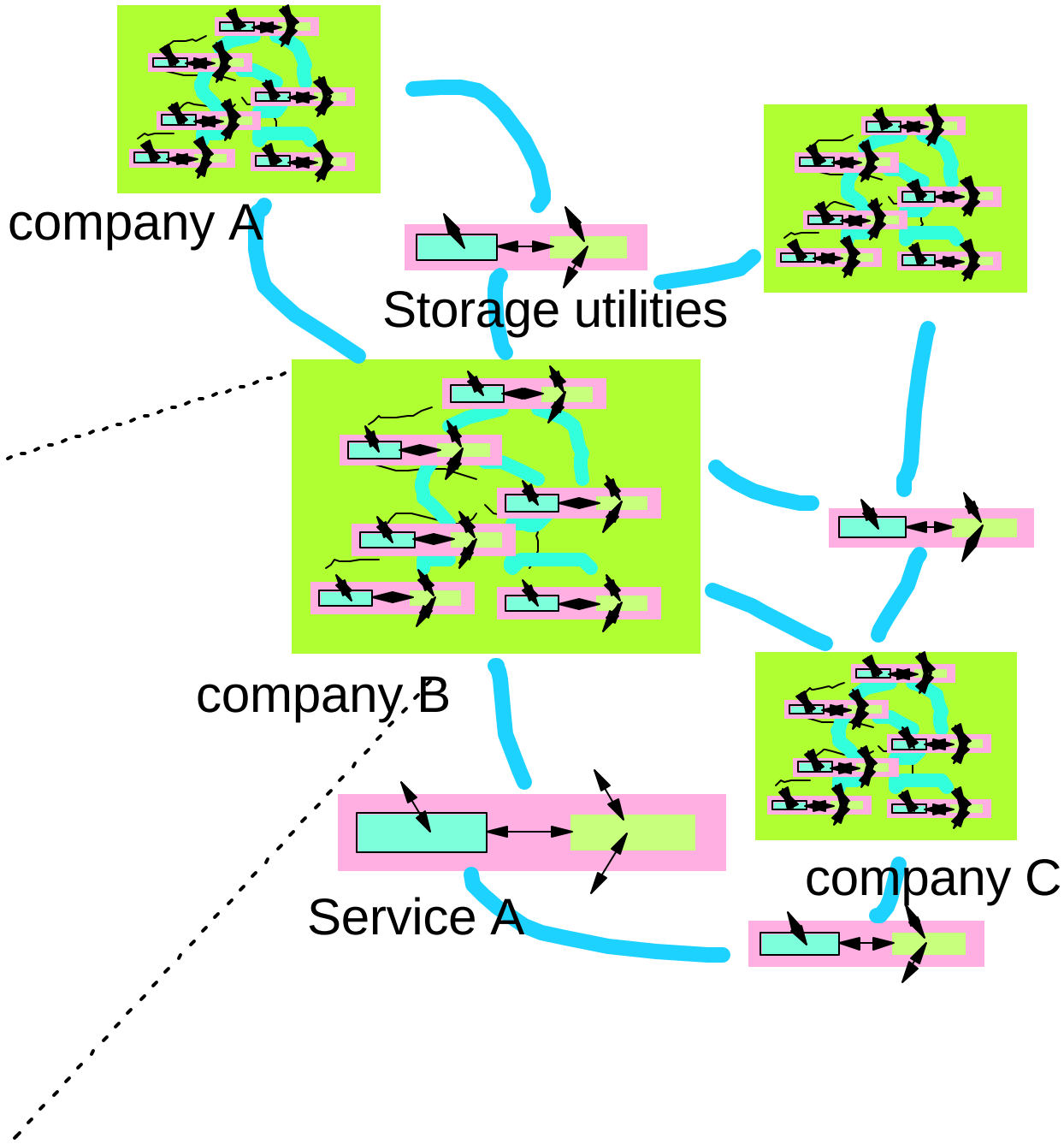


Charles Schwab & Co.
Upgrades/Operator Errors
$70 million new investment.

America Online
Maintenance/Human Error
Cost: $3 million in rebates

AT&T: Software Upgrade
$40 million in rebates

Federated Systems

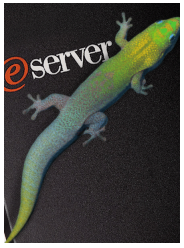


Survivability

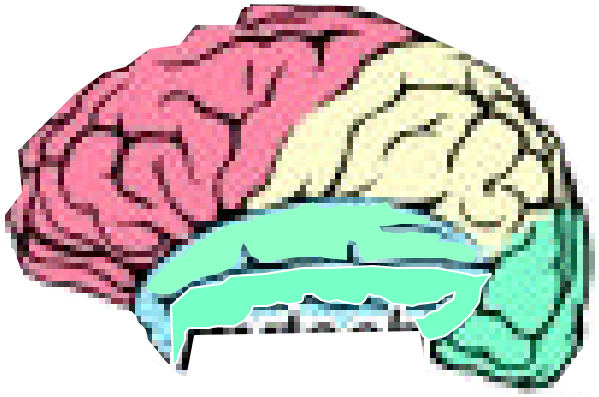
"The ability of a system to dynamically fulfill its mission

- ✓ **in a timely manner** → **self-optimizing**
- ✓ **in the presence of attacks,** → **self-protecting**
- ✓ **changes,** → **self-configuring**
- ✓ **failures or accidents."** → **self-healing**

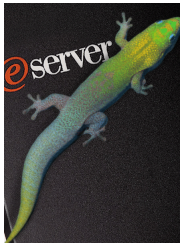
Howard Lipson CERT/CC
(with additions by LS)



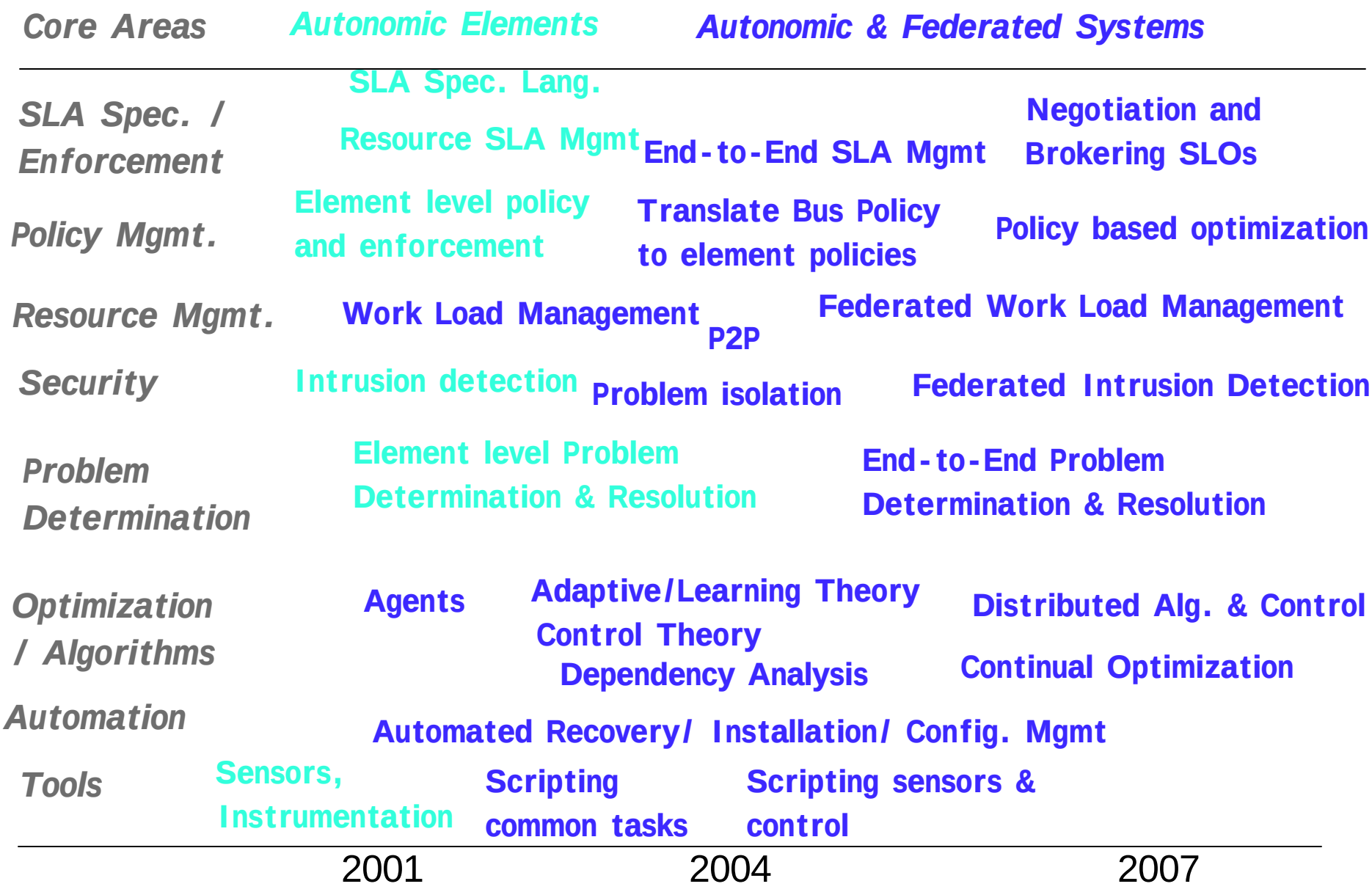
Triune Brain



BRAIN	FUNCTION	eLIZA
Human	Reason	Classic AI Apps & Service
Mammal	Emotion	People / IT staff
Lizard Computing	Survival	Autonomic Self-managing



Technology Roadmap for Autonomic Computing



The (work in progress) Challenge:

Create a global system of interoperable services that can be selected and provisioned dynamically in response to changing business conditions, enabling failure recovery and optimization and ensuring privacy and security, automatically and transparently to the businesses that use these services.