

Metrics in MMP Development and Operations

Larry Mellon
GDC, Spring 2004

Talk Checklist

- Outline *complete*
- Key Points *complete*
- Text draft *complete*
- Rehearsals *incomplete*
 - Add Visuals during Rehearsals *incomplete*
 - Neck down #esper screenshots *incomplete*

Metrics: A Powerful Servant

“I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot express it in numbers your knowledge is a meagre and unsatisfactory kind”

Lord Kelvin, addressing the Institution of Civil Engineers in 1883

But A Dangerous Master

“Figures often beguile me, particularly when I have the arranging of them myself; in which case the remark attributed to Disraeli would often apply with justice and force: “There are three kinds of lies: lies, damned lies and statistics.”

Autobiography of Mark Twain

- What you measure becomes what you optimize: pick carefully
- Cross check the numbers
- GI/GO

What level of Metrics do you need?

- [visual] Scale: Lord Kelvin vs Mark Twain
 - LK: complexity of a system under study requires fine-grain visibility into many variables
 - MT: “Practical Man” measurements: cut to fit, “good enough”, “roughly correct”
- Big metrics systems are expensive
 - Don’t go postal (unless you need to)
 - Build no more than you need (why measure beyond what you care about for either precision, frequency, depth or breadth)

MMP: Go Postal...

- Complexity of implementation
 - Butterfly effect
 - Number of moving parts
- Service business: need to reduce running costs
- Complex social / economic systems
 - Player data essential for design feedback loop

Complex Distributed System

- Hundreds to thousands of processes
- Dynamic, complex inputs
- Realtime constraints
- Hackers

Debugging / optimizing at either micro or macro levels are tricky propositions...

Resource Utilization

- All CPUs must be doing something useful&efficient, all the time
 - Highly dependent on input (the 2nd reason for embedded profilers: what user behaviour is driving this <event> we're seeing)
- Intrinsic scalability: what is the app demanding?
- Achieved scalability: how well is the infrastructure doing against the theoretical ceiling for a given app?

Complex: Social / Economic

- What do people do in-game?
- Where does their in-game money come from?
- What do they spend it on?
- Why?
- “The need to please”
 - What aspects of the game are used the most
 - Are people having fun, right now
- Tuning the gameplay

Service Oriented Business

- Driving Requirements: high reliability & performance
- ROI (value to customer vs cost to build&run)
- Player base (CRM / data mining)
 - Who costs money
 - Who generates money
- Minimize overhead
 - Where do the operational costs go?
 - What costs money
 - What generates money
- Customer Service
 - Who’s being a dick?

“How much fun are people having, and what can we do to make them have more fun?”

Marketing / Community Reps

- Tracking player behaviour
 - \$\$ in, \$\$ out
 - Where do they spend their time
- Tracking results of in-game sponsorship
 - MacDonald's object
- Teasers for marketing & community
 - New Year's Eve: Kisses
- Tracking & guiding community
 - "Metrics that matter"
 - Calvin's Creek: tips

Casinos: Similar Approach

- Highly successful
- Increased revenue per instrumented players
- Lowered costs / Increased profits
- ...

Harrah's "Total Reward"

- One of the biggest success stories for CRM is in fact a sibling game industry: casinos. It is, in fact, the only visible sign of one of the most successful computer-based loyalty schemes ever seen.
- well on the way to becoming a classic business school story to illustrate the transformational use of information technology
 - 26% of customers generate 82% of revenues
 - "Millionaire Maker," which ties regional properties to select "destination" properties through a slot machine contest held at all of Harrah's sites. Satre makes a personal invitation to the company's most loyal customers to participate, and winners of the regional tournaments then fly out to a destination property, such as Lake Tahoe, to participate in the finals. Each one of these contests is independently a valuable promotion and profitable event for each property
 - \$286.3 million in such comps. Harrah's might award hotel vouchers to out-of-state guests, while free show tickets would be more appropriate for customers who make day trips to the casino
- At a Gartner Group conference on CRM in Chicago in September 1999, Tracy Austin highlighted the key areas of benefits and the ROI achieved in the first several years of utilizing the 'patron database' and the 'marketing workbench' (data warehouse). "We have achieved over \$74 million in returns during our first few years of utilizing these exciting new tool and CRM processes within our entire organization"
- John Boushy, CIO of Harrah's, in a speech at the DCI CRM Conference in Chicago in February 2000, stated: "We are achieving over 50% annual return-on-investment in our data warehousing and patron database activities. This is one of the best investments that we have ever made as a corporation and will prove to forge key new business strategies and opportunities in the future."

Driving Requirements

- Ease of use & "Information Management"
 - Adding probes
 - Point&click to find things, speed
 - Automated aggregation of data
- Low RT overhead
 - Don't disrupt the servers under study
 - Positive feedback loops
 - Shrodinger's cat dilemma
 - But, still need massive volumes of information
- Common Infrastructure
 - Less code (at one point, there were about 3 metrics systems)
 - Bonus: allows direct comparison of user actions to load spikes
- [chart: data per event & city to show scope of prob]

Outline

- Background [done]
- Implementation Overview
- Applications of Metrics in TSO
- Wrapup
 - Lessons Learned
 - Conclusions
- Questions

Impl Overview

- Present summary views of data
 - Patterns, collections, comparisons
 - Viewable in timeOrder or dailySummary (e.g. N.Y.Eve kiss charts) (e.g. oscillating out of control & crash, then zoom in on where)
 - Drill-down where required
- Extensible: data-driven, self-organizing
- Hierarchies of views
 - Per process, av per processClass, av per CPU (running N processes)
 - Gives you system & process views, and aggregate one higher to “trouble <here>” triggers&displays
- Basic collection patterns
 - Sum, av, sample_rate, ...
- Summary data means we can collect aggregate-only data: it's most of what you need, and is far cheaper

Esper, v.4

- **Parallel & distributed simulation tool**
 - Hundreds of processors, thousands to ten's of thousands of CPU-consuming unpredictable entities, all in one space
- **Performance optimization**
 - First Esper was just automation to dig thru & summarize 100's of Megs of log files to show me the key patterns (things that point at where a big problem might be living)
 - Needed to correlate against entity actions (heavily drove performance, needed to understand the patterns to optimize the infrastructure) and sometimes change or restrict the entity actions (flow control @ user action level)
- **This Esper dispenses with the raw data phase: probes collect @ the aggregate level**

Implementation Approach: Overview

- **esperProbes: internal to every server process**
 - Count/average values inside a fixed time window
 - Log out values @ end of time_window, reset probes
- **esperFetch: sweeps esper.logs from all processes**
 - Aggregates similar values across process types & probe types
 - Compresses & reports aggregate & process-level data
- **esperDB: auto-register new data & new probe types**
- **DBImporter: many useful items are in the cityDB**
- **esperView: web front end to DB**
 - Standard set of views posted to "Daily Reports" page
 - Flexible report_generator to gen new charts
 - Caching of large graphs (used in turn for archiving)
 - Noise filters (something big you just don't care about right now)

Probe syntax

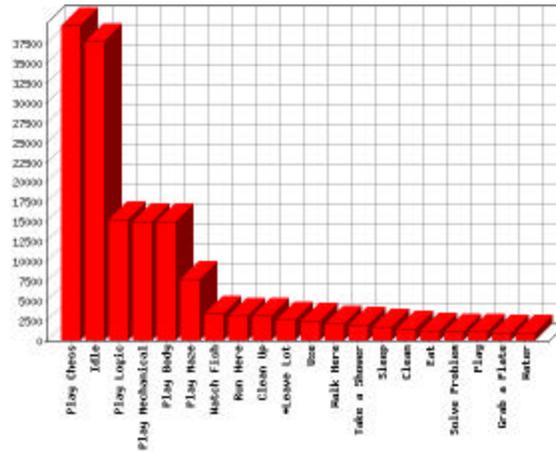
- Name_1.2.3.4 hierarchy
 - Object.interaction.social gets you three types of data from one probe
 - Data driven @ each level
- [pull code snippet for 2 or 3 probes]
- Human-readable intermediate files

Section: Uses of Metrics

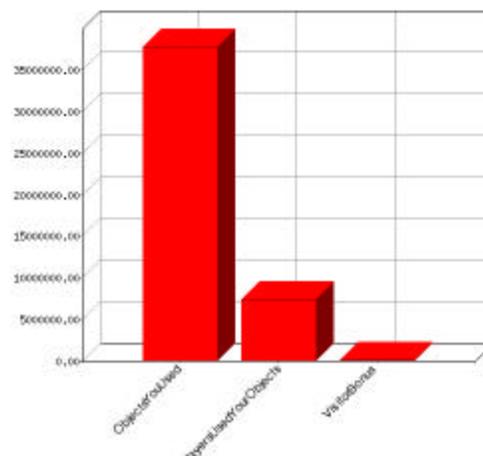
- Load testing
- Player observation
- ...
- [about these charts]
 - The screenshots don't display well, so grab the most meaningful ones & redo in PPT.
 - Sift thru the screenshots for one per type of metrics application

Object Interactions (1st cut)

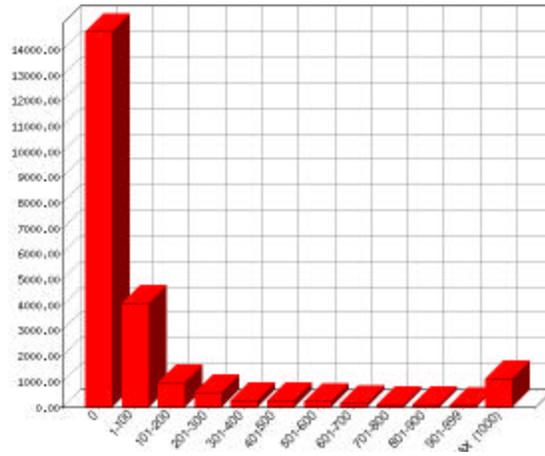
Note the metrics “bug” on top-2



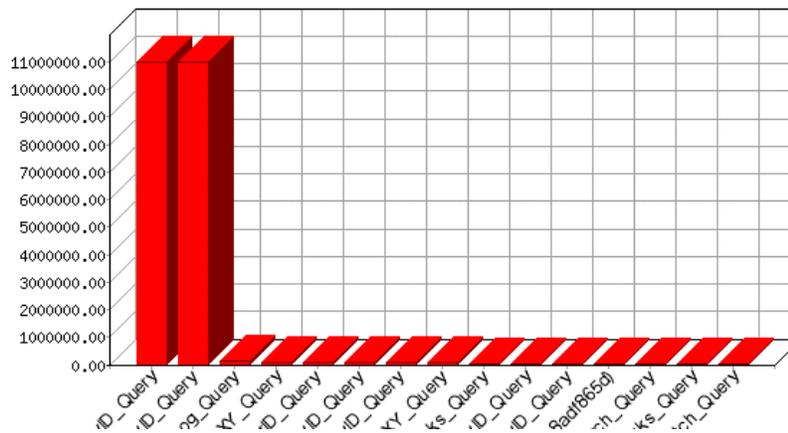
An Unbalanced Economy



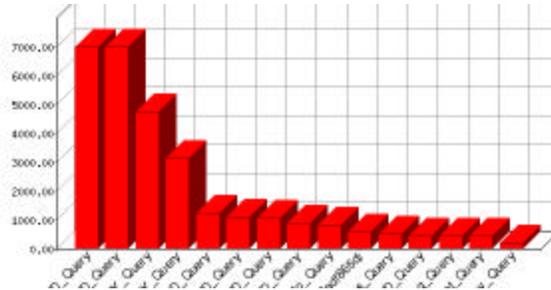
Visitor Bonus (by age of Lot)



DB Concentrator: Prod



DBC: Live



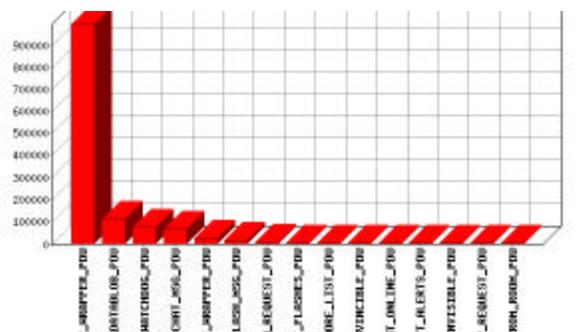
Final totals: NYEve: Kiss Count

Alphaville All Cities (extrapolated)

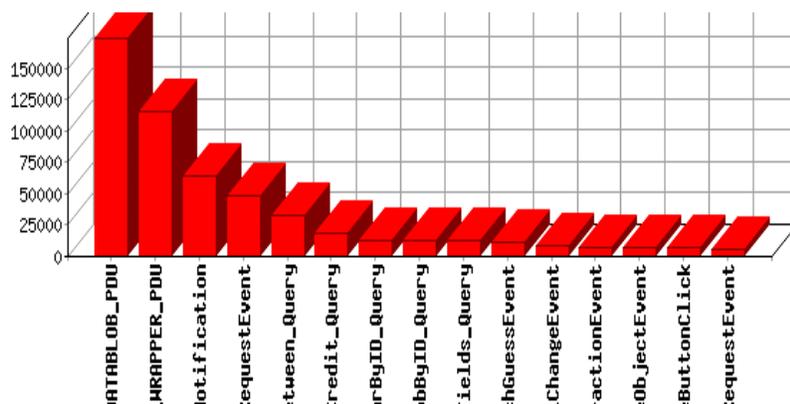
	Alphaville	All Cities (extrapolated)
New Year's Kiss	32,560	271,333
Be Kissed Hotly	7,674	63,950
Be Kissed	5,658	47,150
Be Kissed Sweetly	2,967	24,725
Blow a Kiss	1,639	13,658
Be Kissed Hello	1,161	9,675
Have Hand Kissed	415	3,458
Total	52,074	433,949

Active time range for the New Year's Kiss on Alphaville was 09:00:00
12/31/02 to 11:59:59 1/1/03:

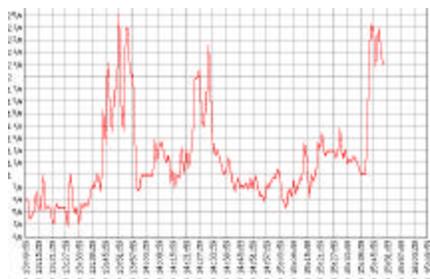
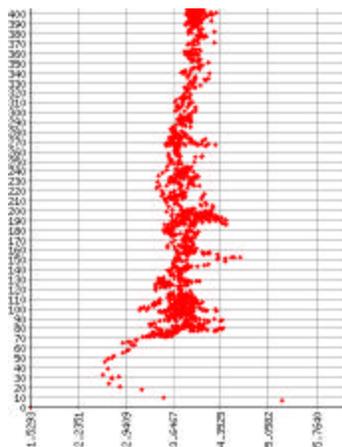
Incoming Packet Types



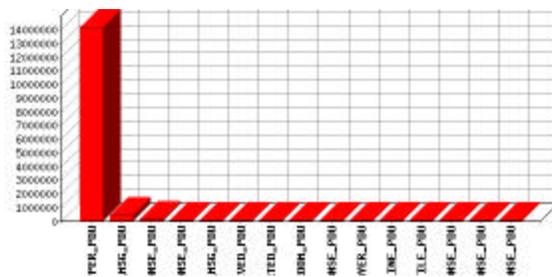
Simulator Overhead (Packet Type)



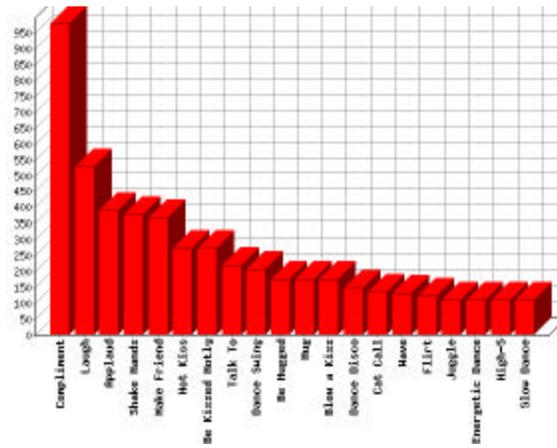
Players/Lot, by players/city



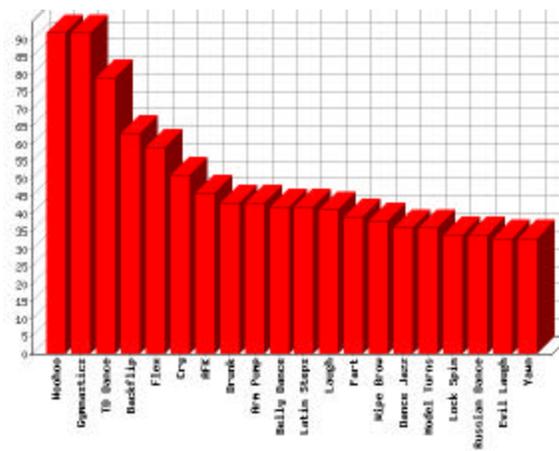
Outgoing PDUs (by Type)



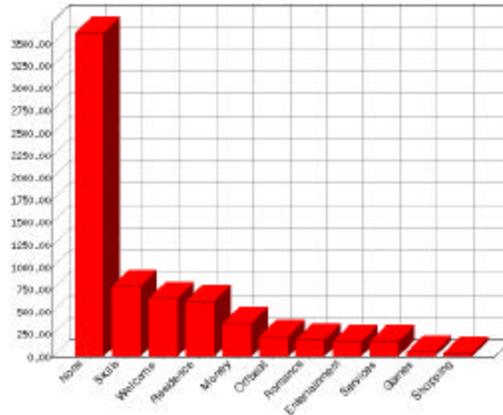
Object Interactions (AlphaVille)



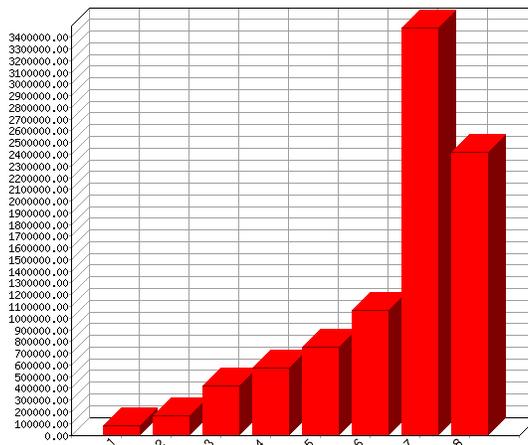
Puppeteering



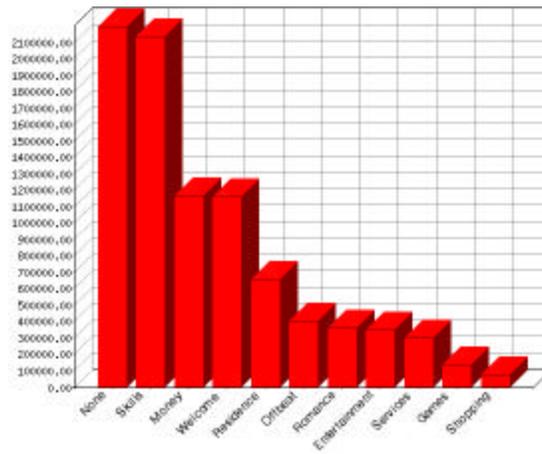
House Categories



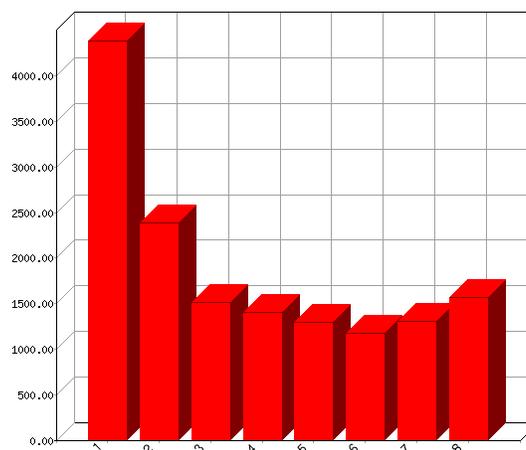
House Value (by Age)



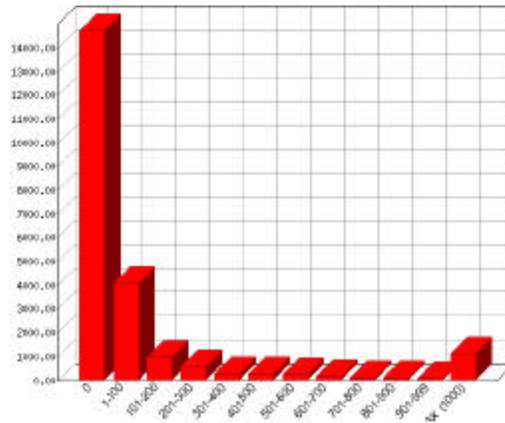
House Value (across city, by Cat)



numPlayers by numRoomMates



numPlayers getting a VisitorBonus



Calibration: Load Testing

- Using esper to measure userLoad @ peak in Live city
- Changing user_behaviour in load testing script (automated testing) to match liveLoad
- Using esper to measure emulatedLoad
 - Tune as required
 - Example: WAH.txt
- Used in turn to measure the infrastructure for completeness “is the infrastructure ready for launch?”
- Visual: monkey See/Do \leftrightarrow liveCity testCity

Scalability / Performance Analysis

- Intrinsic vs achieved
 - Player actions ultimately drive the load. Must understand input patterns to truly optimize system.
 - And, sometimes the best action is to change gameplay to increase intrinsic (e.g. dogfight: all in view? Crowds of people, portal storm, etc?)
- “Tall pole” analysis of packets per day
 - Tune system components accordingly
- What packets cause the heaviest server load?
 - Repeat tuning
- Example: Data Service
- Simulator (and other components): internal
- Per machine: CPU/disk/page_faults/...
 - Directly correlate user_action → packet → simulator_action → CPU hit (e.g. houseLoad higher than expected)

Game Analysis

- Game designers were heavy Esper users
 - Tuning
 - Economy
 - Game play

Economy Analysis

- Where did the money come from?
- Where did it go?
- How much did users play the money sub-game?
- Av amount of \$ made per player over 1st 10 days

Game Play Analysis

- Most popular Interactions / Objects / places
- Length of time in a house
- Chat rate
- Types of characters chosen
- ...
- Direct observation/change_tuning/observe cycle

Marketing

- Press releases
 - Tidbits to catch media / free pub
- Paid sponsorship
 - How many eyes on their brand, and for how long?
- 'Hot' objects / features

Community Management

- Observing user behaviour
- Shifting user's from city to city (generically, managing your users)
 - Calvin's Creek: tipping
- Cheap content: "Metrics that matter"

Customer Service

- Who's being a pain?
- Cheaters / griefers / ...

WrapUp

- Lessons Learned
 - What worked well
 - What didn't
 - What I'd do differently

Lessons Learned

- Don't wait to implement
- Keep light-weight enough to keep live
- Auto summarize
- Had to add some player-level tracking for CSR
 - New players would have been useful too (out of time)
- Ease of use
- Speed
 - Of turnaround on new metrics
 - Of drawing on user's screen
- Excellent compliment to automated testing
 - Repeatable inputs & accurate measurements allow experimentation @ scale
- Automate error checking on inputs
- Too many metrics collection system
 - Lack of a useful central system meant N people went and did one for their (narrowly targeted) needs
- Data Mining on players is very, very cool

Conclusions

- Very useful thing, do it
- Do it early for full benefit
- Make it easy to use

Questions