

IEPM-BW (or PingER on steroids) and the PPDG

Les Cottrell – SLAC

Presented at the PPDG meeting, Toronto, Feb 2002

www.slac.stanford.edu/grp/scs/net/talk/ppdg-feb02.html



Partially funded by DOE/MICS Field Work Proposal on Internet End-to-end Performance Monitoring (IEPM). Supported by IUPAP. PPDG collaborator.



Overview

1. Main issues being addressed by project
2. Other active measurement projects & deployment
3. Deliverables from IEPM-BW
4. Initial results
5. Experiences
6. Forecasting
7. Passive measurements
8. Next steps
9. Scenario

IEPM-BW: Main issues being addressed

- Provide a **simple, robust** infrastructure for:
 - **Continuous/persistent** and one-off measurement of high **network AND application** performance
 - management infrastructure – flexible remote host configuration
- **Optimize impact** of measurements
 - Duration, frequency of active measurements, and use passive
- Integrate standard set of measurements including: ping, traceroute, pipechar, iperf, bbcp ...
- Allow/encourage adding measurement/application tools
- Develop tools to gather, **reduce, analyze, and publicly report** on the measurements:
 - Web accessible data, tables, time series, scatterplots, histograms, forecasts ...
- Compare, evaluate, **validate** various measurement tools and strategies (minimize impact on others, effects of app self rate limiting, QoS, compression...), find better/simpler tools
- Provide simple **forecasting** tools to aid applications and to **adapt the active measurement frequency**
- Provide tool suite for high throughput monitoring and prediction

3

Other active measurement projects

	Surve yor	RIPE	AMP	PingER	NIMI	IEPM-BW
Commun ity	I2	Europe ISPs	NSF	HENP/ ESnet/ ICFA	Research	HENP/ PPDG/ Grid
Coverage	Mainly US	Mainly Europe	Mainly US	72 Countries		US, CA, JP, NL, CH, UK, FR
Metrics	One way delay, loss	One way delay, loss	RTT, loss	RTT, loss	RTT, loss, net thru, mcast	RTT, loss, net & app thru
Persistent	Yes*	Yes	Yes	Yes	On demand	Yes
Topo	Mesh	Mesh	Mesh	Hierarchy		Hierarchy
Data acc.	Request	Member	Public	Public	Member	Public

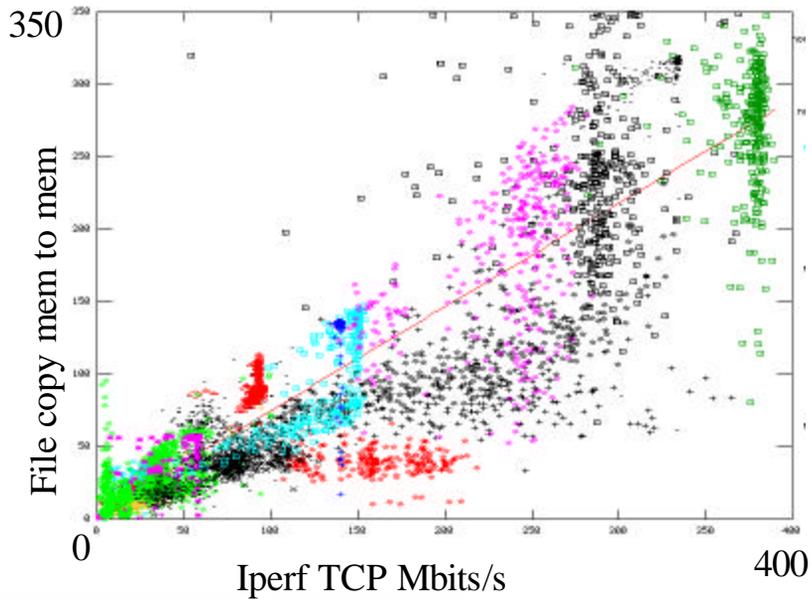
4

IEPM-BW Deployment in PPDG							
	PingER	AMP	Surveyor	RIPE	NIMI	Trace	IEPM-BW
SLAC	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LBNL	Yes				Yes	Yes	yes
UWisc	Yes	Yes	Yes			Yes	yes
FNAL	Yes	Yes	Yes		Yes	Yes	yes
ANL	Yes		Yes			Yes	yes
BNL	Yes		Yes				yes
JLAB	Yes					Yes	yes
Caltech	Yes						yes
SDSC	Yes	Yes					yes

+ { •CERN, IN2P3, INFN(Milan, Rome, Trieste), KEK, RIKEN, NIKHEF, DL, RAL, TRIUMF
•GSFC, LANL, NERSC, ORNL, Rice, Stanford, SOX (Atlanta), UDelaware, UFlorida, UMichigan, UT Dallas

IEPM-BW Deliverables
<ul style="list-style-type: none"> • Understand and identify resources needed to achieve high throughput performance for Grid and other data intensive applications • Provide access to archival and near real-time data and results for eyeballs and applications: <ul style="list-style-type: none"> – planning and expectation setting, see effects of upgrades – assist in trouble-shooting problems by identifying what is impacted, time and magnitude of changes and anomalies – as input for application steering (e.g. data grid bulk data transfer), changing configuration parameters – for forecasting and further analysis • Identify critical changes in performance, record and notify administrators and/or users • Provide a platform for evaluating new SciDAC & base program tools (e.g. pathrate, pathload, GridFTP, INCITE ...) • Provide measurement/analysis/reporting suite for Grid & hi-perf sites ⁶

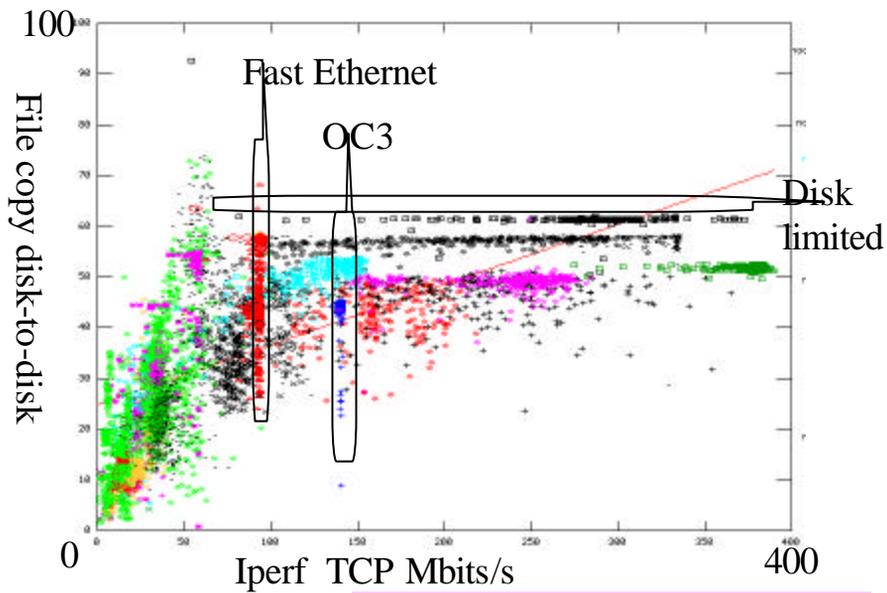
E.g. Iperf vs File copy (mem-to-mem)



File copy mem to mem ~ 72% iperf

9

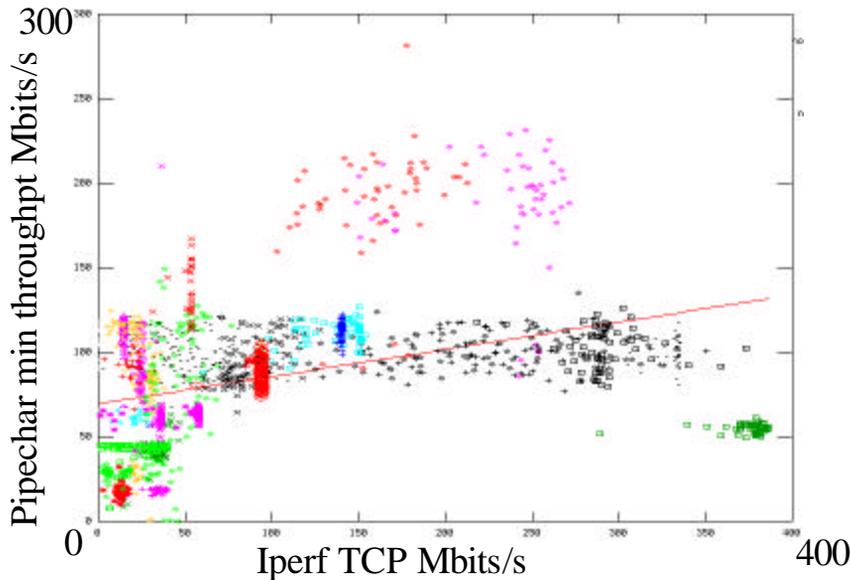
E.g. Iperf vs file copy disk to disk



Over 60Mbits/s iperf >> file copy

10

E.g. iperf vs pipechar



11

Forecasting

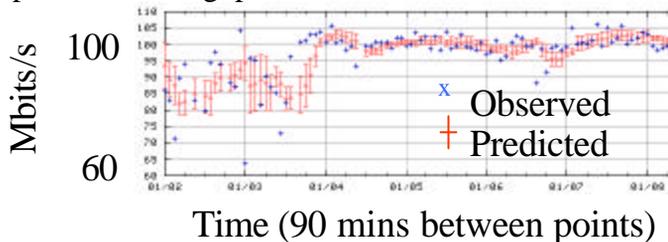
- Given access to the data one can do real-time forecasting for
 - TCP bandwidth, file transfer/copy throughput
 - E.g. NWS, *Predicting the Performance of Wide Area Data Transfers* by Vazhkudai, Schopf & Foster
- Developing simple prototype using average of previous measurements
 - Validate predictions versus observations
 - Get better estimates to adapt **frequency** of active measurements & reduce impact
 - Also look at ping RTTs and route information
 - Look at need for diurnal corrections
 - Use for steering applications
- Working with NWS for more sophisticated forecasting
- Can also use on demand bandwidth estimators (e.g. pipechar, but need to know range of applicability)

12

Forecast results

Predict = Moving average of last 5 measurements $\pm \sigma$

Iperf TCP throughput SLAC to Wisconsin, Jan '02



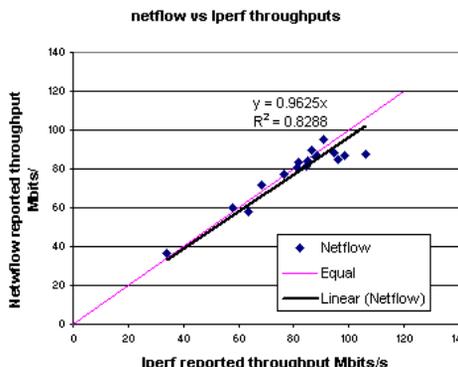
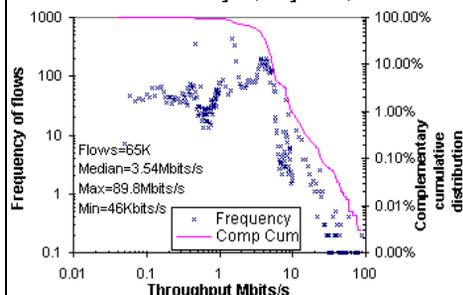
% average error = $\text{average}(\text{abs}(\text{observe} - \text{predict}) / \text{observe})$

33 nodes	Iperf TCP	Bbcp mem	Bbcp disk	bbftp	pipechar
Average % error	13% +- 11%	23% +- 18%	15% +- 13%	14% +- 12%	13% +- 8%

Passive (Netflow) data

- Use Netflow measurements from border router
 - Netflow records time, duration, bytes, packets etc./flow
 - Calculate throughput from Bytes/duration for big flows
 - Validate vs. iperf

SLAC border Netflow throughput distribution for TCP flows with > 10MBytes, July 12-26, 01



Scenario

- BaBar user wants to transfer large volume (e.g. TByte) of data from SLAC to IN2P3:
 - Select initial windows and streams from a table of pre-measured optimal values, or use an on demand tool (extended iperf), or reasonable default if none available
 - Application uses data volume to be transferred and simple forecast to estimate how much time is needed
 - Forecasts from active archive, Netflow, on demand use one-end bandwidth estimation tools (e.g. pipechar, NWS TCP throughput estimator)
 - If estimate duration is longer than some threshold, then more careful duration estimate is made using diurnal forecasting
 - Application reports to user who decides whether to proceed
 - Application turns on QBSS and starts transferring
- For long measurements, provide progress feedback, using progress so far, Netflow measurements of this flow for last few half hours, diurnal corrections etc.
 - If falling behind required duration, turn off QBSS, go to best effort
 - If throughput drops off below some threshold, check for other sites₁₇

More Information

- IEPM/PingER home site:
 - www-iepm.slac.stanford.edu/
- IEPM/BW site
 - www-iepm.slac.stanford.edu/bw
- SC2001 & high throughput measurements
 - www-iepm.slac.stanford.edu/monitoring/bulk/sc2001/
- QBSS measurements
 - www-iepm.slac.stanford.edu/monitoring/qbss/measure.html
- Netflow
 - <http://www.cisco.com/warp/public/732/Tech/netflow/>
 - www.slac.stanford.edu/comp/net/netflow/SLAC-Netflow.html