An Application of Central Limit Theorem to Wide Area Network Service Level Agreement Analyses

Sean Farney
Global Network Architect
seanfarney@ieee.org

Abstract[1] --
In this paper I analyze Round Trip Delay (RTD) Service Level Agreements (SLA) on a global MPLS Wide Area Network. Specifically, latency in milliseconds measured between a Data Center facility in Boston and 20 remote locations. The monthly mean of the actual delay is compared to the maximum delay allowed under the terms of the contract, the "latency SLA", and the Central Limit Theorem is applied to assess the probability of exceeding the SLA. The expectation is that the data will establish or challenge the efficacy of the provider guarantees. After performing the analysis, it is clear that the latency SLAs are ineffectual- the Theorem demonstrates that the probabilities of experiencing the conditions which would invoke SLA violation that are so low as to be nonexistent. We see an average distance from the expected mean equal to eight σ, falling within three σ in only five instances. Given such results, I suggest that a more efficient model for SLA construction and measurement is required. Finally, I question the need for managed bandwidth services if latency is perceived as a competitive advantage.

Introduction
Managed Network Service Providers (NSP) supply the bandwidth, transport, equipment, and management services to connect disparate locations across the corporate enterprise. To ensure levels of quality for traffic transiting these networks, Service Level Agreements for latency, availability, and packet loss are included in the contractual arrangements between provider and customer. If these service levels are not met, the customer is entitled to financial remuneration. Usually, there is reimbursement of a certain pre-determined percentage of the monthly cost on a per site basis. Obviously, providers are motivated to avoid violations in order to retain as much operating profit as possible.

The carrier monitors SLA performance via network management systems (NMS). In our example, the NSP uses a proprietary NMS to send Simple Network Management Protocol (SNMP) polls to remote routers, and records the roundtrip latency in milliseconds. This happens every 15 minutes, daily, and is tabulated at the end of every month, resulting in sample size n=2,976. In our example, the Boston Data Center facility hosts shared data and applications for the entire enterprise. SLA measurements are performed between this location and 20 sites spanning 11countries and 4 continents.

One of the challenges of operating a geographically distributed infrastructure is making sense of the abundance of collected data. The network manager is faced with important questions: Are SLAs effective at serving business needs and the infrastructure portfolio? Do they elevate network performance or restrain it? Are SLAs part of the vendor value equation? How are they constructed before finalizing a contract and enforced afterwards? In today's technology driven economy, IT can create a clear competitive edge. Properly answering these questions has evolved from a titular back office consideration into a critical component of corporate data communication strategy. So there is a need for more exacting means of assessing network performance guarantees. For this we turn to statistical analyses.

Analysis and Findings

[1] Paper inspired by material from MITP 431 *Probability and Communications Networks*, taught by Dr. Abraham Haddad, Professor of Electrical and Computer Engineering, McCormick School, Northwestern University

To perform the analysis, I extracted January's SNMP information from an internal NMS located in the Boston Data Center[2] and formatted it into Microsoft Excel. I then derived the mean, standard deviation, variance, IQR, and other relevant information[3].

Before looking at the results we can make some predictions. The Law of Large Numbers tells us that as *n* becomes larger, the variance required to influence the mean becomes very large. With our large sample, we know the average will be driven to the mean and will be resistant to all but a very large number of outliers[4]. These two ideas can be summarized as:

$$\mu_Y = E\{Y\} = E\{\frac{1}{n}\sum_{i=1}^{n} X_i\} = \frac{1}{n}E\{\sum_{i=1}^{n} X_i\} = \frac{1}{n}\sum_{i=1}^{n} E\{X_i\} = \frac{n\mu}{n} = \mu$$

$$\sigma_Y^2 = \mathrm{Var}\{\frac{1}{n}\sum_{i=1}^{n} X_i\} = \frac{1}{n^2}\mathrm{Var}\{\sum_{i=1}^{n} X_i\} = \frac{1}{n^2}\sum_{i=1}^{n} Var\{X_i\} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$ [5]

To illustrate this, let's look at results from the Dallas office. The true mean, or expected value, of the polling samples from Boston to Dallas for January was 60.58ms. The standard deviation is only 4.69ms and variance is 22.00. It's a relatively tight group, which is not surprising considering the high *n*. The vendor guarantees that sample mean, the latency SLA, will be <145ms, or more than twice the expected mean! More precisely, the difference is 145-60.58=84.42 ms. So for μ to reach 145ms 84.42ms must be added to every sample. Or, 373.85ms must be added to all samples during a 7 day period (lower *n*, higher variance):

*Boston Data Center router to **Dallas** router, latency, milliseconds*	
Sample Mean, μ, E(X)	60.58
Standard Deviation, σ	4.69
Variance, σ2	22.00
Max	89.63
SLA Mean, (X)	145.00
#σ's SLA is from μ	18.00
To exceed SLA, # of ms added to every sample for one week	**373.85**
To exceed SLA, # of ms added to every sample	**84.42**

[6]

From the network user perspective, this could have devastating implications. Latency would have to skyrocket to 2,616ms for an 8 hour period (*n=96*), a level which would render the network completely unusable, in order to invoke SLA violation. Knowing the behavior the network would have to exhibit to reach the SLA levels, we must determine the chance of this event occurring. For this we apply the Central Limit Theorem.

The Central Limit Theorem teaches us that if sampling from a population with an unknown probability distribution, the sampling distribution of the sample mean will approximate a normal if *n* is large.[7] By

---

[2] Solarwinds Orion Network Performance Monitoring NMS, sample period 01 January 00:01- 31 January 23:59
[3] See Appendix I, Site-to-site Summary. Source Excel data available electronically
[4] The mean is specifically considered not resistant to outliers, but in this the large *n* makes them irrelevant.
[5] Dr. Abraham Haddad, Professor of Electrical and Computer Engineering, McCormick School, Northwestern University MITP 431 *Probability and Communications Networks*, lecture notes pg. 115
[6] See Appendix II, SLA Gap Analysis for complete results
[7] Pg. 239, Montgomery D., and Runger G., (2006) *Applied Statistics and Probability for Engineers,* 3rd Edition, John Wiley & Sons, Inc.

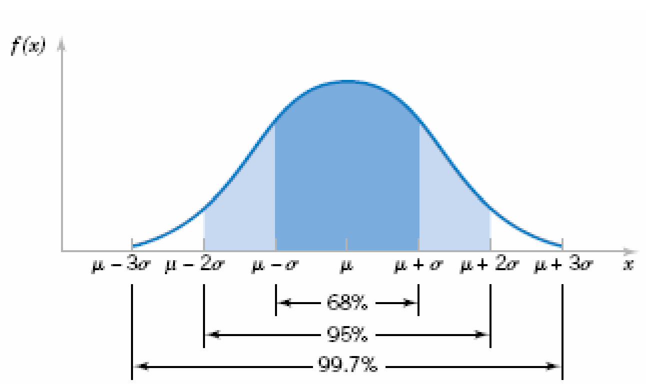
knowing the expected value through testing, and the SLA sample mean, we can compute the exact probability of this occurring. I used a modified Central Limit Theorem which uses distance, σ, away from mean as the key variable:

$$\alpha = \#\sigma$$

$$P(Z \geq \alpha) \cong \frac{1}{\sqrt{2\Pi a}} e^{-\frac{a^2}{2}}$$

[8]

The importance of standard deviation vis-à-vis probability is best illustrated by the Standard Normal Distribution:



[9]

As depicted above, the density of the probability function is greatest about the mean, μ. Logically then, as you move farther away from μ, the density diminishes, eventually reaching an approximated zero. The probabilities within the distribution are bounded by the number of standard deviations, σ, from the mean. Since inclusion within 3σ is .997, the probability of attaining a value outside of 3σ is very low.

I computed the σ value between expected mean and SLA mean and applied the Central Limit Theorem to arrive at the probabilities. The results were shocking.[10] For only three sites is SLA impingement possible. For the others it is not even probable. It is evident that the service level agreements are generally worthless on this network, guarding against latency levels that will never occur in 17 of 20 sites. It's is easy to see why. Once the true mean becomes approximately 10 σ from the SLA mean, the probability becomes essentially zero[11] The sites that have achievable probabilities for SLA violation are Jakarta, 7%, Mumbai, 12%, Shanghai, 58%. For only these locations can we consider the possibility that the RTD levels may be appropriate.

Conclusion

The analysis suggests that the only thing these service level agreements guarantee is that the service provider will never have to pay for SLA encroachment. This is because the provider's sampling methodology, 2,976 per month, wields The Law of Large Numbers offensively to insure no revenues returned to the customer. It seems the key to accurate and valid SLA design and enforcement is the lowering of *n*. Since there is an inverse relationship between *n* and variance, if *n* could be lowered from

---

[8] Dr. Abraham Haddad, Professor of Electrical and Computer Engineering, McCormick School

[9] Pg. 111. Montgomery D., and Runger G., (2006) *Applied Statistics and Probability for Engineers,* 3rd Edition, John Wiley & Sons, Inc.

[10] See Appendix I, Site-to-site Summary. Source Excel data available electronically

[11] The probability exceeds 10^{-31}, a restraint in Microsoft Excel, which for my purposes can safely be assumed to be zero

2,976 to 96, (30 days to 1 day) for example, the effect of variance would be dramatic[12], and a much more realistic picture of the network experience. Central Limit Theorem can be employed to insure the gulf between the real mean and the SLA mean never exceeds 1.5-2σ. Experience based models could be created to generate acceptable SLA numbers. This would involve defining worst-case scenarios for specific network services. For example, if an application or location cannot tolerate latency >300ms for more than two hours on any given day, a model can be built with these parameters to return a maximal μ, and construct a tight SLA. Transmission delays are a given on computer networks and there is some overhead which can never be eliminated[13]. But having an intimate knowledge of how performance parameters are engineered empowers the consumer of network services and provides tremendous leverage in both negotiations and operations.

A residual effect of the determination that WAN service level agreements may be ineffective is that it could offer an opportunity to pursue a new data communications strategy. Corporate customers pay a premium for managed private network services because of the perceived benefit of service level agreements. However, if these SLAs actually provide no value, it would be the Network Manager's duty to investigate public network services that cost considerably less. In fact, on the same network we studied, in a comparison of both latency and monthly $/Mb between managed (private MPLS carrier) and unmanaged (public ISP) bandwidth, the public bandwidth comes out quite favorably.[14] Wide Area Networks that carry voice traffic and have strict jitter and packet loss requirements could never pursue a public option for transport, but for those that do not, the cost savings of an alternate communications scheme could be substantial.

Insights

The ubiquity of the wide area connectivity in today's business environment raises network performance to a fundamental level of importance. The simultaneous distribution of global infrastructure resources and consolidation of technical architectures, are pushing perimeters outward at an alarming rate. Combined, these factors mandate predictable, reliable, and high-performing super-distributed enterprises. So knowing exactly how the network performs is seminal to the corporate IT mission. The mathematics involved in these analyses is amateur. But the implications are grand. By applying fundamental statistical analyses, our level of understanding, and subsequently control, of infrastructure systems rises dramatically. Whether it's Poisson arrivals into a switch, Central Limit Theory to predict SLA adherence, or a simple Normal Distribution, probability is an empowering tool in the IT professional's repertoire, and will fast become a necessary skill for advancing technology.

---

[12] Since variance is a function of the square root

[13] $T = T_q + T_s + T_p$, Total transmission time on a data network = queueing delay + serialization delay + propagation delay

[14] See Appendix III, Public vs. Private Bandwidth Comparison

Appendix I: Site-to-site summary

*Boston Data Center router to **Atlanta** router, latency, milliseconds*	
Sample Mean, μ, E(X)	36.29
Standard Deviation, σ	4.65
Variance, σ2	21.59
Median	34.71
Average Deviation	3.40
Max	82.5
1st Quartile	33.29
3rd Quartile	38.43
IQR	5.14
SLA Mean, (X)	83
#σ's SLA is from μ	10.05
P{E(X)>(X)} **Probability of exceeding SLA**	0.000000000000000000000004470612

*Boston Data Center router to **Boston** router, latency, milliseconds*	
Sample Mean, μ, E(X)	10.94
Standard Deviation, σ	2.61
Variance, σ2	6.80
Median	10.25
Average Deviation	1.89
Max	24.63
1st Quartile	9.13
3rd Quartile	11.86
IQR	2.73
SLA Mean, (X)	83
#σ's SLA is from μ	27.63
P{E(X)>(X)} **Probability of exceeding SLA**	0.000000000000000000000000000000

*Boston Data Center router to **Chicago** router, latency, milliseconds*	
Sample Mean, μ, E(X)	37.59
Standard Deviation, σ	4.09
Variance, σ2	16.70
Median	36.00
Average Deviation	3.21
Max	62.5
1st Quartile	34.63
3rd Quartile	39.57
IQR	4.94
SLA Mean, (X)	92
#σ's SLA is from μ	13.31
P{E(X)>(X)} **Probability of exceeding SLA**	0.000000000000000000000000000000

*Boston Data Center router to **Dallas** router, latency, milliseconds*	
Sample Mean, μ, E(X)	60.58
Standard Deviation, σ	4.69
Variance, σ2	22.00
Median	58.88
Average Deviation	3.60
Max	89.63
1st Quartile	57.5
3rd Quartile	62.86
IQR	5.36
SLA Mean, (X)	145
#σ's SLA is from μ	18.00
P{E(X)>(X)} **Probability of exceeding SLA**	0.000000000000000000000000000000

*Boston Data Center router to **Hong Kong** router, latency, milliseconds*	
Sample Mean, μ, E(X)	240.66
Standard Deviation, σ	12.39
Variance, σ2	153.57
Median	240.43
Average Deviation	9.16
Max	331.43
1st Quartile	232.5
3rd Quartile	247.29
IQR	14.79
SLA Mean, (X)	322
#σ's SLA is from μ	6.56
P{E(X)>(X)} **Probability of exceeding SLA**	0.000000000026886901713419900000

*Boston Data Center router to **Jakarta** router, latency, milliseconds*	
Sample Mean, μ, E(X)	300.64
Standard Deviation, σ	46.85
Variance, σ2	2195.37
Median	290.13
Average Deviation	20.11
Max	1194.75
1st Quartile	285
3rd Quartile	298.43
IQR	13.43
SLA Mean, (X)	373
#σ's SLA is from μ	1.54
P{E(X)>(X)} **Probability of exceeding SLA**	0.078408911511951500000000000000

*Boston Data Center router to **Kuala Lumpur** router, latency, milliseconds*	
Sample Mean, μ, E(X)	300.11
Standard Deviation, σ	15.60
Variance, σ2	243.35
Median	298.43
Average Deviation	9.10
Max	487.63
1st Quartile	294
3rd Quartile	304
IQR	10
SLA Mean, (X)	374
#σ's SLA is from μ	4.74
P{E(X)>(X)} **Probability of exceeding SLA**	0.000001130298132273690000000000

*Boston Data Center router to **Los Angeles** router, latency, milliseconds*	
Sample Mean, μ, E(X)	86.63
Standard Deviation, σ	4.55
Variance, σ2	20.71
Median	85.38
Average Deviation	3.34
Max	117.6
1st Quartile	83.88
3rd Quartile	88.57
IQR	4.69
SLA Mean, (X)	150
#σ's SLA is from μ	13.93
P{E(X)>(X)} **Probability of exceeding SLA**	0.000000000000000000000000000000

*Boston Data Center router to **Melbourne** router, latency, milliseconds*	
Sample Mean, μ, E(X)	271.00
Standard Deviation, σ	22.55
Variance, σ2	508.56
Median	263.29
Average Deviation	15.32
Max	391
1st Quartile	258.57
3rd Quartile	271.88
IQR	13.31
SLA Mean, (X)	360
#σ's SLA is from μ	3.95
P{E(X)>(X)} Probability of exceeding SLA	0.000041966263329571400000000000

*Boston Data Center router to **Mumbai** router, latency, milliseconds*	
Sample Mean, μ, E(X)	245.43
Standard Deviation, σ	60.00
Variance, σ2	3599.88
Median	237.00
Average Deviation	20.00
Max	2392.57
1st Quartile	228.71
3rd Quartile	246.5175
IQR	17.8075
SLA Mean, (X)	324
#σ's SLA is from μ	1.31
P{E(X)>(X)} Probability of exceeding SLA	0.129268954960728000000000000000

*Boston Data Center router to **Munich** router, latency, milliseconds*	
Sample Mean, μ, E(X)	108.82
Standard Deviation, σ	8.24
Variance, σ2	67.94
Median	106.43
Average Deviation	5.54
Max	199.71
1st Quartile	104.14
3rd Quartile	111.43
IQR	7.29
SLA Mean, (X)	184
#σ's SLA is from μ	9.12
P{E(X)>(X)} Probability of exceeding SLA	0.000000000000000000037508509647

*Boston Data Center router to **New York** router, latency, milliseconds*	
Sample Mean, μ, E(X)	17.05
Standard Deviation, σ	3.54
Variance, σ2	12.55
Median	15.86
Average Deviation	2.62
Max	36.71
1st Quartile	14.57
3rd Quartile	18.13
IQR	3.56
SLA Mean, (X)	61
#σ's SLA is from μ	12.41
P{E(X)>(X)} Probability of exceeding SLA	0.000000000000000000000000000000

*Boston Data Center router to **Singapore** router, latency, milliseconds*	
Sample Mean, μ, E(X)	274.85
Standard Deviation, σ	12.24
Variance, σ2	149.91
Median	273.57
Average Deviation	9.79
Max	347.29
1st Quartile	266.25
3rd Quartile	283.0325
IQR	16.7825
SLA Mean, (X)	359
#σ's SLA is from μ	6.87
P{E(X)>(X)} Probability of exceeding SLA	0.000000000003215538669874840000

*Boston Data Center router to **Seoul** router, latency, milliseconds*	
Sample Mean, μ, E(X)	229.34
Standard Deviation, σ	23.74
Variance, σ2	563.59
Median	221.71
Average Deviation	15.63
Max	508.38
1st Quartile	215.29
3rd Quartile	234.865
IQR	19.575
SLA Mean, (X)	304
#σ's SLA is from μ	3.14
P{E(X)>(X)} Probability of exceeding SLA	0.000903312023690128000000000000

*Boston Data Center router to **San Francisco** router, latency, milliseconds*	
Sample Mean, μ, E(X)	95.36
Standard Deviation, σ	9.60
Variance, σ2	92.11
Median	93.43
Average Deviation	4.31
Max	320.5
1st Quartile	91.71
3rd Quartile	97.57
IQR	5.86
SLA Mean, (X)	164
#σ's SLA is from μ	7.15
P{E(X)>(X)} Probability of exceeding SLA	0.000000000000434516432626647000

*Boston Data Center router to **Shanghai** router, latency, milliseconds*	
Sample Mean, μ, E(X)	310.58
Standard Deviation, σ	97.80
Variance, σ2	9565.33
Median	278.13
Average Deviation	52.11
Max	2611
1st Quartile	269.63
3rd Quartile	307.8125
IQR	38.1825
SLA Mean, (X)	367
#σ's SLA is from μ	0.58
P{E(X)>(X)} Probability of exceeding SLA	0.585501976754279000000000000000

Boston Data Center router to ***Sydney*** *router, latency, milliseconds*	
Sample Mean, μ, E(X)	263.12
Standard Deviation, σ	23.91
Variance, σ2	571.68
Median	254.14
Average Deviation	16.66
Max	393.71
1st Quartile	249.75
3rd Quartile	262.57
IQR	12.82
SLA Mean, (X)	344
#σ's SLA is from μ	3.38
P{E(X)>(X)} **Probability of exceeding SLA**	0.000386346056224959000000000000

Boston Data Center router to ***Tokyo*** *router, latency, milliseconds*	
Sample Mean, μ, E(X)	199.99
Standard Deviation, σ	5.60
Variance, σ2	31.38
Median	198.86
Average Deviation	4.25
Max	239
1st Quartile	196.615
3rd Quartile	203.25
IQR	6.635
SLA Mean, (X)	260
#σ's SLA is from μ	10.71
P{E(X)>(X)} **Probability of exceeding SLA**	0.000000000000000000000000004475

Boston Data Center router to ***Toronto*** *router, latency, milliseconds*	
Sample Mean, μ, E(X)	24.33
Standard Deviation, σ	27.23
Variance, σ2	741.22
Median	34.86
Average Deviation	23.18
Max	414.57
1st Quartile	-2
3rd Quartile	40.75
IQR	42.75
SLA Mean, (X)	95
#σ's SLA is from μ	2.60
P{E(X)>(X)} **Probability of exceeding SLA**	0.005290519156430780000000000000

Boston Data Center router to ***Washington*** *router, latency, milliseconds*	
Sample Mean, μ, E(X)	20.40
Standard Deviation, σ	7.35
Variance, σ2	54.08
Median	20.43
Average Deviation	4.25
Max	43.86
1st Quartile	19.14
3rd Quartile	23.38
IQR	4.24
SLA Mean, (X)	68
#σ's SLA is from μ	6.47
P{E(X)>(X)} **Probability of exceeding SLA**	0.000000000049285760920406800000

Appendix II: SLA gap analysis

Atlanta	
To exceed SLA, # of ms added to every sample for one week	207
To exceed SLA, # of ms added to every sample	47

Boston	
To exceed SLA, # of ms added to every sample for one week	319
To exceed SLA, # of ms added to every sample	72

Chicago	
To exceed SLA, # of ms added to every sample for one week	241
To exceed SLA, # of ms added to every sample	54

Dallas	
To exceed SLA, # of ms added to every sample for one week	374
To exceed SLA, # of ms added to every sample	84

Hong Kong	
To exceed SLA, # of ms added to every sample for one week	360
To exceed SLA, # of ms added to every sample	81

Jakarta	
To exceed SLA, # of ms added to every sample for one week	320
To exceed SLA, # of ms added to every sample	72

Kuala Lumpur	
To exceed SLA, # of ms added to every sample for one week	327
To exceed SLA, # of ms added to every sample	74

Los Angeles	
To exceed SLA, # of ms added to every sample for one week	281
To exceed SLA, # of ms added to every sample	63

Melbourne	
To exceed SLA, # of ms added to every sample for one week	394
To exceed SLA, # of ms added to every sample	89

Mumbai	
To exceed SLA, # of ms added to every sample for one week	348
To exceed SLA, # of ms added to every sample	79

Munich	
To exceed SLA, # of ms added to every sample for one week	333
To exceed SLA, # of ms added to every sample	75

New York	
To exceed SLA, # of ms added to every sample for one week	195
To exceed SLA, # of ms added to every sample	44

Singapore	
To exceed SLA, # of ms added to every sample for one week	373
To exceed SLA, # of ms added to every sample	84

Seoul	
To exceed SLA, # of ms added to every sample for one week	331
To exceed SLA, # of ms added to every sample	75

San Francisco	
To exceed SLA, # of ms added to every sample for one week	304
To exceed SLA, # of ms added to every sample	69

Shanghai	
To exceed SLA, # of ms added to every sample for one week	250
To exceed SLA, # of ms added to every sample	56

Sydney	
To exceed SLA, # of ms added to every sample for one week	358
To exceed SLA, # of ms added to every sample	81

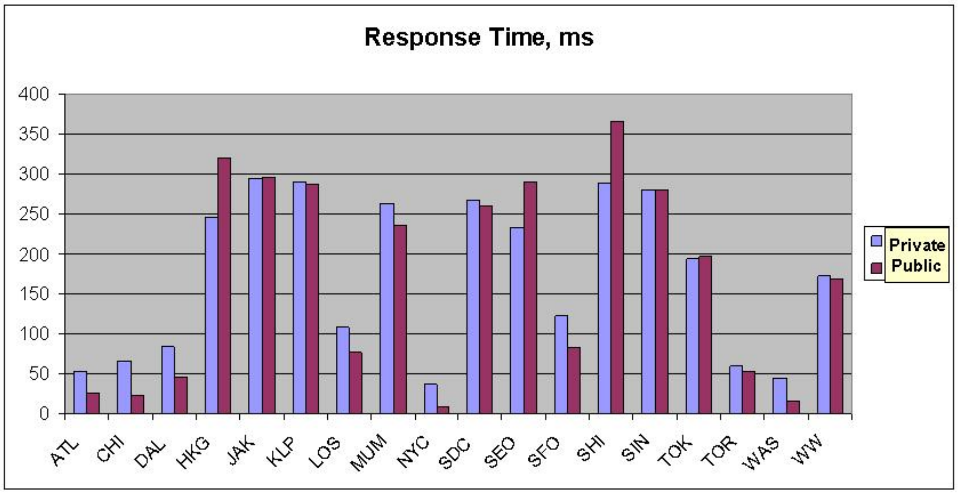
Tokyo	
To exceed SLA, # of ms added to every sample for one week	266
To exceed SLA, # of ms added to every sample	60

Toronto	
To exceed SLA, # of ms added to every sample for one week	313
To exceed SLA, # of ms added to every sample	71

Washington DC	
To exceed SLA, # of ms added to every sample for one week	211
To exceed SLA, # of ms added to every sample	48

Appendix III: Public vs. Private Bandwidth comparison

LATENCY: ISP's HAVE SLIGHT ADVANTAGE



WW average: Private Bandwidth 173ms, Public Bandwidth 169ms

[redacted]