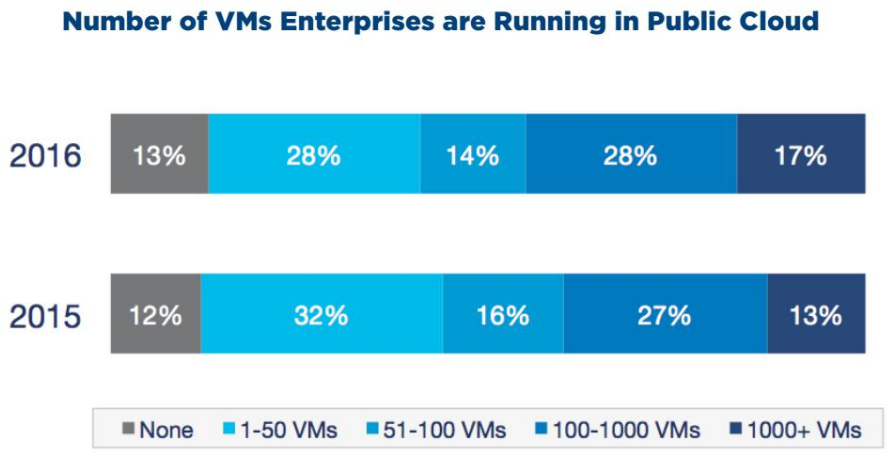
第二部分

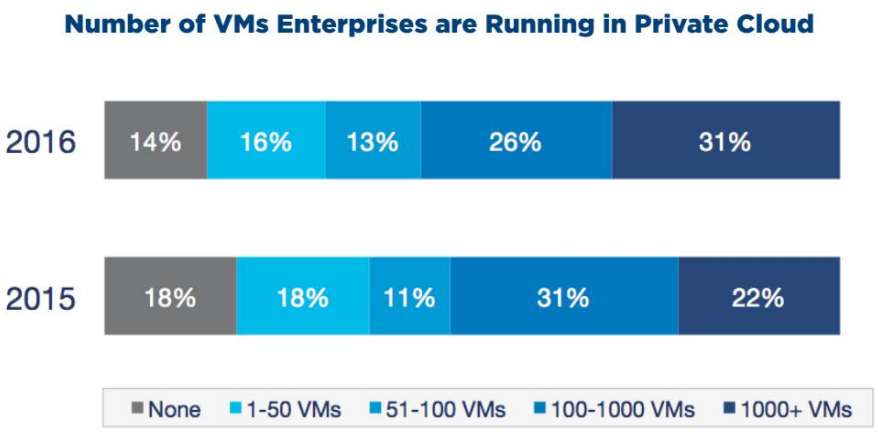
存储服务质量精确保障

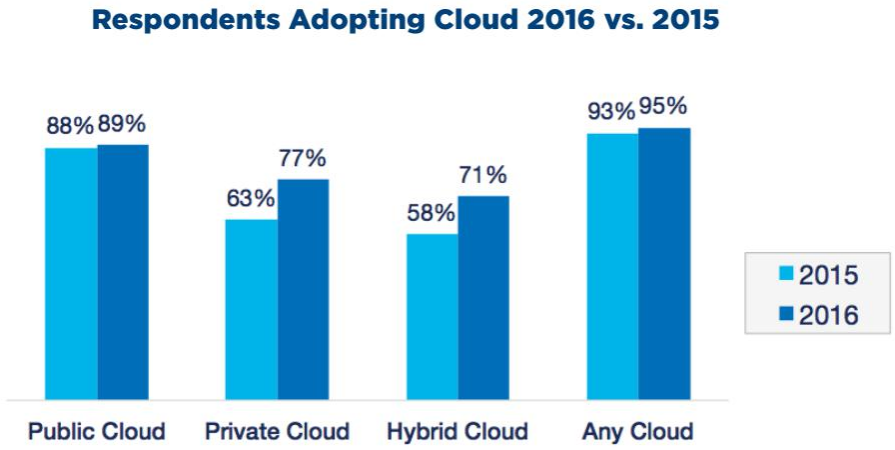
1. Customizable SLO and Its Near-Precise Enforcement for Storage Bandwidth[J]. ACM Transactions on Storage (TOS), 2017.
2. SASLO: Support User-Customized SLO Policy via Programmable End-to-End VM-Oriented IO Control[C]. International Conference on Cloud Computing and Big Data (CCBD), 2015.

背景



- 各类云服务应用比例逐年提升
- 云中租赁大规模VM集群的企业级用户的比例增长较快





2016年 RightScale “State of the Cloud Report”

概念

SLO, SLA & SLI Terminology

- **SLO - Service level objective** is agreed as a means of measuring the performance of the Service Provider.
- **SLA - Service Level Agreement** specifies what service is to be provided, how it is supported, times, locations, costs, performance, and responsibilities of the parties involved. SLOs are specific measurable characteristics of the SLA such as availability, throughput, frequency, response time, or quality.
- **SLI - Service Level Indicator** is a measure of the service level provided by a service provider to a customer. SLIs form the basis of Service Level Objectives (SLOs), which in turn form the basis of Service Level Agreements (SLAs).

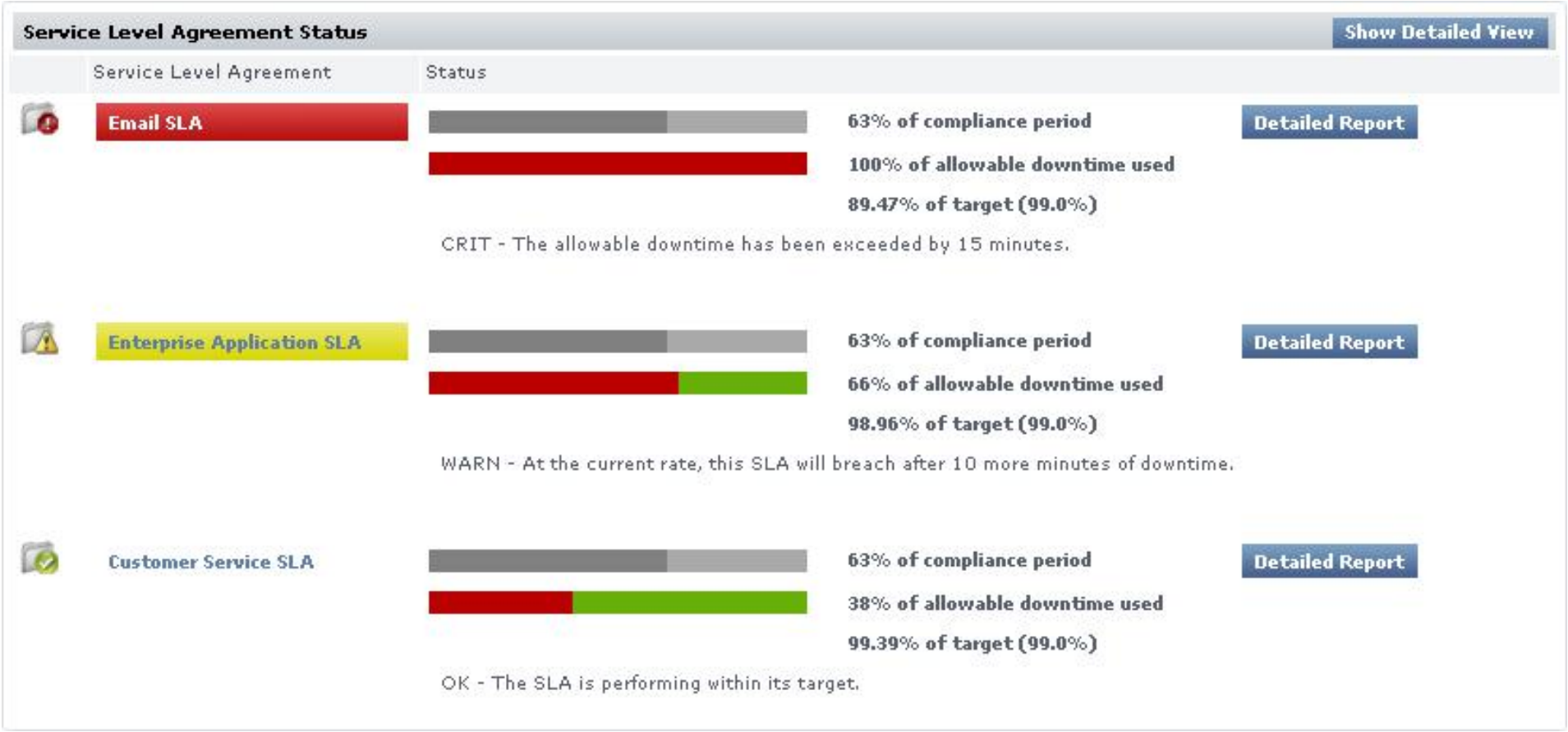
性能评价标准

性能指标	评价标准	性能评价标准	数学定义
I/O延时(L)	均值误差(E)	I/O延时均值误差(E_L)	$E_L = (L_{ave} - L_{SLO})/L_{SLO} * 100\%)$
	服从率($SCrate$)	I/O延时SLO服从率($SCrate_L$)	$SCrate_L = F(\delta_1 \leq L \leq \delta_2) * 100\%$
	百分位性能(PC)	I/O延时的百分位性能(PC_L)	$PC_L = F^{-1}(X\%), 0 \leq X \leq 100$
I/O吞吐率(T)	均值误差(E)	I/O吞吐率均值误差(E_T)	$E_T = (T_{ave} - T_{SLO})/T_{SLO} * 100\%)$
	服从率($SCrate$)	I/O吞吐率SLO服从率($SCrate_T$)	$SCrate_T = F(\delta_1 \leq T \leq \delta_2) * 100\%$
	性能波动($PFratio$)	I/O吞吐率性能波动($PFratio_T$)	$PFratio_T = \frac{SD(T_{actual}(k))}{T_{SLO}} * 100\%$
I/O带宽(B)	均值误差(E)	I/O带宽均值误差(E_B)	$E_B = (B_{ave} - B_{SLO})/B_{SLO} * 100\%)$
	绝对性能误差(AE)	I/O带宽绝对性能误差(AE_B)	$AE_B = (\sum_{i=1}^{T} \lvert G_B^{(t_i)} \rvert)/T * 100\%)$

举例

Example Service Level Objectives	
General Comparisons	**Service levels** (*X* percent answer/*Y* seconds)
Emergency services	100/0
Service level objectives "high"	90/20, 85/15, 90/15
Service level objectives "moderate"	80/20, 80/30, 90/60
Service level objectives "modest"	70/60, 80/120, 80/300

应用场景


Service Level Agreement Status
Show Detailed View
Service Level Agreement
Status
Email SLA
63% of compliance period
100% of allowable downtime used
89.47% of target (99.0%)
Detailed Report
CRIT - The allowable downtime has been exceeded by 15 minutes.
Enterprise Application SLA
63% of compliance period
66% of allowable downtime used
98.96% of target (99.0%)
Detailed Report
WARN - At the current rate, this SLA will breach after 10 more minutes of downtime.
Customer Service SLA
63% of compliance period
38% of allowable downtime used
99.39% of target (99.0%)
Detailed Report
OK - The SLA is performing within its target.

研究现状

- 目前的I/O吞吐率/带宽SLO保障和优化方案没有充分考虑I/O请求队列的突发性波动对I/O延时性能的影响，导致I/O吞吐率/带宽和I/O延时指标之间的SLO保障行为和性能优化未实现有效隔离。
- 现有的I/O延时SLO保障方案主要基于冗余资源或最坏情况下的资源供给模式，难以支持精确I/O延时，尤其是高百分位尾延时SLO的执行。

吞吐率/带宽SLO精确保障方法

SASLO系统的设计 Stable and Accurate SLO

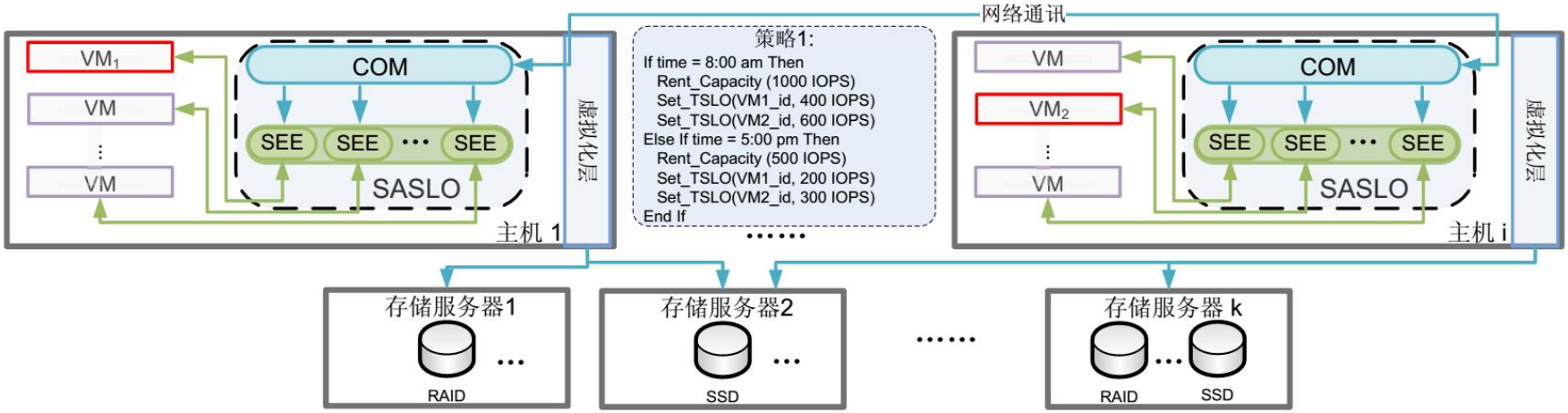


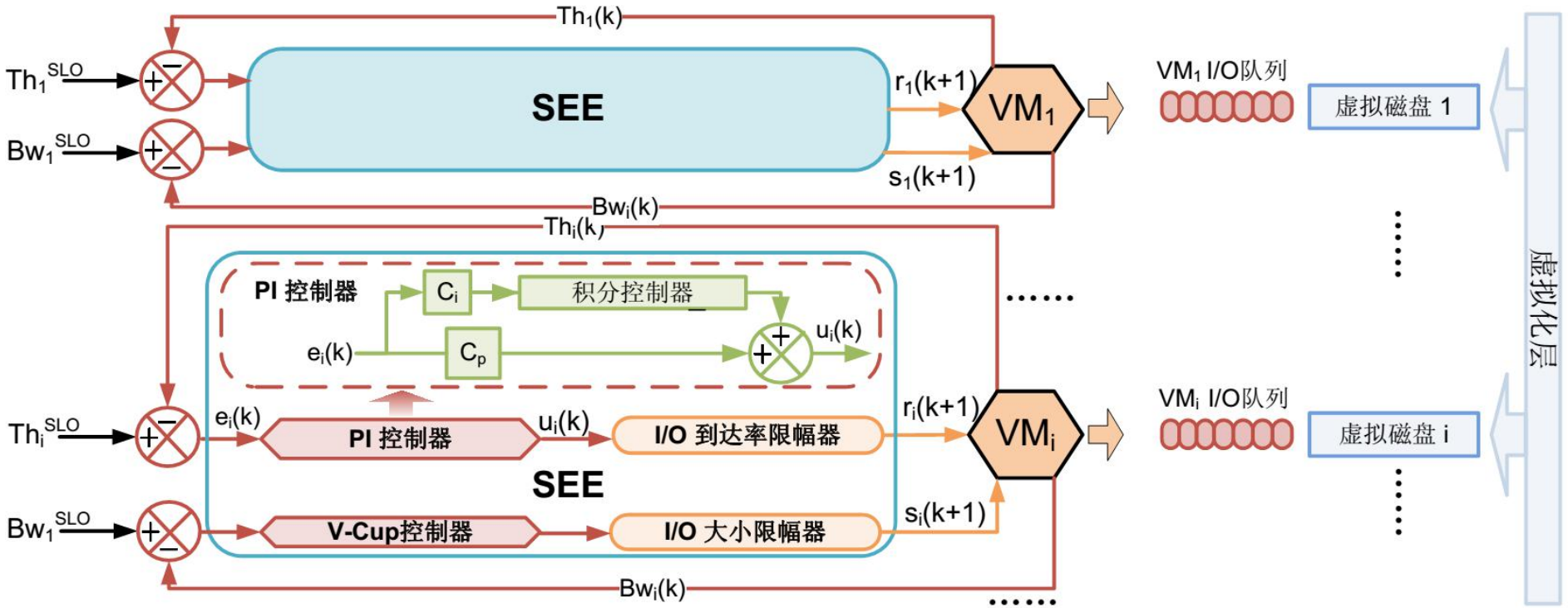
表 2.1 面向用户订制SLO的API

#	API
1	**Set_TSLO (VM_id i, Throughput_slo t)** 将ID为*i*的虚拟机的I/O吞吐率SLO目标值设定为*t*
2	**Set_BSLO (VM_id i, Bandwidth_slo t)** 将ID为*i*的虚拟机的I/O带宽SLO目标值设定为*t*

SASLO主要包括通讯层(COM) 和SLO执行引擎 (SEE)

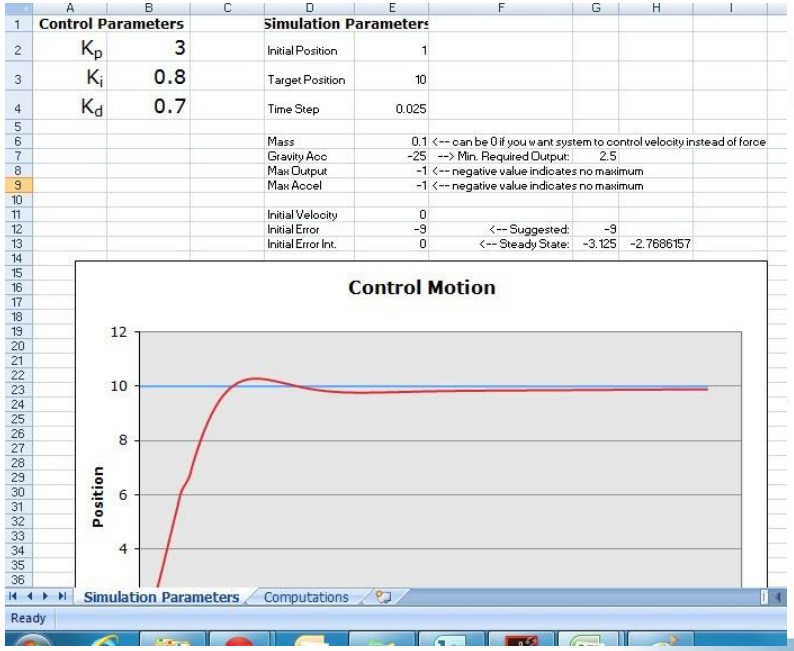
吞吐率/带宽SLO精确保障方法

- **比例积分(PI) 控制器**: 将实际虚拟机的I/O吞吐率收敛于SLO目标值。
- **V-Cup控制器**: 通过I/O请求大小波动控制进行I/O带宽SLO精确保障。



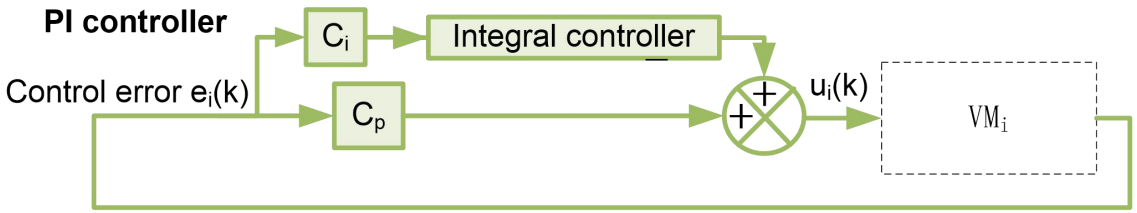
吞吐率/带宽SLO精确保障方法

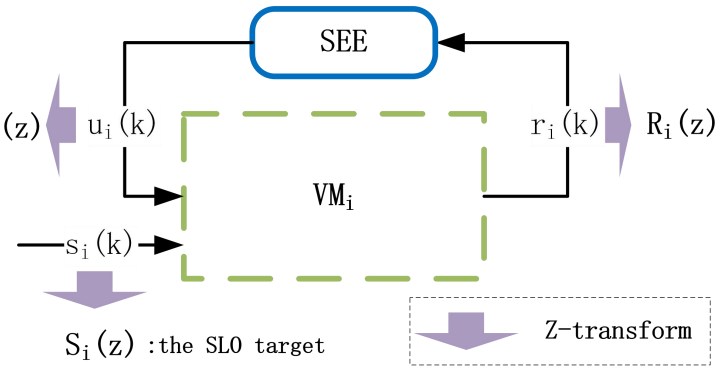
- 关于PI控制器
  - https://automationforum.in/t/pid-simulator-free-tools-collection/557



吞吐率/带宽SLO精确保障方法

比例积分(PI)控制器





$$u_i^p(k) = C_P \times e_i(k)$$

$$u_i^I(k) = u_i^I(k-1) + C_I \times e_i(k)$$

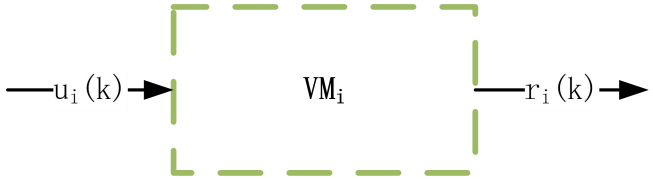
$$u_i(k) = u_i^p(k) + u_i^I(k)$$

比例积分控制模型

闭环传递函数

采用自回归移动平均 (ARMA)模型描述VM输入ui(k)对VM输出ri(k)的影响。

$$F_i(z) = \frac{R_i(z)}{S_i(z)} = \frac{[(C_p + C_I) \times z - C_p] \times G_i(z)}{(z-1) + [(C_p + C_I) \times z - C_p] \times G_i(z)}$$



传递函数

根据控制理论，当闭环系统稳定的时候，PI控制具有0稳态误差

$$r_i(k+1) = a_i^k \times r_i(k) + b_i^k \times u_i(k)$$

$$G_i(z) = \frac{R_i(z)}{U_i(z)} \Rightarrow G_i(z) = \frac{b_i^k}{z - a_i^k}$$

$$E_{ss} = \lim_{z \to 1}(z-1)(S_i(z) - R_i(z))$$

$$= \lim_{z \to 1} \frac{(z-1)^2 \times (z - a^k) \times S_i(z)}{z^2 + [(C_P + C_I) \times b_i^k - (1 + a_i^k)] \times z + a_i^k - C_P \times b_i^k}$$

吞吐率/带宽SLO精确保障方法

V-Cup控制器

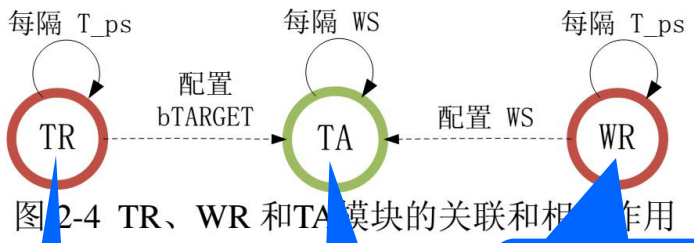


图2-4 TR、WR 和TA模块的关联和相互作用

目标调整模块 (TR)

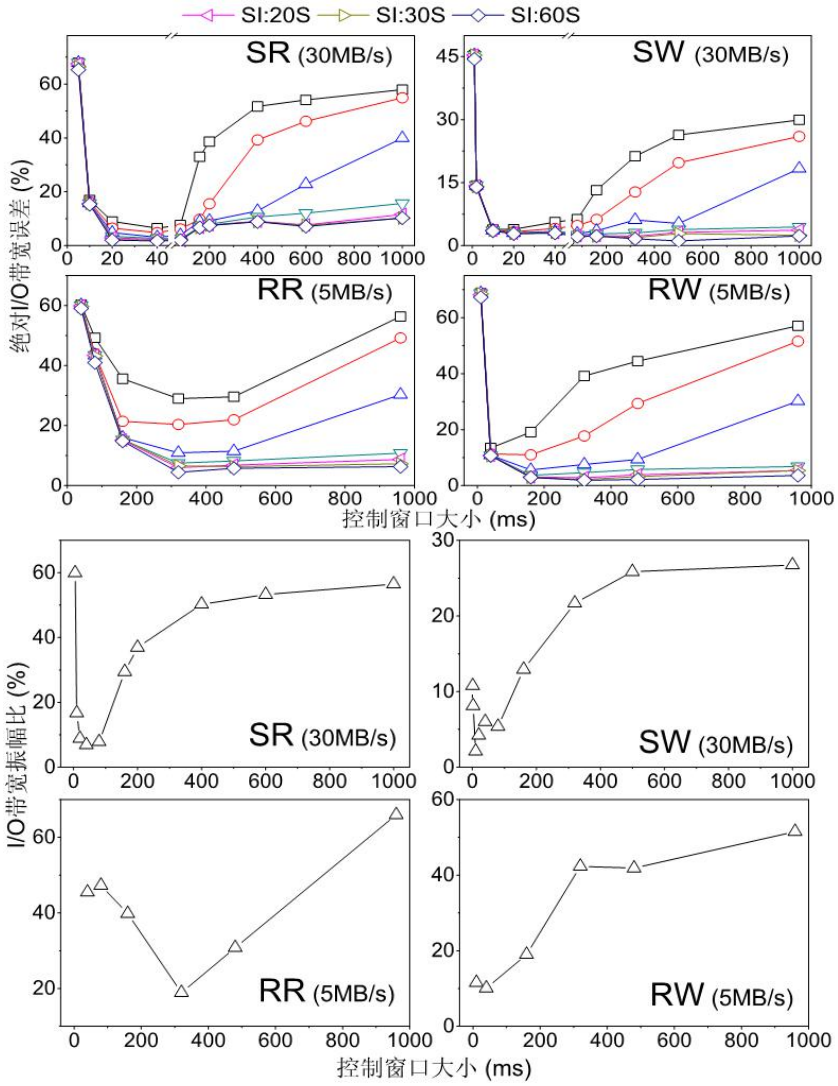
I/O带宽控制模块(TA)

控制时间窗口调整模块 (WR)

负责优化TA的控制窗口大小

I/O带宽的细粒度调整

目标调整模块（TR）通过动态修正TA模块采用的目标值来控制I/O带宽误差。



吞吐率/带宽SLO精确保障方法

V-Cup控制器

ALGORITHM 1: 控制时间窗口调整(WR) 算法

Input: 统计时段T_ps得出的I/O带宽振幅比(V_ps), 统计时段T_ps得出的绝对I/O带宽误差(β_ps) 和I/O带宽SLO (Bw_{SLO}).

Output: 控制窗口大小WS的次优解(ws).

```
1  ws_max = 1024 ms; ws_min = 32 ms; ws = ws_max;
2  repeat
3      if (V_ps − U ≤ last_V) and (β_ps ≤ B_threshold);
4      then
5          ws_max = ws;
6      else
7          ws_min = ws;
8      end
9      ws = (ws_min + ws_max) / 2;
10 until (ws_max − ws_min ≤ UN_WIN) or ((V_ps ≤ BEST_V) and
   (β_ps ≤ B_threshold));
11 return ws
```

假设在算法第一次迭代时，控制窗口大小WS的初始搜索空间大小为n，则该值在第二次迭代时为n/2。

以此类推，在后续的迭代中，搜索的空间大小分别为n/4, n/8, … , $n/(2^k)$。其中，k是最大的迭代次数。

所以，WR算法的时间复杂度可以表示为O(logn)。

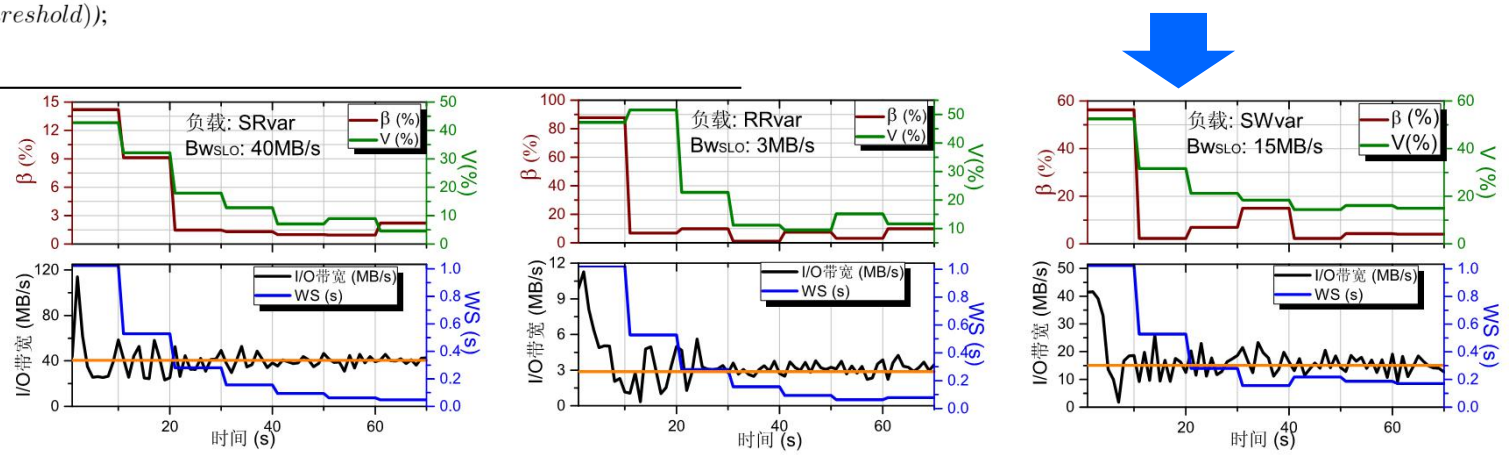


图 2-6 三个典型负载的控制窗口大小调整过程

吞吐率/带宽SLO精确保障方法

性能测试

测试平台介绍：

主机: 2台 PowerLeader PR2760T servers
*) 2 Intel Xeon E5620 quad-core 处理器,
*) 12GB 内存,
*) 10Gbps NIC (Intel 82598EB).

存储: 2类存储子系统
*) 16-disk (7200RPM, 250GB) RAID 0 磁盘阵列
*) Fusion-io ioScale 2, 825GB Multi Level Cell (MLC) SSD

虚拟化软件: Xen 4.2

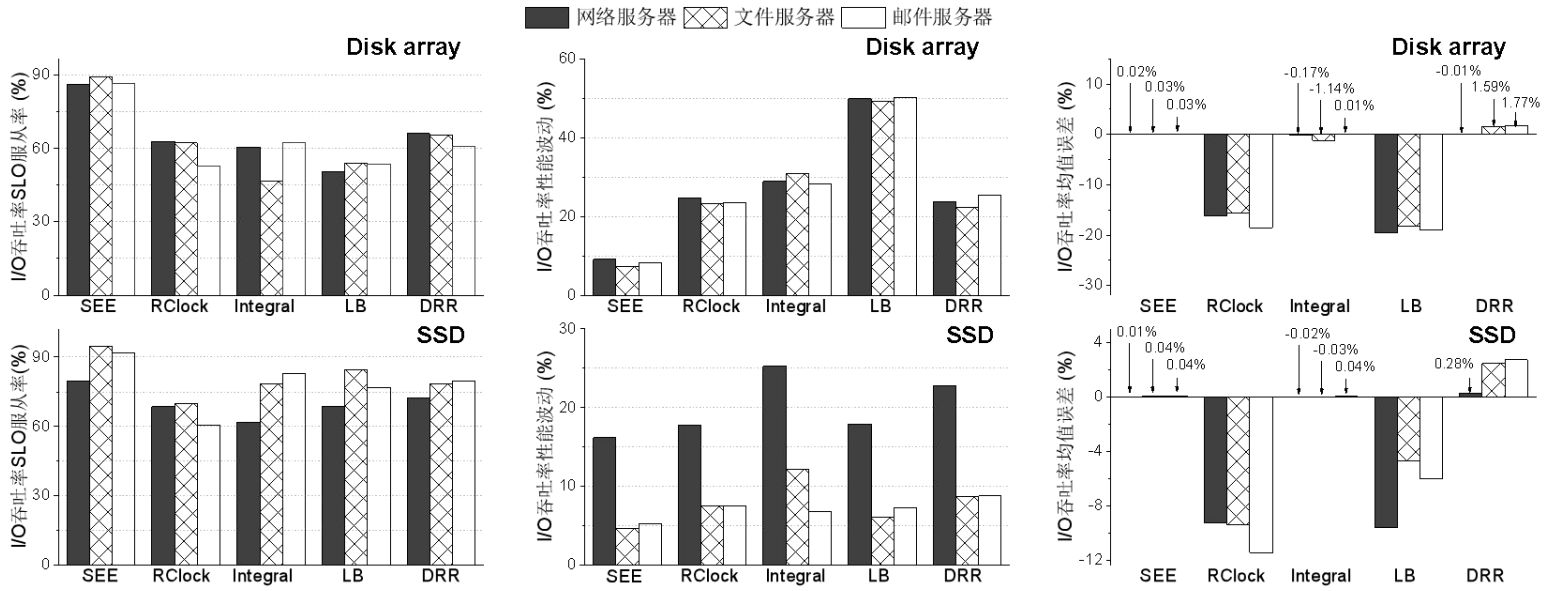
测试目标:

*) 验证SASLO框架支持下I/O吞吐率/带宽的SLO保障精度；
*) 验证SASLO响应SLO秒级变化的能力；
*) 验证SASLO对用户不同SLO策略的支持能力；
*) 考察SASLO对不同数量的聚合虚拟机进行SLO保障时产生的精度差异；
*) 考察SASLO框架支持下I/O吞吐率SLO保障对I/O延时性能的影响。

性能测试

I/O吞吐率SLO保障精度比较

简称	参考算法	说明
RClock	Time-stamp based IO control	按请求的时间戳顺序进行I/O调度。
DRR	Deficit round-robin	一类基于GPS公平调度算法的改进版，它有助于降低控制误差。
LB	Leaky bucket	一种用于IO 节流控制的调度算法
Integral	Integral control	一种反馈控制算法，它保证控制输出的变化和控制误差的积分成比例。

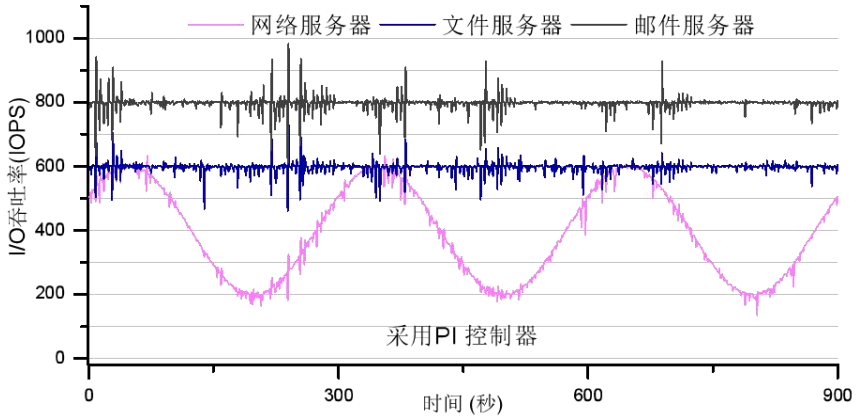


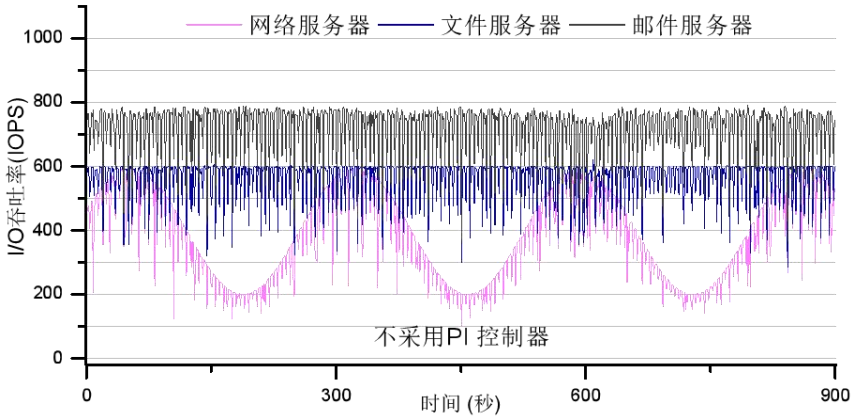
SLO执行引擎（SEE） 在磁盘阵列和SSD作为存储设备时均有较好的SLO保障精度和稳定性。

性能测试

响应SLO秒级变化

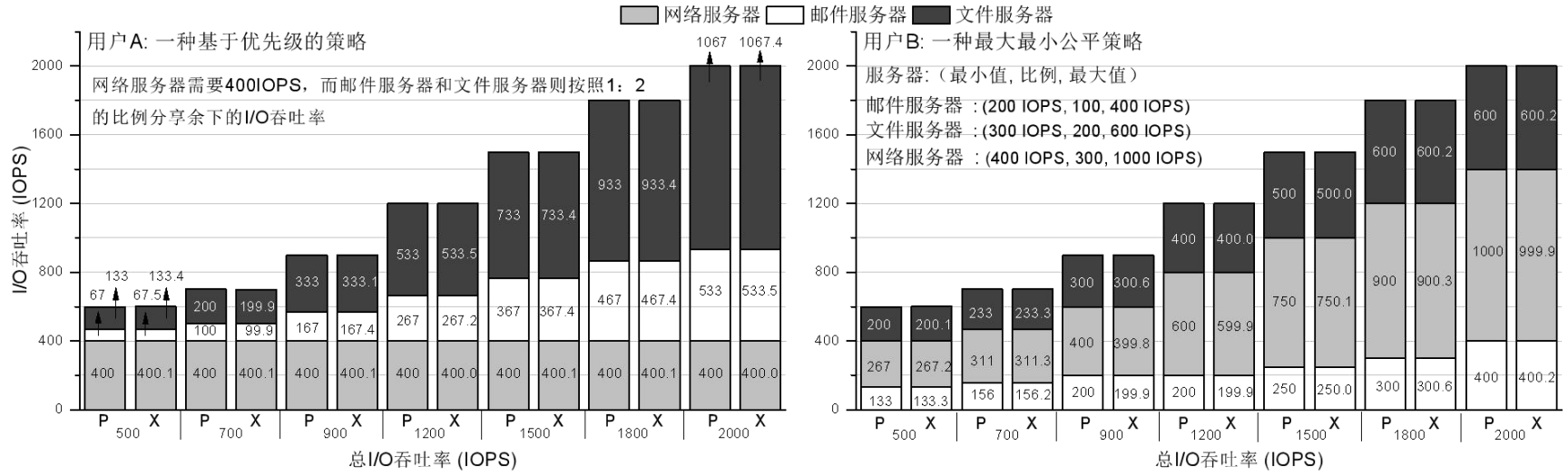
采用比例积分(PI)模型的SLO执行引擎（SEE）对变化的SLO序列具有较好的时间响应。





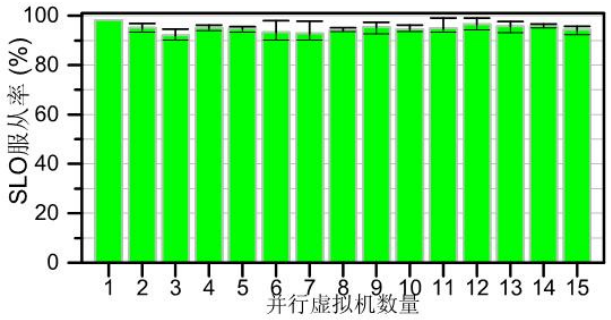
对用户不同SLO策略的支持

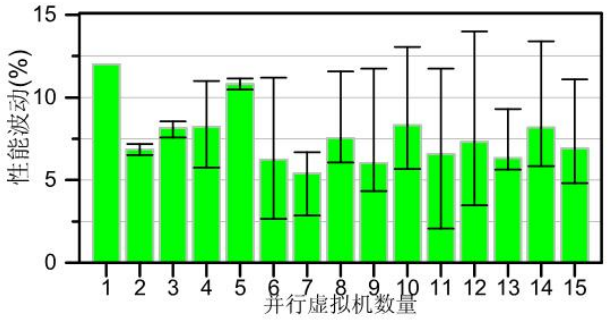
能准确的执行由不同用户策略发出的SLO序列

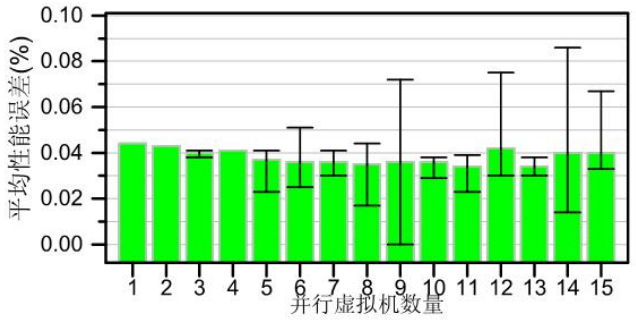


性能测试

对不同数量的并行虚拟机进行SLO保障时产生的精度差异

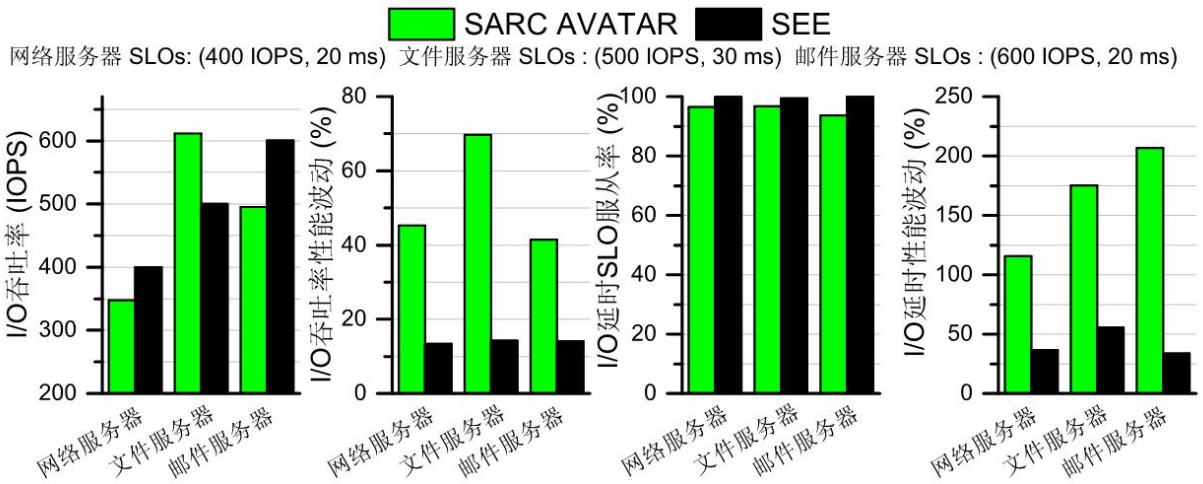






只要存储资源足够（即最大资源利用率< 100%），在SASLO的控制下并行虚拟机数量的增加几乎不会对SLO服从率、性能波动和均值误差产生影响。

I/O吞吐率SLO保障对I/O延时性能的影响:



3台虚拟机在SASLO的SLO执行引擎(SEE)控制下秒级I/O延时的SLO服从率分别为99.92%、99.42% 和99.83%；而在SARC AVATAR系统调度下为96.42%、96.75% 和93.58%。且SEE控制下各虚拟机秒级I/O延时性能波动均远低于SARC AVATAR。

本讲小结

- 为何需要SLO精确保障
- 如何实现SLO精确保障
- 交叉运用控制论经典方法进行创新
  - 提出面向虚拟机的SLO精确保障和基于SLO约束性能优化的解决方案，以实现多性能指标的精确保障。