Smart Client … Smart SLA !

Wladimiro Bedin
BEDIN Shop Systems
www.akite.net

3/1/2011

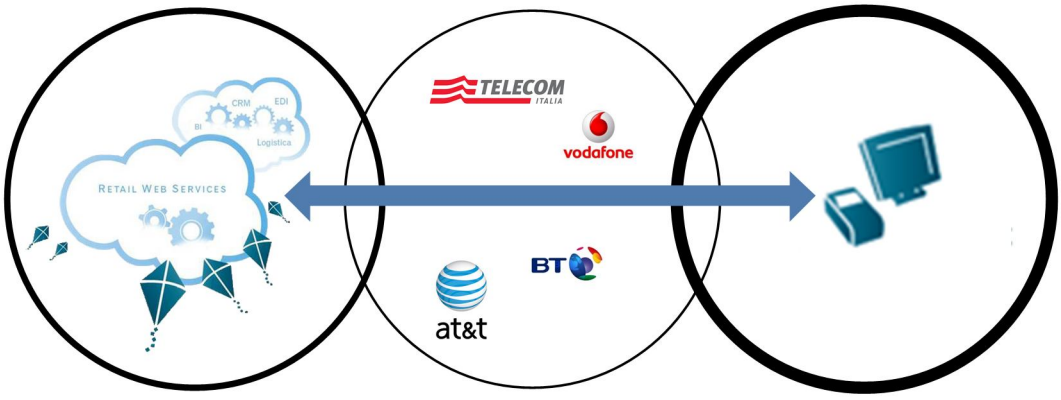
Lo SLA Totale

Nel valutare la qualità dei servizi erogati da Internet spesso ci si concentra sul **Livello di Servizio** (**SLA** da Service Level Agreement) della piattaforma di Cloud Computing da cui vengono distribuiti e si trascura il fatto che il risultato finale dipende anche dallo SLA della connettività fornita dalle compagnie telefoniche e infine da quello dei PC o thin client utilizzati in periferia.

Il livello di servizio di una postazione di cassa deve essere superiore a quello di molti altri servizi commerciali, perché il Retail è caratterizzata da transazioni relativamente piccole, frequenti e non posticipabili. Se a causa di un blackout informatico non fosse possibile emettere scontrini, non si perdono solo vendite, ma addirittura clienti. Diversamente, in un ambiente B2B, spesso l'emissione di una fattura può essere posticipata di uno o più giorni senza gravi disservizi per il cliente finale.

Lo SLA è la % del tempo con servizio regolare, in altre parole la probabilità che tutto funzioni regolarmente. La differenza a 100 è invece la probabilità di un blocco del servizio. Ad esempio, uno SLA superiore al 99,9 % equivale alla probabilità di un guasto inferiore allo 0,1%.

Quando il servizio viene erogato da sistemi complessi che possono avere dei blackout, passa attraverso una linea che si può interrompere ed infine è utilizzato da un PC, è intuitivo ritenere che **la catena non può essere più robusta dell'anello più debole**. In termini matematici, la probabilità di un servizio regolare è pari al prodotto dei singoli SLA.



Per costruire dei ragionamenti quantitativi facciamo le seguenti ipotesi:

$SLA_{Cloud} > 99,9\ \%$ $\quad$ $SLA_{Linea} > 99\%$ $\quad$ $SLA_{PC} > 99,99\%$

I dati sono ipotetici, tranne quelli del Cloud Computing che coincidono con il minimo garantito da Microsoft per Windows Azure. Secondo tali ipotesi, la probabilità di un regolare servizio è

$$\mathbf{SLA_{Totale}} = SLA_{Cloud} * SLA_{Linea} * SLA_{PC} > 99,9\% * 99\% * 99,99\% = \mathbf{98},89\ \%$$

Anche se lo SLA della piattaforma Cloud passasse all'eccezionale valore del 99,999%, la situazione non migliorerebbe in modo apprezzabile. Infatti:

$$SLA'_{Totale} = SLA'_{Cloud} * SLA_{Linea} * SLA_{PC} > 99,999\% * 99\% * 99,99\% = \mathbf{98},989\ \%$$

Quindi con un enorme sforzo per aumentare di cento volte lo SLA del Cloud, si passerebbe dal **98**,89 al **98**,989 % appena. **Del resto è intuitivo capire che è poco utile rafforzare ulteriormente gli anelli più robusti e che invece ci si deve concentrare su quelli più deboli**.

Potrebbe sembrare ingeneroso verso le compagnie di telecomunicazioni averle indicate come anello debole della catena. Chiaramente molte sono in grado di fornire connettività ad alto SLA, oppure è possibile prevedere una linea alternativa via UMTS che svincola dai problemi dell'ultimo miglio, ma aumentando i costi e la complessità. Quando la linea aggiuntiva si dovrà risvegliare dal letargo, lo farà istantaneamente e senza problemi ? La difficoltà non si pone tanto per la sede della catena, vista la bassa incidenza sui costi totali di un secondo o terzo canale verso Internet, ma per i singoli negozi, specialmente se piccoli e numerosi, alcuni dei quali potrebbero non essere nemmeno raggiunti da una normale linea ADSL.

Un aspetto importante delle mancate erogazioni di servizi **non è solo la loro frequenza, ma anche la loro durata**. Nel caso delle normali linee telefoniche è esperienza comune che i guasti siano poco frequenti, ma lunghi, a volte anche più di 24 ore.

Inoltre, quando si considera lo SLA, spesso si trascura di considerare **il periodo su cui viene misurato,** che può essere il mese oppure l'anno. Un guasto di 1 giorno su 1 mese rappresenta una probabilità maggiore del 3,3% (1/30) e quindi un SLA inferiore al 96,7 % mentre, se misurato su base annua, la probabilità crolla allo 0,28 % (1/360) e quindi lo SLA passa ad un più presentabile 99,72 %. **Chiaramente il mese è il periodo massimo su cui va misurato lo SLA.**

Per quanto riguarda i PC, ipotizziamo prodotti professionali, sotto gruppo di continuità e soggetti ad una corretta gestione. E' esperienza comune che, in questa situazione, il PC sia l'anello più robusto della catena. Lo SLA del 99,99% utilizzato nell'esempio rappresenta solo un ordine di grandezza relativo. Inoltre quando è presente più di una postazione di cassa e l'architettura, come quella di **aKite**, non ha singoli punti bloccanti (single point of failure) come server di barriera casse o router, il guasto di un PC o di una sua periferica rappresenta un degrado del servizio, ma non una interruzione. Per questo, nel resto delle nostre considerazioni, gli impatti dei guasti PC sul servizio vengono trascurati.

Lo SLA nei Punti di Vendita è condannato a valori bassi oppure a costi elevati ?

Invece che cercare vie di uscita all'interno di una visione pre-internet, BEDIN Shop Systems ha scelto di passare decisamente ad una visione post-internet basata sui servizi, dividendo quelli che sono frequenti, indispensabili e basati su dati relativamente stabili (prezzi, promozioni ...) dagli altri, richiesti meno frequentemente, basati su dati relativamente variabili e da condividere all'interno dell'intera catena (saldi, giacenze …). I primi sono forniti da un database presente in ogni postazione e i secondi da Cloud.

Gli **Smart Client** sono applicativi leggeri ed **internet-centrici**, il cui deployment è completamente automatico partendo da una pagina web, e quindi a costo quasi nullo e tempi ridotti. Possono utilizzare servizi locali e remoti contemporaneamente, quindi vendere ed eseguire le normali operazioni anche sconnessi, con l'ulteriore e fondamentale vantaggio della velocità. Quando un cliente ha fatto la sua scelta, non vuole perdere tempo in coda alla cassa.

Se in alcune catene i servizi centralizzati fossero ritenuti troppo vitali per essere soggetti allo SLA di una normale linea, basta aggiungerne una di back-up, come si è costretti a fare per le soluzioni basate su thin clients. Ma anche in questo caso, una soluzione basata su Smart Client come **aKite** ha i vantaggi della maggiore scalabilità e della già citata velocità di servizio.

Lo Smart SLA

Partendo da uno SLA della piattaforma di Cloud Computing del 99,9%, l'inevitabile degrado dovuto alla connettività abbassa il valore disponibile in periferia. Per sviluppare le successive considerazioni, ipotizziamo che i valori siano compresi tra:

- 98,9 % con una sola linea (SLA_{Cloud} 99,9 % * SLA_{1Linea} 99 %)
- 99,89 % con una seconda linea di backup (SLA_{Cloud} 99,9 % * SLA_{2Linee} 99,99 %).
  I passaggi matematici per dimostrare che due linee con SLA del 99 % ciascuna offrono un valore totale del 99,99 % sono semplici e vengono omessi.

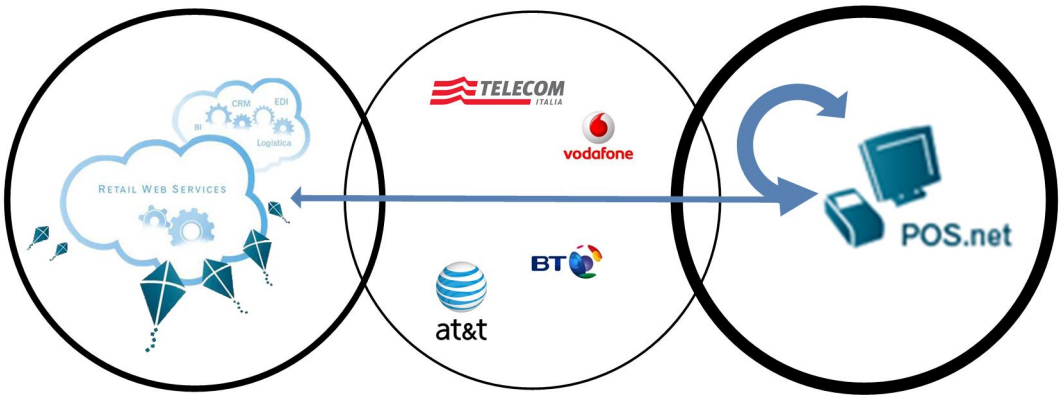
Lo SLA di un servizio basato su Thin Client o su Browser è quindi compreso fra:

$$98{,}9\ \% < SLA_{ThinClient} < 99{,}89\ \%$$

Guardando la situazione dal punto di vista della Probabilità di un guasto, questa risulta essere compresa tra i valori:

$$0{,}11\ \% < P_{Cloud+Linea} < 1{,}1\ \%$$

In una architettura distribuita ed orientata ai servizi come quella di **aKite**, basata su servizi locali e remoti, il problema si pone solamente quando la connettività manca nel momento in cui si deve erogare un servizio centralizzato. Se la connettività manca quando vengono erogati i normali servizi, non si ha alcuna disfunzione.



La frequenza con cui, all'interno dei Punti di Vendita, vengono erogati i servizi speciali che arrivano solo da Cloud dipende dal settore merceologico e, all'interno dello stesso, dal tipo di catena. Si può ipotizzare che incidano dall' 1% fino ad un massimo del 10%.

$$1\ \% < P_{Serv.OnLine} < 10\ \%$$

La probabilità che i due eventi si verifichino contemporaneamente è pari al prodotto delle singole probabilità ed è compresa nel seguente intervallo :

$$0{,}11\ \% * 1\ \% < P_{servizioKO} < 1{,}1\ \% * 10\ \%$$
$$0{,}0011\ \% < P_{servizioKO} < 0{,}11\ \%$$

Trasformato in termini di SLA

$$\mathbf{99{,}89\ \% < SLA_{SmartClient} < 99{,}9989\ \%}$$

Dove il caso peggiore è quello di singola linea e alta frequenza delle operazioni OnLine e il caso migliore è quello della doppia linea e basso ricorso ai servizi centralizzati.

Ricordo che siamo partiti da uno SLA compreso fra 98,89 % e 99,89 % e il miglioramento è stato da uno a due ordini di grandezza (da 10 a 100 volte).

Come si può migliorare lo SLA di un servizio senza intervenire sui singoli componenti attraverso cui viene erogato? Conoscendo anche superficialmente i principi e le leggi di probabilità e fisica, questa sembra una sfida alla logica.

La risposta intuitiva è che **la catena non può essere più robusta dell'anello più debole, ma si può ridurre la probabilità di rottura utilizzandola il meno possibile**. E' quello che abbiamo fatto in **aKite** con i servizi locali, con gli ulteriori e fondamentali vantaggi di maggiore velocità di servizio, maggiore scalabilità e minori costi di erogazione.