

VALUTAZIONE DI PRESTAZIONI

- E' utile ad ogni fase della vita di un "computer system": progetto, costruzione, vendita, uso, aggiornamento
- I sistemi da valutare sono così diversificati che non è possibile definire uno standard in tale area. occorre quindi selezionare:
 - - la corretta misura di prestazioni,
 - - il corretto ambiente di misura,
 - - la tecnica di misura più adatta.
- Ogni valutazione richiede un'intima conoscenza del sistema modellato ed un'accurata selezione di metodologia, workload, e tools.
- Il primo passo è la definizione del problema reale e la sua conversione in una forma in cui sia possibile usare le tecniche e tools più adatti.

Principali Errori

- **Assenza di obiettivi:** non esistono modelli general purpose. Ogni modello deve essere sviluppato con un chiaro obiettivo in mente. E' importante capire il problema ed identificare il problema da risolvere.
- **Approccio non sistematico:** parametri, variabili da misurare e workload non possono essere scelti in modo arbitrario.
- **Performance metrics inadatte**
- **Workload non rappresentativo** delle condizioni reali.
- **Tecnica di valutazione inadatta.** Le tre tecniche utilizzabili sono Simulazione, Modelli analitici, Misure.
- **Trascurare parametri importanti.**

- **Livello di dettaglio non appropriato.** Occorre evitare formulazioni del problema troppo dettagliate o troppo generiche.
- **Errata analisi dei risultati.**
- **Assenza di analisi di sensitività'.**
- **Trattamento inadatto dei valori singolari (outliers).**
- **Inadatta presentazione dei risultati.**
- **Omissione di assunzioni e limitazioni.**

Selezionare una tecnica di valutazione

| CRITERIO | MODELLI ANALITICI | SIMULAZIONE | MISURE |
|---|---------------------------------|--------------------------|---|
| Stadio del sistema in cui si può usare la tecnica | In qualunque stadio | In qualunque stadio | Dopo che il sistema è stato realizzato |
| Tempo richiesto dalla tecnica | Breve (per un esperto analista) | Medio | Variabile a seconda la complessità del sistema |
| I tools usati nelle varie tecniche | Gli analisti | I linguaggi dei computer | Gli strumenti di misura |
| Accuratezza | Bassa | Moderata | Variabile a seconda i strumenti usati e il tipo di misura (diretta o indiretta) |
| Compromesso tra complessità della tecnica e bontà della valutazione | Facile | Moderato | Difficile (perché mettere insieme la soluzione è spesso abbastanza complicato) |
| Costi | Bassi | Medi | Elevati |
| Vendibilità | Bassa | Media | Elevata |

Misure, Modelli Analitici, Simulazione

quando si possono usare:

- Le misure sono possibili solo se esiste qualcosa di simile al sistema da valutare.
- Modelli analitici e simulazioni possono sostituire le misure in assenza di sistemi disponibili.

che precisione hanno:

- I modelli analitici sono imprecisi (impongono semplificazioni)
- Le simulazioni sono più precise ma possono richiedere molto tempo.
- Le misure possono non fornire risultati accurati a causa della unicità (non ripetibilità) di alcuni parametri.

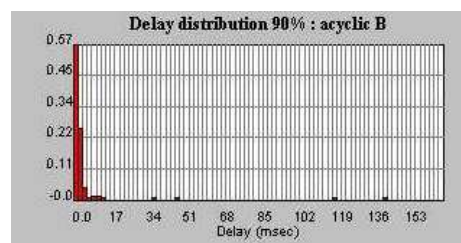
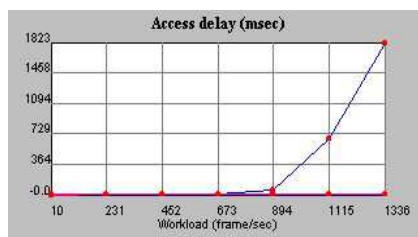
Legame fra i vari parametri

- I modelli analitici permettono di evidenziare l'effetto mutuo di più parametri.
- Con le simulazioni a volte non è chiaro il trade-off fra diversi parametri.
- Le misure rendono difficile interpretare il legame fra vari parametri.

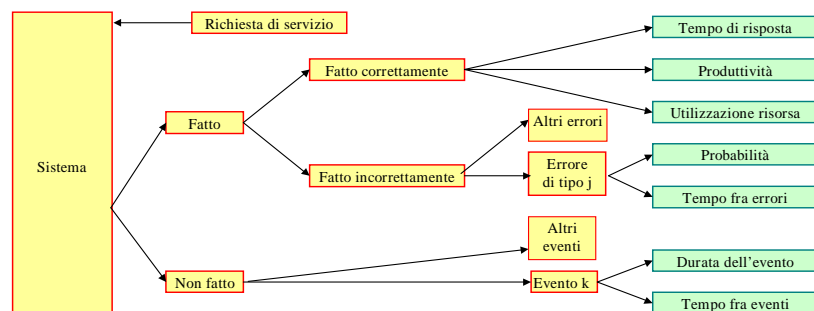
Due o più tecniche possono essere usate in modo sequenziale. Ad esempio un modello analitico trova il range adatto dei parametri e la simulazione studia le prestazioni in quel range.

Valore medio o valore istantaneo?

- Per molte “metrics” il valore medio è quello importante nel senso che le prestazioni del sistema vengono di norma valutate in termini di comportamento medio a regime.
- Tuttavia la variabilità tra i vari valori ottenuti in alcuni casi può essere pericolosa. Ad esempio, in una rete per controllo di processo, il valore del tempo di ritardo medio è un'informazione insufficiente se si ha un'elevata variabilità di valori.
- Un basso valore medio non esclude che ci possano essere diversi valori singoli molto più elevati del valore medio stesso.



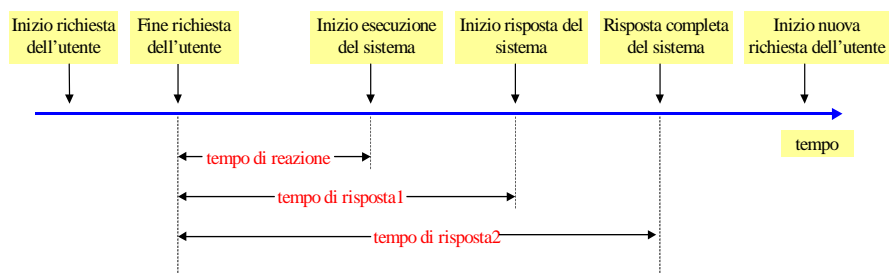
Selezionare le “Performance metrics”.



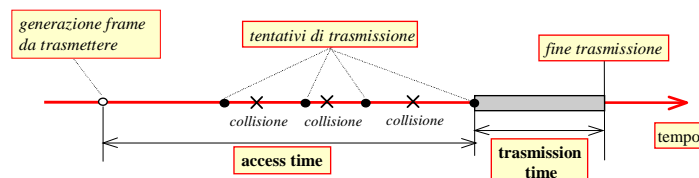
- Un sistema può effettuare un servizio in modo **corretto**, **incorretto**, o **non effettuarlo**.
- Se il sistema esegue il servizio correttamente, le metrics sono chiamate **responsiveness**, **productivity**, **utilization**.
- Se il sistema non opera correttamente è utile classificare gli errori e determinare la **probabilità** di ciascuna **classe di errori**.
- Se il sistema non funziona (unavailable) è utile classificare i **modelli di fallimento** e determinare la **probabilità** di ciascuna classe.

Performance metrics più usate

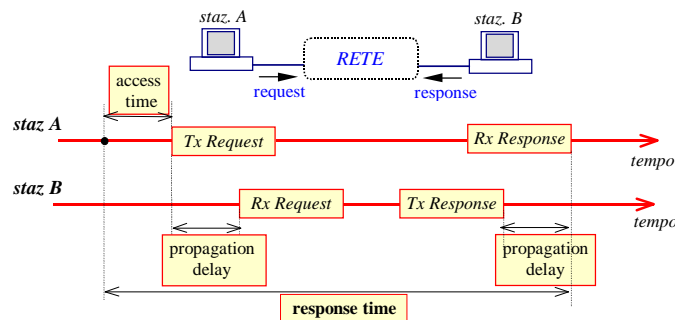
- Response time: intervallo fra la richiesta dell'utente e la risposta del sistema.
- Reaction time: intervallo fra la sottomissione di una richiesta e l'inizio della sua esecuzione.
- Stretch factor: rapporto fra il Response time ad un certo carico e quello a minimo carico.



- Con riferimento alle LAN è utile definire "l'**Access delay**" cioè il tempo che intercorre fra quando un host ha un pacchetto da trasmettere e quando acquisisce l'uso del canale.
- Tale ritardo dipende dalle condizioni del traffico sulla rete e dal particolare protocollo di MAC utilizzato.
- Nel caso di protocolli di tipo CSMA/CD, le collisioni possono influenzare fortemente il tempo di risposta complessivo.



- Il tempo di risposta, nel caso delle reti, può essere definito come il **tempo necessario perché un utente riceva la risposta ad una richiesta inviata ad un altro host**. Esso è composto da:
 - tempo necessario per trasmettere la richiesta (**access time + Tx time**).
 - tempi di propagazione del segnale (**andata e ritorno**)
 - tempo necessario al destinatario per trasmettere la risposta (**access time + Tx time**)
 - Tempo di reazione alla richiesta, da parte del destinatario. (spesso trascurabile)



• **Throughput** : frequenza a cui le richieste possono essere servite dal sistema.

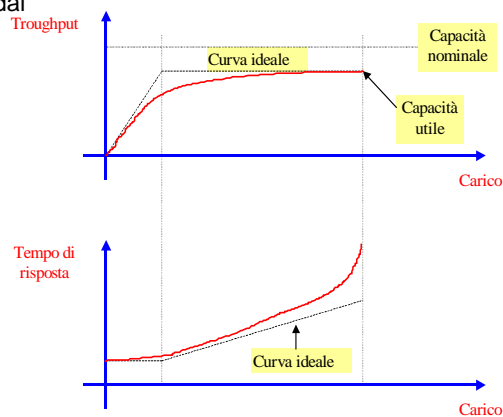
• **Nominal capacity**: massimo throughput in condizioni ideali (bandwidth per le reti).

• **Usable capacity**: massimo throughput ottenibile senza superare un delay prefissato.

• **Efficienza**: rapporto fra massimo throughput (usable capacity) e "nominal capacity".

• **Utilizzazione**: frazione di tempo in cui una risorsa e' impegnata per servire una richiesta.

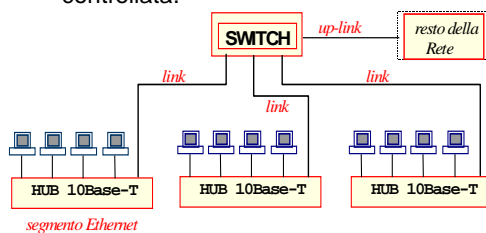
• **Reliability**: tempo medio fra gli errori.



Availability: frazione del tempo in cui un sistema è disponibile per le richieste dell'utente.

WORKLOAD 1/4

- Il workload consiste di richieste di servizi oppure di utilizzo di risorse da parte degli utenti di un sistema.
- L'ambiente vero dell'utente (real-user environment) non è in genere ripetibile, pertanto per realizzare il giusto workload è necessario studiare tale ambiente, osservarne le caratteristiche chiave e costruirne un modello da potere usare ripetutamente.
- **Real workload:** e' quello osservato su un sistema durante le normali operazioni. Non e' ripetibile.
- **Syntetic workload:** ha caratteristiche simili a quelle del "Real workload" di cui costituisce un modello. Puo' essere applicato ripetutamente, in maniera controllata.



System under test (SUT):
denota l'insieme di componenti che si stanno valutando.

Component under test:
denota il singolo componente del SUT considerato.

WORKLOAD 2/4

I parametri che caratterizzano il workload devono dipendere dal workload e non dal sistema. Quelle caratteristiche che hanno un impatto significativo sulle prestazioni del sistema vanno incluse fra i parametri del workload;

- La richiesta più frequente.
- utilizzare un *miscuglio di richieste a frequenza diversa*
- utilizzare un *Trace di richieste di un sistema reale*
- Utilizzare *richieste mediate di servizi con una prefissata probabilità di distribuzione nel tempo: tipicamente distribuzioni esponenziali o gaussiane.*

WORKLOAD 3/4

- **Periodico-** Nei sistemi campionati il traffico è soprattutto costituito da campioni generati in maniera periodica. Il workload sarà quindi costituito da pacchetti generati con una frequenza costante, pari a quella di campionamento.
- **Asincrono-** Questo tipo di workload è legato a fenomeni non periodici (tipicamente allarmi). In tal caso i pacchetti dati saranno normalmente di piccola dimensione, ma la frequenza di generazione sarà effettuata mediante una opportuna funzione random. Questo tipo di traffico di solito coesiste con quello periodico (ma ovviamente è molto minore) e va dimensionato in funzione delle condizioni operative che si vogliono rappresentare.

WORKLOAD 4/4

- **Distribuzione di Bernoulli:** una variabile può assumere solo i valori $X=0$, $X=1$ che determinano fallimento o successo.
 P = probabilità di successo, $1-P$ = probabilità di fallimento.
La distribuzione di Bernoulli modella il verificarsi o meno di un evento e viene utilizzata, ad esempio, per modellare la probabilità che un pacchetto venga danneggiato dal rumore.
- **Distribuzione Binomiale:** Il numero di successi X in una sequenza di n tentativi di Bernoulli ha una distribuzione binomiale, adatta a modellare il numero di successi in una sequenza di n tentativi indipendenti. Ad esempio, il numero di bit disturbati in una frame, ecc.
- **Distribuzione esponenziale:** È la sola distribuzione continua *memoryless*, questo significa che ogni campione che noi generiamo non ha memoria del campione precedente. Viene usata per modellare il tempo fra due eventi successivi indipendenti (tipicamente per la generazione del traffico in una rete di calcolatori).

Valutazione attraverso la simulazione

- La simulazione è utile soprattutto quando il sistema da valutare non è disponibile.
- Attraverso la simulazione è possibile confrontare alternative diverse con diversi workload ed ambienti. Una volta realizzato il modello del sistema da simulare, è possibile, cambiando il workload e quindi le condizioni al contorno, eseguire diverse valutazioni sullo stesso sistema.
- Realizzare l'ambiente desiderato, in un sistema reale è molto complesso (a volte impossibile), costoso e richiede molto tempo. Con un simulatore, implementare le diverse condizioni al contorno richiede solo la configurazione di un certo numero di parametri.
- Va comunque ricordato il grosso rischio della simulazione: *l'errore è sempre in agguato*. Occorre sempre verificare con molta attenzione il modello realizzato e *non fidarsi mai ciecamente dei risultati ottenuti*.
- Alla fine della simulazione, è possibile ottenere delle curve che forniscono una rappresentazione visiva dei risultati.

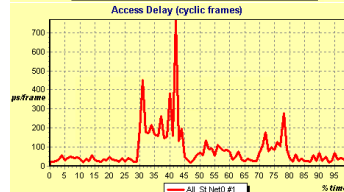
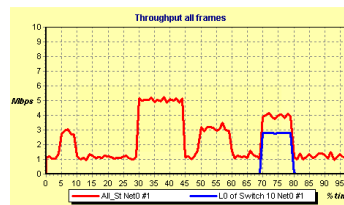
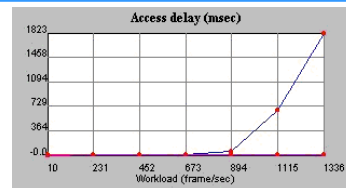
Simulazione stazionaria e transitori

- La simulazione stazionaria ricava il valore degli indici prestazionali prescelti, a regime, per un prefissato punto di lavoro.

La presenza di transitori non può essere evidenziata, e può falsare i risultati ottenuti.

- Una simulazione che evidenzi la risposta di un sistema ad eventi transitori deve rappresentare il comportamento del sistema nel tempo.

Ciò non esclude la realizzazione di statistiche o regime, a fine simulazione.



Implementazione di un simulatore

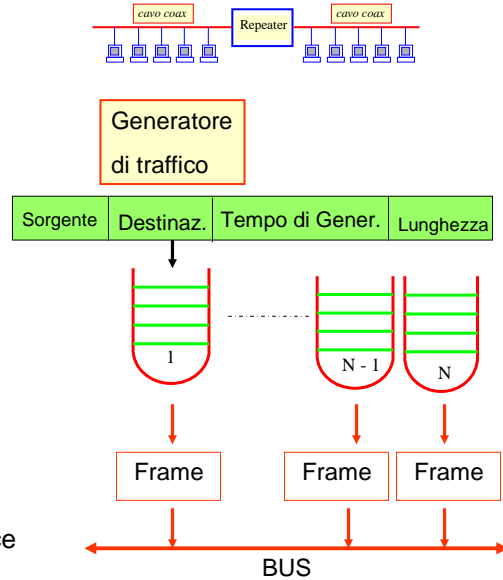
Esempio: rete Ethernet

Elementi critici da modellare:

- Generatori di traffico
- code in ogni host
- trasmissione/rivelazione delle collisioni.
- Event scheduler.
- Time.

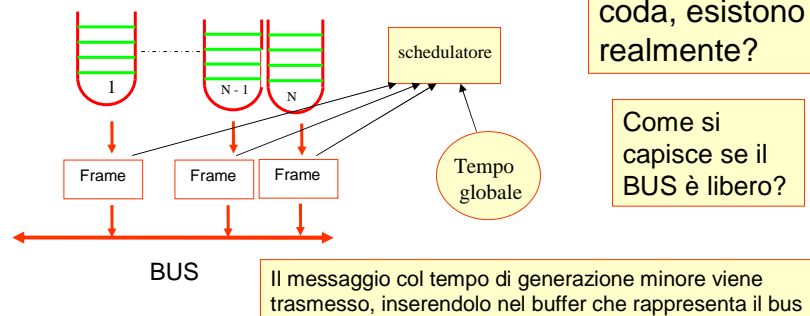
Le input routines, initialization routines e report generator non rappresentano un problema.

- Il **generatore di traffico** è una routine che, utilizzando una opportuna funzione di distribuzione temporale, inserisce in coda le frame da spedire.

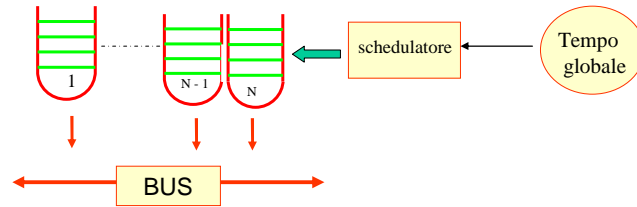


Trasmissione delle frames

- Quando il **bus è libero**, tutti gli host che hanno un messaggio in coda cercano di trasmetterlo.
- Lo scheduler **deve capire quali host hanno un messaggio pronto** da trasmettere. Lo fa confrontando il tempo globale con il tempo di generazione dei messaggi in testa in ogni coda.



Rivelazione delle collisioni



- Quando due o più frames sono trasmesse contemporaneamente (o quasi) si ha una collisione.
- Lo schedulatore rivela una collisione confrontando il tempo di inizio trasmissione con quello di generazione delle frames in testa. Deve esaminare tutte le code.

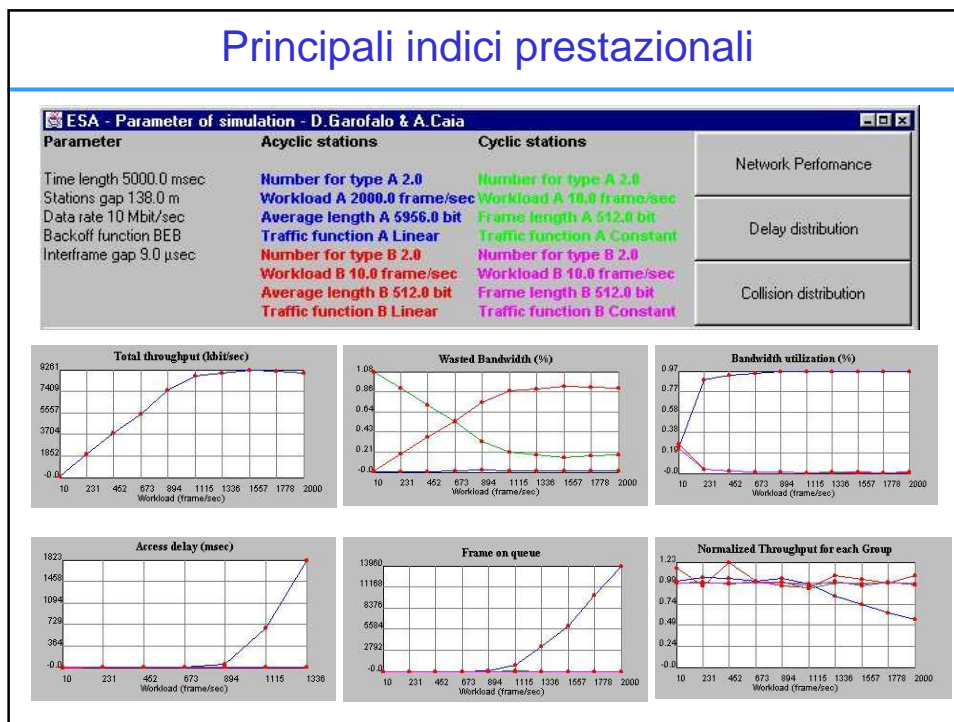
$\text{Tempo_inizio_transmiss.} - \text{tempo di generaz.} < \text{slot_time} \Rightarrow \text{COLLISIONE}$

- Algoritmo di backoff esponenziale
- Il tempo di generazione va aggiornato come:
 $\text{istante_di_collisione} + \text{ritardo_di_backoff}$

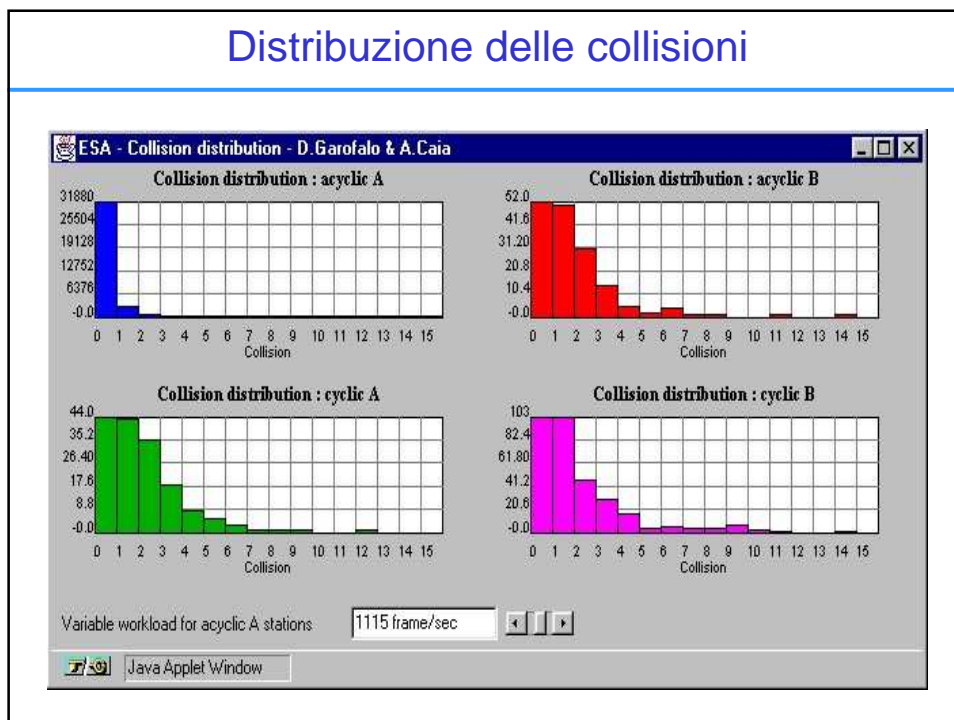
Clock e Scheduler

- Il **clock** è l'elemento di riferimento del simulatore.
- **Avanza a scatti**, di una quantità pari alla durata dell'evento attivato.
- Lo scheduler confronta il clock con gli eventi possibili e lo incrementa della durata dell'evento da attivare.
- Se l'evento attivato è una trasmissione, **verifica se c'è una collisione**, incrementa il clock della durata della frame (o del tempo perduto nella collisione), ed effettua le operazioni sui dati, necessarie per le statistiche finali.
- **Effettua una nuova scansione delle code** per individuare la nuova frame da trasmettere.

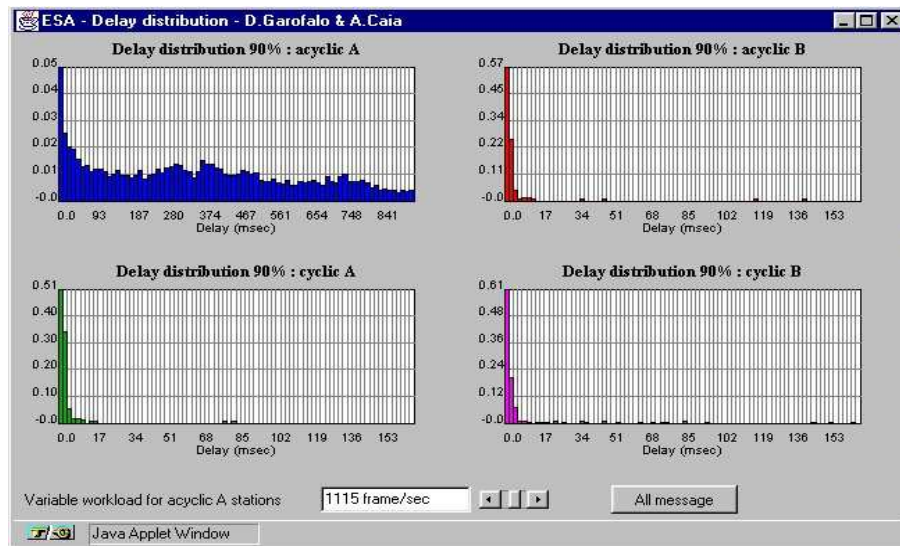
Principali indici prestazionali



Distribuzione delle collisioni



Distribuzione dei tempi di ritardo



Analisi degli errori nella simulazione

- Uno dei pericoli nella valutazione di prestazioni è la possibilità di commettere errori:
- **Livello di dettaglio inappropriato**
- **Linguaggio inadatto**
- **Modello non valido**
- **Condizioni iniziali errate**
- **Tempo di simulazione troppo breve**

Tipi di simulazione

- **Montecarlo Simulation**

E' una simulazione statica. Viene usata per analizzare dei fenomeni di tipo probabilistico e tempo invarianti. Adatta per fenomeni descivibili analiticamente.

- **Trace Driven Simulation**

In questo tipo di simulazione usiamo come ingresso una *trace* di eventi di un sistema reale per ottenere una *trace* di uscita. Vantaggi: credibilità, workload accurato, facilità di confronto, somiglianza con implementazioni reali.

- **Discrete Event Simulation**

Si implementa un modello a stati discreti del sistema, cioè un modello in cui lo stato del sistema evolve a scatti anziché in modo continuo. In questo modo possiamo far avanzare lo stato del sistema in funzione degli eventi che ci interessa mettere in risalto. OMNet++, NS2,3

Analisi dei risultati

Durante lo sviluppo di un modello di simulazione occorre garantire che esso sia rappresentativo del sistema reale (**validazione del modello**) e che esso sia implementato in modo corretto (**verifica del modello**).

La validazione del modello è **legata alla correttezza delle assunzioni fatte** sul comportamento del sistema (cioè al fatto che abbiamo modellato il sistema in modo corretto o il nostro modello contiene qualche errore).

La verifica (debugging sul modello implementato) è **relativa alla correttezza della sua implementazione**, cioè al SW.

Tecniche di validazione del modello

- Assunzioni
- Valori dei parametri di ingresso e distribuzioni
- Valori d'uscita e conclusioni
- Intuizione di esperti
- Misura sui sistemi reali
- Risultati teorici

Tecniche di verifica del modello

- Modular design
- Top-down design
- Antibugging
- Modelli deterministici
- Eseguire dei casi semplificati
- Continuity test

Misure sui sistemi reali

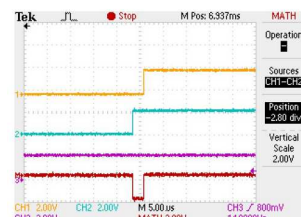
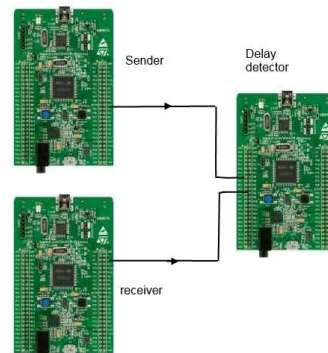
Effettuare delle misure su sistemi reali è più difficile che valutare un sistema mediante la simulazione. Operando su un sistema fisico non c'è bisogno di realizzarne il modello, (con i relativi possibili errori) ma nascono problemi che in un sistema simulato non sono presenti.

Elenchiamone alcuni e vediamo come risolverli:

- Misura dei tempi
- Misura del Throughput
- Definizione dell'ambiente operativo
- Valutazione del Physical layer e del mezzo fisico.

Misura del tempo

- In un sistema reale, poiché i clock dei vari nodi sono fra di loro scorrelati non possiamo confrontare le letture effettuate nei vari nodi.
- - Misurare il tempo di andata e ritorno (esempio invio di un pacchetto e ricezione dell'Ack).
- - Misurare il tempo impiegato da una singola trasmissione, mediante l'ausilio di un terzo dispositivo



Misura del Throughput

- Basta realizzare nel sender un modulo software che generi traffico e contare nel receiver il numero di pacchetti ricevuti correttamente.
- Il Throughput può essere calcolato contando il numero di pacchetti moltiplicati per la loro dimensione.
- Se il protocollo di comunicazione prevede l'uso automatico dei retry, può essere utile contarli (se l'HW lo consente) in modo da avere una statistica sul numero medio di ritrasmissioni effettuate.
- Una misura del Throughput deve tenere conto di uno scenario realistico (in cui sono presenti diversi sender) in modo da tenere conto di possibili collisioni o della necessità di effettuare una schedulazione delle trasmissioni da parte di un Master.
- Uno scenario formato solo da un sender ed un receiver è adatto a calcolare il Max throughput che il sistema può fornire in condizioni ideali.

Definizione dell'ambiente operativo

- L'ambiente operativo può essere estremamente variabile a seconda dell'applicazione considerata.
- Le caratteristiche più interessanti possono riguardare:
- **Tipologie di traffico**: percentuale di traffico periodico ed asincrono.
- **Densità dei nodi** e valori del workload
- **Rumore ambientale**
- **Mobilità dei nodi** (ad es. applicazioni di vehicle to Vehicle V2V communication)
- Presenza di **ostacoli metallici** (tipicamente in ambienti industriali)

Valutazione del Physical layer 1/3

- La struttura del physical layer e le caratteristiche del mezzo fisico influenzano molto le prestazioni del sistema.
- In accordo al modello a livelli, le prestazioni dell'applicazione sono influenzate da quelle di tutti i livelli sottostanti.
- Il physical layer è il livello più basso e mette a disposizione un canale digitale con un certo bit rate. Più elevato è il bit rate offerto dal livello Phy maggiore è la banda disponibile per i livelli superiori.

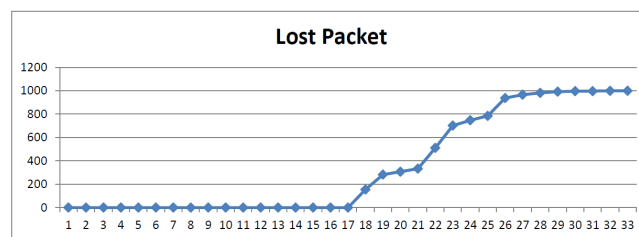
Valutazione del Physical layer 2/3



- L' nRF24L01+ è un ricetrasmittitore wireless a chip singolo a 2,4 GHz con embedded baseband protocol engine .
- La principale valutazione è quella relativa alla distanza coperta sia indoor che in open air. Influenza la topologia della rete nel senso che in base alla distanza coperta si può optare per una rete single-hop o multi-hop.

Valutazione del Physical layer 3/3

- Può essere interessante valutare il numero di pacchetti persi al variare della distanza .
- Le misure possono essere ripetute per diversi valori del bit rate (per il dispositivo considerato varia da 250Kbps a 2 Mbps)
- Le misure possono essere ripetute a diverse frequenze (visto che il dispositivo permette il frequency hopping) per vedere se alcune frequenze sono più favorevoli di altre, ed in presenza di stazioni interferenti, su canali vicini.



Il transitorio iniziale e la durata della valutazione.

- Dopo lo sviluppo del modello occorre decidere quante osservazioni iniziali devono essere scaricate prima che il modello sia in uno stato stazionario, e quanto tempo debba durare una simulazione.
 - Questi due aspetti non vanno considerati nelle valutazioni in condizioni transitorie, poiché si visualizza il comportamento di un sistema nel tempo.
 - Questi due punti sono invece cruciali nell'esecuzione di una simulazione o di una misura in condizioni stazionarie: infatti, durante il transitorio iniziale, prima che il sistema si porti a regime, si ottengono dati che in genere non sono significativi poiché le condizioni operative non sono stabili. E' inoltre importante comprendere quanto tempo debba durare la simulazione, in modo da collezionare una quantità adeguata di dati, da analizzare.
- **Transient removal:** rimozione del transitorio iniziale
 - **Stopping criterion:** scegliere i criteri per bloccare la simulazione.

Eliminazione del transitorio iniziale

- Quando in una valutazione interessa solo la *Steady-state performance* (Performance nello stato stazionario) i risultati della parte iniziale della valutazione (transitorio iniziale) vanno rimossi.
- **Long runs:** la simulazione dura talmente a lungo che le condizioni iniziali non influenzano i risultati. Non garantisce il risultato.
- **Proper initialization:** consiste nel fare iniziare la simulazione in uno stato vicino a quello stazionario. Difficile da usare.
- **Initial data deletion:** durante la fase stazionaria, la media dei valori ottenuti non cambia molto. Si eliminano valori iniziali fin quando la loro eliminazione provoca variazioni sensibili nella media.
- **Batch means:** una lunga simulazione è divisa in parti (batch) di uguale durata. Si eliminano i batch relativi al transitorio iniziale utilizzando nelle statistiche solo i dati relativi ai batch in condizioni stazionarie, mediando fra i vari batch.

Arresto della simulazione

- Per determinare il tempo di simulazione, cioè il tempo necessario affinché il sistema sia a regime, ed abbia prodotto un numero sufficiente di campioni per le statistiche non esistono regole precise.
- In genere basta eseguire due simulazioni alle stesse condizioni ma con durate diverse; se i risultati ottenuti sono simili allora vuol dire che siamo a regime.
- A questo punto occorre contare il numero di campioni della grandezza di interesse (pacchetti spediti, numero di collisioni, ecc.) .
- Se il numero è sufficiente per una statistica significativa (almeno qualche centinaia/migliaia di campioni) si sceglie il tempo di simulazione inferiore fra i due.

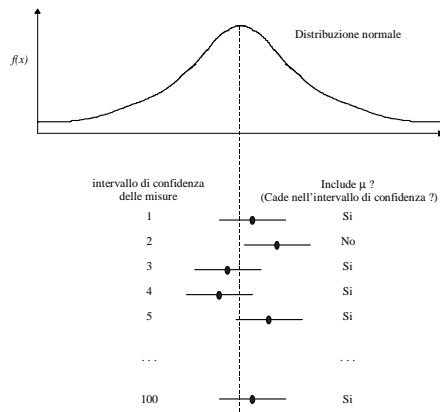
Analisi dei sistemi da campioni 1/4

- Quando eseguiamo delle misure sui sistemi otteniamo delle misure puntiformi, nel senso che le valutazioni di prestazioni che noi eseguiamo non sono relative al comportamento del sistema nella sua globalità, ma a specifici punti di lavoro del sistema.
- Non è possibile ricavare le caratteristiche esatte di un sistema basandosi solo sull'osservazione di alcuni campioni.
- Dato un campione (X_1, X_2, \dots, X_n), indichiamo con X_m la media relativa al campione. Tale media è diversa da quella μ della popolazione di dati misurati (il valore μ si ottiene considerando un numero infinito di valori).
- X_m è il valore misurabile, definito *Statistica*
- μ è il valore cercato, definito *Parametro*

Analisi dei sistemi da campioni 2/4

- La stima di un parametro di una popolazione, data da un solo numero (per esempio, il valore X_m) è definita **Stima Puntuale** del parametro stesso.
- Se la stima è data da due numeri che si possono considerare, uno maggiore ed uno minore del parametro reale, allora si parla di **Stima per Intervallo**.
- Definiamo *Intervallo di confidenza* un intervallo (C_1, C_2) che con una elevata probabilità $P = 1 - \alpha$ contiene la media μ della popolazione.

Analisi dei sistemi da campioni 3/4



Supposto di avere eseguito cento misure, il numero di quelle che cadono dentro sarà $100(1-\alpha)$, mentre quelle che cadono fuori saranno $100(\alpha)$.

$100(1-\alpha)$ rappresenta la percentuale di confidenza. **$(1-\alpha)$** si chiama *coefficiente di confidenza*

Analisi dei sistemi da campioni 4/4

- <https://www.mccallum-layton.co.uk/tools/statistic-calculators/confidence-interval-for-mean-calculator/>

3.1 Calculator

Enter Sample Size

Enter Sample Observed Mean

Enter Sample Observed Standard Deviation

Select Desired Confidence Level (%)

3.1.1 Results

3.1.1.1 Confidence Interval: ± 0.04

3.1.1.2 Range for the true population mean: 32.96 to 33.04

mette a disposizione un tool per il calcolo dell'intervallo di confidenza una volta nota la dimensione del campione, il valor medio, la deviazione standard ed il livello di confidenza desiderato.