



# Università degli Studi di Roma
# *Tor Vergata*

Facoltà di Scienze Matematiche Fisiche e Naturali

*Corso di Laurea in Fisica*

Anno accademico 2008/2009

Tesi di Laurea

***"Evoluzione dei servizi informatici forniti dal Centro di Calcolo di un grande Laboratorio di ricerca in fisica fondamentale"***

Laureando
**Massimo Pistoni**

Relatore
**Prof. Piergiorgio Picozza**
Università degli studi di Roma
Tor Vergata

Correlatore
**Dott. Sergio Bertolucci**
I.N.F.N. – Laboratori
Nazionali di Frascati

Consegnata: Maggio 2010

***Dedicata a mia madre e a mio padre***

*Un ringraziamento particolare a **Sergio Bertolucci**,*
*per avermi dato gli stimoli necessari per raggiungere importanti traguardi*

*Ringrazio lo staff del Servizio di Calcolo dei LNF*
*per la realizzazione dei servizi descritti e per il supporto fornito*
*a compimento di questo lavoro:*
***Sandro Angius***
***Claudio Bisegni***
***Mario Masciarelli***
***Dael Maselli***
***Claudio Soprano***
***Dario Spigone***

*Desidero ringraziare **Nunzio Amanzi** e **Antonino Passarelli***
*che hanno collaborato con lo staff del Servizio di Calcolo*
*alla realizzazione di alcuni importanti servizi*

*Desidero ringraziare **Danilo Babusci***
*per il supporto fornito sul capitolo introduttivo*
*e **Elisabetta Vilucchi***
*per il supporto fornito sul capitolo dedicato alle griglie computazionali*

*Un ringraziamento speciale a mia **moglie** e a mio **figlio***
*che hanno saputo supportarmi in questi anni di studio,*
*sopportando le mie frequenti indisponibilità*

*Un saluto carico di commozione a*
***Mario Masciarelli***
*che ci ha lasciato tragicamente e prematuramente,*
*ma il cui magnifico ricordo rimarrà indelebile nel mio cuore.*

# Indice

# Elenco delle figure

# Elenco delle tabelle

# 1. Introduzione

## *1.1 Presentazione dell'ambiente di lavoro*

Il lavoro presentato nelle seguenti pagine è frutto di un'attività di stage svolta, presso i LNF (Laboratori Nazionali di Frascati) dell'INFN (Istituto Nazionale di Fisica Nucleare).

## *1.2 L'Istituto Nazionale di Fisica Nucleare*

L'INFN è un'organizzazione dedicata allo studio dei costituenti fondamentali della materia e svolge ricerca teorica e sperimentale nel campo della fisica subnucleare, nucleare e astroparticellare.

La ricerca fondamentale in queste aree richiede l'uso di strumentazione e tecnologie di frontiera, che l'INFN sviluppa sia nei suoi Laboratori che in collaborazione con il mondo dell'industria.

Inoltre l'INFN promuove l'applicazione delle professionalità, dei metodi e delle tecniche sperimentali sviluppate nel corso delle proprie attività di ricerca, anche per la ricerca in altri campi, come la medicina, la preservazione del patrimonio artistico e la protezione ambientale.

Queste attività sono sempre state condotte in stretta collaborazione con il mondo accademico. Gruppi di scienziati delle Università di Roma, Firenze, Milano e Padova, fondarono l'INFN nel 1951 con l'obiettivo di continuare il percorso di ricerca teorica e sperimentale nella fisica nucleare, già stabilito negli anni 1930 da Enrico Fermi e la sua scuola.

Già alla fine degli anni 1950, l'INFN progettò e costruì il primo acceleratore di particelle italiano, l'elettrosincrotrone sviluppato a Frascati, il luogo di nascita dei primi Laboratori Nazionali dell'Istituto (rif. Figura 1-1).

Durante lo stesso periodo, l'INFN cominciò a partecipare alle attività di ricerca nella costruzione e l'uso del più grande e potente acceleratore di particelle mai costruito fino ad allora, condotte al CERN, l'Organizzazione Europea per la Ricerca Nucleare, a Ginevra.

Oggi, i ricercatori dell'INFN danno un importante contributo alla ricerca non soltanto in vari laboratori Europei, ma anche in numerosi centri di ricerca mondiali.

### 1.2.1 Da AdA al LEP

Le prime decadi del 1900 testimoniano un'intensa attività nello studio della struttura dell'atomo. L'osservazione dei decadimenti radioattivi e l'interazione di fasci di particelle con la materia hanno permesso ai fisici di effettuare scoperte fondamentali.

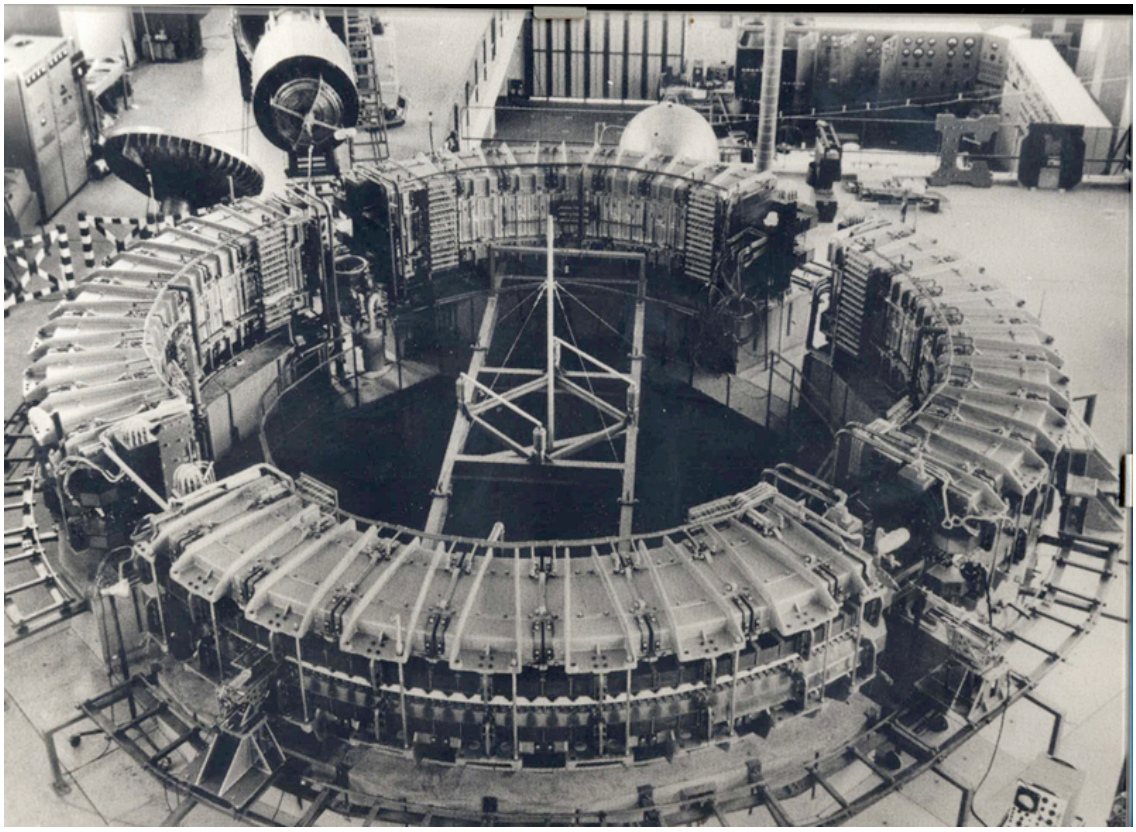


**Figura 1-1: L'Elettro-Sincrotrone costruito a Frascati tra il 1957 e il 1959 era in grado di accelerare gli elettroni fino ad 1,1 GeV di energia.**

Allo stesso tempo, attraverso lo studio dei raggi cosmici, divenne possibile osservare particelle che non erano state viste fino ad allora.

Successivi sviluppi tecnologici facilitarono la costruzione di acceleratori, il cui uso ha permesso l'esplorazione dei costituenti fondamentali del nucleo atomico e la creazione di nuove forme di materia. Negli acceleratori lineari, le particelle sono dirette su un obiettivo fisso (*target*), mentre negli acceleratori circolari, due fasci di particelle, accelerati in direzioni opposte, vengono fatti collidere l'uno con l'altro.

In entrambi i casi, l'energia acquistata dalle particelle prima dell'impatto, è convertita in nuove forme di materia durante l'interazione, e ciò rende possibile l'osservazione di nuove particelle.

Il primo acceleratore di tipo concettualmente più elegante, AdA (Anello di Annichilazione, rif. Figura 1-2), fu disegnato e costruito durante gli anni 1960 ai Laboratori Nazionali di Frascati. In AdA, il cui diametro era di circa un metro, venivano fatti collidere due fasci, uno di elettroni e uno di positroni. L'uso di macchine di questo tipo, basate sulla collisione di materia e antimateria, è responsabile delle maggiori scoperte degli ultimi trenta anni nel campo della fisica delle particelle, e ha consentito la dettagliata verifica della maggioranza delle predizioni teoriche.

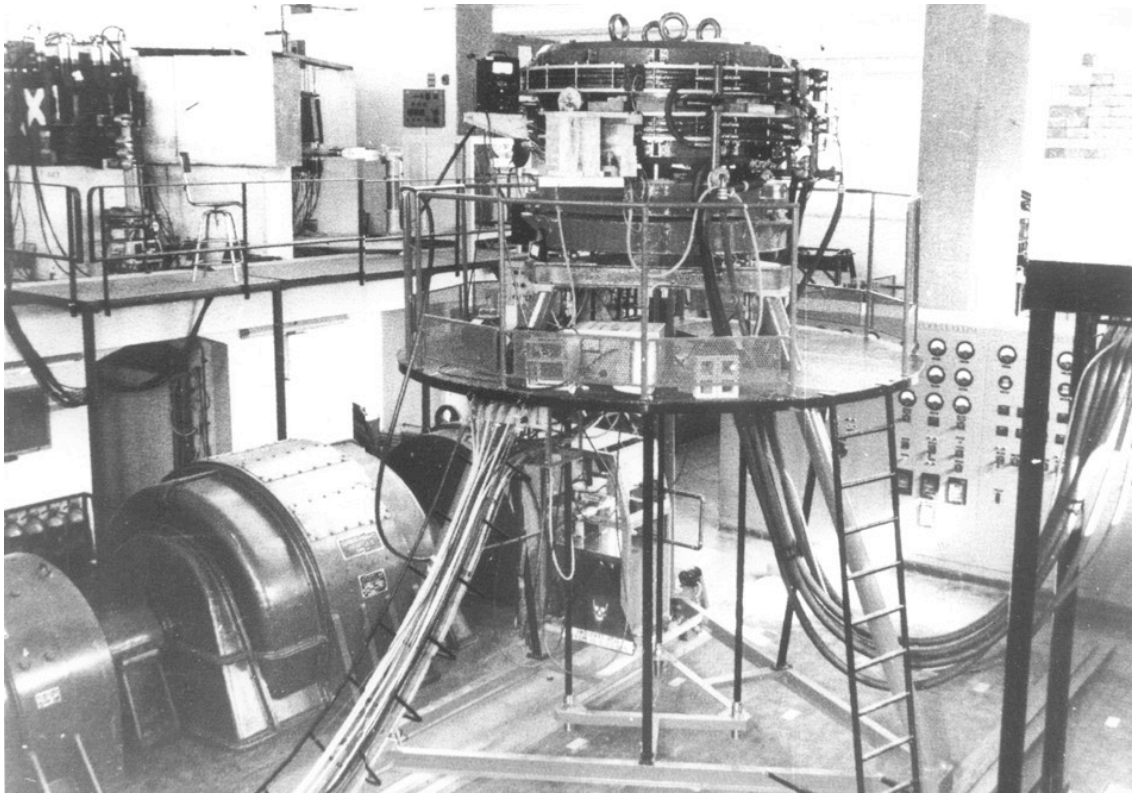


**Figura 1-2: AdA, il primo anello di annichilazione a Frascati**

Negli anni, allo scopo di studiare il comportamento delle particelle ad energie sempre più alte, sono stati costruiti acceleratori sempre più grandi, da AdA (con la sua circonferenza di un metro) all'anello sotterraneo LEP (*Large Electron Positron collider*) in uso al CERN dal 1989 al 2000, con la sua circonferenza di 27 chilometri (rif. Figura 1-3). L'uso del LEP ha permesso lo studio delle particelle W e Z, che sono responsabili del decadimento radioattivo β, e supportano il Modello Standard, il modello teorico sul quale è basata la moderna descrizione del mondo subnucleare.

**Figura 1-3: Il tunnel del LEP al CERN**

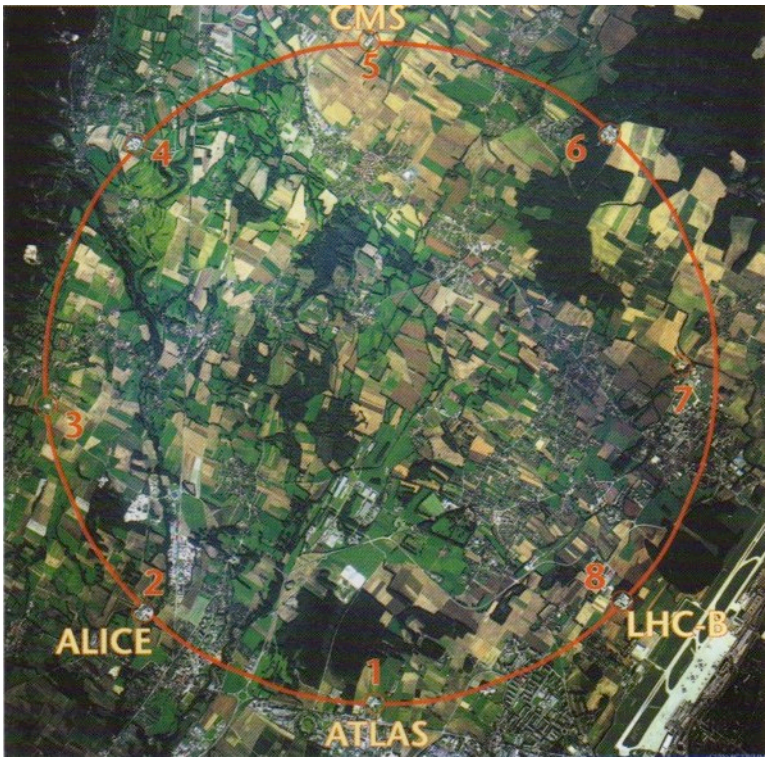
La ricerca al LEP ha aperto la strada verso una nuova generazione di esperimenti che mirano alla scoperta del solo pezzo mancante rimasto della teoria: la particella di *Higgs*.

LHC (*Large Hadronic Collider*), il nuovo acceleratore del CERN, è un collisore protone – protone e ha come scopo fondamentale proprio la verifica sperimentale dell'esistenza di questa particella. Ci si aspetta inoltre che con LHC sarà possibile ottenere prove dell'esistenza di nuove particelle previste dall'estensione supersimmetrica del modello standard.

### 1.2.2 LHC: la frontiera dell'energia

Alla fine degli anni 2000, terminava l'era del LEP, dopo oltre dieci anni di intensa attività. Circa 2000 ricercatori di tutto il mondo, di cui 300 dell'INFN, parteciparono alla costruzione e al coordinamento dei quattro esperimenti del LEP: ALEPH, DELPHI, L3 e OPAL. Il LEP consentì la verifica del Modello Standard con una precisione mai raggiunta prima.

La costruzione di LHC nel tunnel del LEP (di 27 km di circonferenza) è giunta al termine. È molto recente la notizia delle prime collisioni tra fasci di protoni ad un'energia nel centro di massa pari a 7 TeV. I suoi quattro esperimenti, ALICE, ATLAS, CMS, LHC-B, dovrebbero partire all'inizio del 2011.




Questi esperimenti, non solo permetteranno di verificare i dati relativi alla particella di Higgs, ma focalizzeranno la loro attenzione su altri problemi irrisolti, come la violazione della simmetria materia-antimateria e le caratteristiche di certi stati di materia quali il plasma quark-gluone.

L'interazione tra due fasci di protoni nell'anello sotterraneo produrrà 40 milioni di collisioni al secondo, ognuna delle quali produrrà migliaia di particelle. Queste particelle saranno tracciate da vari differenti rivelatori in grado di fornire decine di milioni di informazioni per ogni collisione.

La richiesta potenza di calcolo risultante dall'enorme quantità di dati prodotti dagli esperimenti e la necessità di distribuire questi dati nel mondo attraverso la rete, ha condotto i ricercatori verso lo sviluppo di nuovi modelli di calcolo: le griglie computazionali (GRID).

### 1.2.3 ATLAS e CMS

Gli esperimenti ATLAS e CMS hanno gli stessi obiettivi: acquisire dati fondamentali alla ricerca della particella Higgs, e scoprire le particelle supersimmetriche previste nelle estensioni del Modello Standard.

I due esperimenti si distinguono per le differenti tecnologie utilizzate per osservare le traiettorie e le misure delle energie delle particelle prodotte nelle collisioni.

L'esperimento ATLAS è il risultato della collaborazione di circa 2000 scienziati di 150 istituti in 34 paesi. Molte caratteristiche dell'apparato sperimentale rappresentano dei veri e propri record mondiali. Ad esempio: la misura della posizione delle particelle con un'accuratezza di 14 micron, 80 km di cavi superconduttore e il più grande circuito magnetico mai costruito, di 26 metri di lunghezza, costruito in Italia.

L'INFN è anche responsabile della realizzazione del magnete superconduttore di CMS, che produrrà un campo magnetico di 4 Tesla, 80.000 volte superiore al campo magnetico terrestre. Tale campo magnetico è in grado di immagazzinare tanta energia da fondere una tonnellata di acciaio.

### 1.2.4 LHC-B e ALICE

I mesoni B, particelle che contengono quark b, rivestono un posto fondamentale per la comprensione della violazione delle simmetria tra la quantità di materia e di antimateria nell'universo. LHC-B esaminerà i processi per i quali i mesoni B decadono, e osserverà tutte le particelle prodotte nelle collisioni con precisione estremamente alta.

In alcuni periodi dell'operatività di LHC, fasci di nuclei di piombo rimpiazzeranno i fasci di protoni. L'acceleratore sarà quindi usato per le collisioni di questi fasci, creando energie 300 volte superiori a quelle ottenibili con i fasci di protoni, alte abbastanza per riprodurre le condizioni esistenti durante i primi istanti di vita dell'universo.

L'esperimento ALICE studierà il particolare stato che la materia assume a queste energie, noto come plasma di quark-gluoni.

### 1.2.5 Organizzazione dell'INFN

L'attività di ricerca dell'INFN si basa su due tipi di strutture di ricerca complementari: le sezioni e i Laboratori Nazionali. Le 20 sezioni hanno sedi in

dipartimenti universitari e realizzano il collegamento diretto tra l'Istituto e le Università. I 4 Laboratori, con sede a Catania, Frascati, Legnaro e Gran Sasso, ospitano grandi apparecchiature e infrastrutture della comunità scientifica nazionale e internazionale.


**Figura 1-4: L'INFN sul territorio nazionale**

Il personale dell'INFN conta circa 2000 dipendenti propri e quasi 2000 dipendenti universitari coinvolti nell'attività dell'Istituto e 1300 giovani tra laureandi, borsisti e dottorandi.

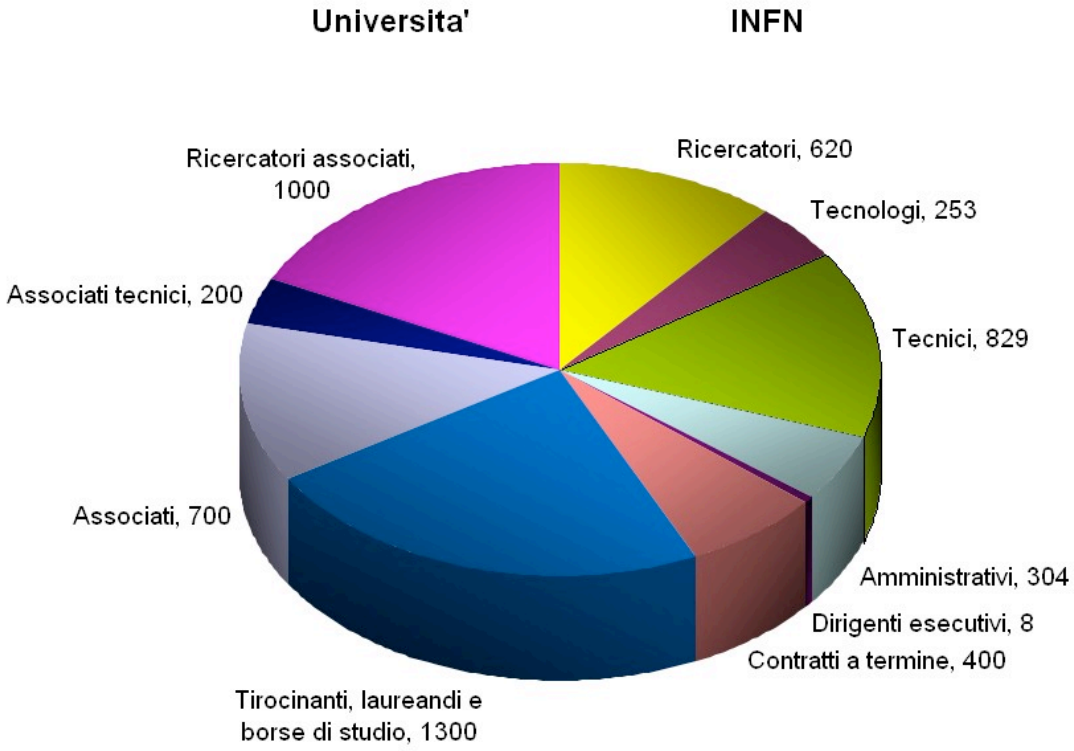



**Figura 1-5: Composizione del personale dell'INFN**

L'organizzazione dell'INFN rappresenta un efficace equilibrio fra gestione centralizzata e decentralizzata ed è frutto di consuetudini consolidate negli anni.

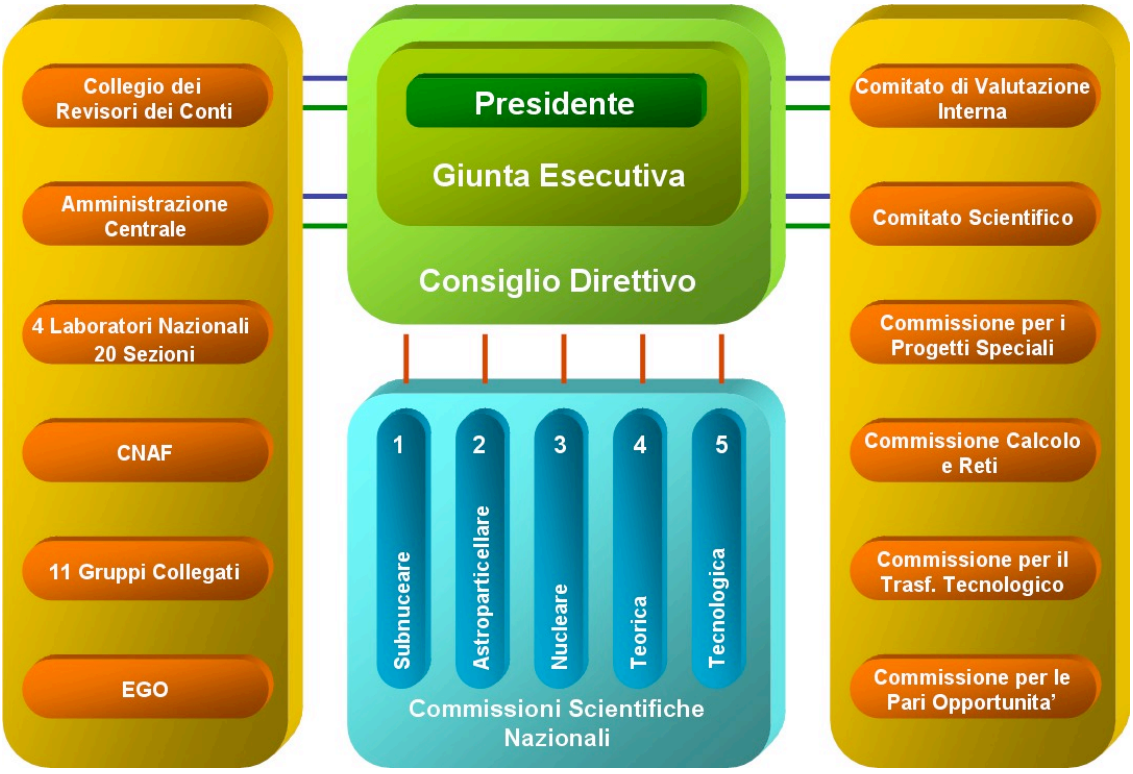



**Figura 1-6: Diagramma dell'organizzazione dell'INFN**

L'organo decisionale dell'Istituto è il Consiglio Direttivo, costituito dal Presidente e dalla Giunta Esecutiva, dai quattro Direttori dei Laboratori Nazionali e 20 Direttori delle Sezioni, da rappresentanti del MIUR, del Ministero dell'Industria, del CNR e dell'ENEA.

L'attuazione delle decisioni del Consiglio compete, secondo i casi, al Presidente, alla Giunta, ai Direttori di Laboratorio o di Sezione per l'organizzazione delle attività a livello locale, il tutto con l'ausilio dei dirigenti dell'Amministrazione Centrale.

## *1.3 I Laboratori Nazionali di Frascati*

I Laboratori Nazionali di Frascati, fondati nel 1955, sono i più grandi tra i 4 laboratori dell'INFN.

### 1.3.1 Cenni storici: ADONE

Successivamente alla realizzazione dell'elettro-sincrotrone e della macchina acceleratrice AdA già accennate nei paragrafi precedenti, ai LNF nel 1961 si decise di costruire una nuova macchina, ADONE, un acceleratore di elettroni e positroni all'energia di 1,5 GeV. Una volta ultimata la realizzazione, nel 1969, ADONE deteneva il record mondiale per l'energia.

I risultati ottenuti grazie alla nuova macchina, contribuirono ad aprire nuove frontiere nella fisica delle particelle elementari. Un risultato particolarmente rilevante fu l'abbondante produzione di particelle adroniche dalla collisione tra elettroni e positroni; risultato inatteso e anomalo rispetto ai modelli teorici dell'epoca.

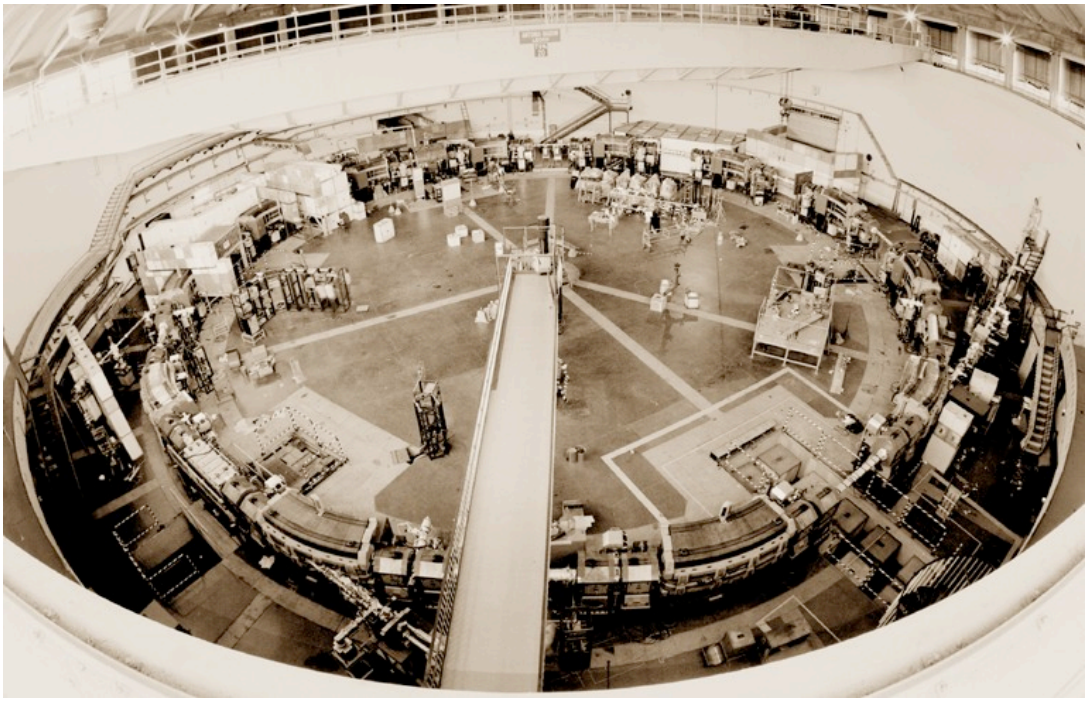


**Figura 1-7: ADONE: acceleratore di elettroni e positroni**

L'interpretazione di tale fenomeno pose la base sperimentale per l'attuale teoria delle interazioni nucleari forti: Cromodinamica quantistica (QCD).

Nel novembre 1974 ADONE contribuì a confermare la scoperta della particella J/Ψ (massa di 3,1 GeV), prodotta simultaneamente pochi giorni prima nei laboratori di Stanford e Brookhaven (USA).

Da qui fu dimostrata l'esistenza di un nuovo tipo di quark, chiamato *charm*, la cui esistenza era stata ipotizzata da Glashow, Iliopoulos e Maiani,.

Oltre alla ricerca nella fisica delle particelle subnucleari, ADONE fu anche usato per studiare la fisica nucleare e la struttura della materia. Grazie alla produzione di luce di sincrotrone, ADONE ha permesso di effettuare studi in molti campi di ricerca: dalla fisica dei solidi, alla chimica, alla biologia e alla medicina.

L'attività di ADONE è terminata nel 1993 per lasciare spazio alla nuova macchina acceleratrice di nuova generazione: DAFNE.

### 1.3.2 DAΦNE

L'attività principale a Frascati durante gli ultimi anni è stata la costruzione di una nuova macchina: DAFNE, acceleratore ad alta luminosità per la collisione di elettroni e positroni (rif. Figura 1-8).

**Figura 1-8: DAFNE: acceleratore elettroni-positroni ad alta luminosità**

DAFNE (o DAΦNE, Double Annular Φ Factory for Nice Experiments) è costituita da un doppio anello per le collisioni elettroni-positroni con un'energia di 0,51 GeV per fascio. È la prima di una nuova generazione di macchine dedicate allo

studio ad alta precisione di fenomeni estremamente rari. A tale scopo questi acceleratori devono essere in grado di produrre un grande numero di collisioni tra i fasci ed avere un'energia precisamente definita.

La prima di queste due caratteristiche è chiamata luminosità. DAFNE ha raggiunto nel 2009 una luminosità di circa $5 * 10^{32}$ $cm^{-2}sec^{-1}$, che è cinquanta volte maggiore della massima luminosità ottenuta alle stesse energie.

Uno degli aspetti innovativi di DAFNE è che gli elettroni e i positroni circolano in due anelli distinti per ridurre le reciproche interferenze. I due fasci si attraversano soltanto in due punti dove si intersecano ad un piccolo angolo orizzontale. Ciò permette un numero particolarmente alto di pacchetti (*bunch*) circolanti.

All'interno degli anelli principali di DAFNE (lunghi 100 metri), 120 bunch di particelle collidono ogni $3 * 10^{-9}$ s ad un angolo di 25 mrad. Nei punti di interazione, le loro dimensioni sono circa 4mm * 40μm * 6 cm. Ogni bunch è composto da 90 miliardi di particelle le cui collisioni producono circa 5000 particelle Φ al secondo.

Per assicurare tali alte performance, DAFNE usa tecnologie estremamente avanzate. Il vuoto nella macchina è particolarmente alto (inferiore a $10^{-12}$ atm) per evitare collisioni tra i fasci e le particelle di gas residue nella camera da vuoto dove circolano i fasci (*beam pipe*).

Un sofisticato sistema di *feedback* elettronico misura in continuazione la posizione delle orbite e corregge le traiettorie dei bunch. Un sistema di allineamento molto preciso mantiene i fasci di dimensioni compatte. Sofisticate cavità a radiofrequenza compensano l'energia persa dai fasci ad ogni giro. Un efficiente sistema di controllo progettato e costruito ai LNF raccoglie continuamente informazioni sullo stato di ogni componente di macchina e sulle caratteristiche del fascio.

DAFNE include anche un acceleratore lineare (LINAC) tramite il quale gli elettroni e i positroni vengono accelerati all'energia richiesta, e un piccolo anello di accumulazione, tramite il quale i fasci acquisiscono le caratteristiche ottimali prima di essere iniettati nella beam pipe.

DAFNE è anche un'interessante sorgente di luce di sincrotrone nella regione dell'ultravioletto e dei raggi X. A Frascati la ricerca con la luce di sincrotrone di DAFNE, segue una consolidata tradizione iniziata con ADONE.

### 1.3.3 Gli esperimenti locali

DAFNE produce collisioni tra elettroni e positroni con un'energia totale di 1,02 GeV. Al momento della collisione, gli elettroni e i positroni si annichilano, ovvero scompaiono e si trasformano in energia pura. Questa energia viene immediatamente riconvertita in massa, ovvero in nuove particelle, in accordo con la nota relazione di Einstein $E=mc^2$.

All'energia di 1,02 GeV elettroni e positroni danno vita ad una particella, la Φ, che ha infatti energia di 1,02 GeV. La Φ non è una particella stabile, ma decade

spontaneamente, dando origine ad altre due particelle: i kaoni o mesoni K. Esistono 4 tipi di mesoni K: $K^+$, $K^-$, $K^0_S$ e $K^0_L$.

Questi mesoni K sono le particelle usate nei tre esperimenti che in questi anni hanno preso dati:

- KLOE (K LOng Experiment),
- FINUDA (FIsica NUcleare a DAFNE),
- DEAR (DAΦNE Exotic Atom Research).

#### 1.3.3.1 KLOE

KLOE è mirato a studiare la sottile differenza che esiste tra la materia e l'antimateria che si evidenzia nel decadimento dei mesoni K. In particolare, i mesoni $K^+$ e $K^-$ sono reciprocamente antiparticelle, mentre $K^0_S$ e $K^0_L$ differiscono soltanto rispetto alla simmetria CP, che risulta dalla combinazione della trasformazione di coniugazione di carica (C) e di quella di parità (P). Se la simmetria CP fosse rispettata, tutte le leggi conosciute della natura non farebbero alcuna distinzione tra la materia e l'antimateria e il numero di particelle e antiparticelle non muterebbe nei processi che coinvolgono la loro creazione e distruzione. La conservazione del numero delle particelle e antiparticelle può essere violata soltanto se una delle due specie è privilegiata rispetto all'altra; e ciò avviene se la simmetria CP è violata.

Comunque, l'universo è composto solo da particelle e benché siano state condotte molte ricerche sulla terra o nello spazio, non sono state rilevate tracce significative dell'esistenza dell'antimateria nel cosmo. La teoria moderna dell'origine dell'universo, il Big Bang, parte con l'ipotesi che l'universo primordiale fosse inizialmente formato equamente da particelle e antiparticelle, ed è grazie alla violazione di CP che le particelle possono aumentare rispetto alle antiparticelle, con l'evoluzione dell'universo.

La violazione di CP nel decadimento dei K fu scoperta nel 1964. Il suo valore e il motivo per cui essa avvenga, tuttavia, non sono note con abbastanza precisione. KLOE, misurando la violazione di CP con una precisione di $10^{-4}$, sarà in grado di aggiungere dati che possono essere cruciali per capire l'evoluzione dell'universo.

KLOE registra la tipologia dei decadimenti dei K neutri. I K prodotti dalla Φ sono infatti particelle instabili. I tempi di decadimento (tempi di vita medi) variano, in funzione del tipo, da alcuni centesimi ad alcune decine di miliardesimi di secondo. Questi si trasformano in pioni (mesoni π), che hanno massa inferiore.

Esistono 3 tipi di pioni: $\pi^+$, $\pi^-$ e $\pi^0$. Prima del decadimento, i K percorrono un cammino che va da pochi centimetri ad alcuni metri, a partire dal punto in cui sono stati prodotti.

KLOE registra il decadimento dei K in due o tre π. Tali decadimenti avvengono con frequenze leggermente differenti, in funzione del tipo di K che li origina. Queste differenze, dovute alla violazione di CP, sono molto piccole e sono trovate in rari casi (circa uno su mille). Per effettuare le misure con la precisione richiesta, KLOE deve analizzare molti miliardi di decadimenti.

Il rivelatore KLOE racchiude la zona di interazione con un apparato cilindrico largo 4m e lungo 4 metri. L'apparato consiste di due elementi principali: una camera a deriva (*drift chamber*), per misurare le caratteristiche dei $\pi$ carichi e tracciarne le traiettorie, e un calorimetro elettromagnetico per i $\pi$ neutri. Questi rivelatori, come anche quelli usati in altri esperimenti a DAFNE, hanno caratteristiche tecniche molto avanzate. La camera a deriva è in grado di misurare l'energia dei $\pi$ con una precisione dello 0,3% e ricostruisce il punto di decadimento dei K con un errore di circa 3 mm. Il calorimetro misura l'energia dei $\pi$ neutri con una precisione del 15% e ricostruisce il punto di decadimento con un errore di circa 1 cm. Queste operazioni sono eseguite per ogni singolo decadimento dei K, generati dai 5000 $\Phi$ prodotti in un secondo.

L'alta frequenza di eventi, ciascuno dei quali con grandi quantità di informazioni, richiede necessariamente l'uso di sistemi di elettronica estremamente veloci e sistemi di acquisizione, immagazzinamento ed elaborazione estremamente potenti.

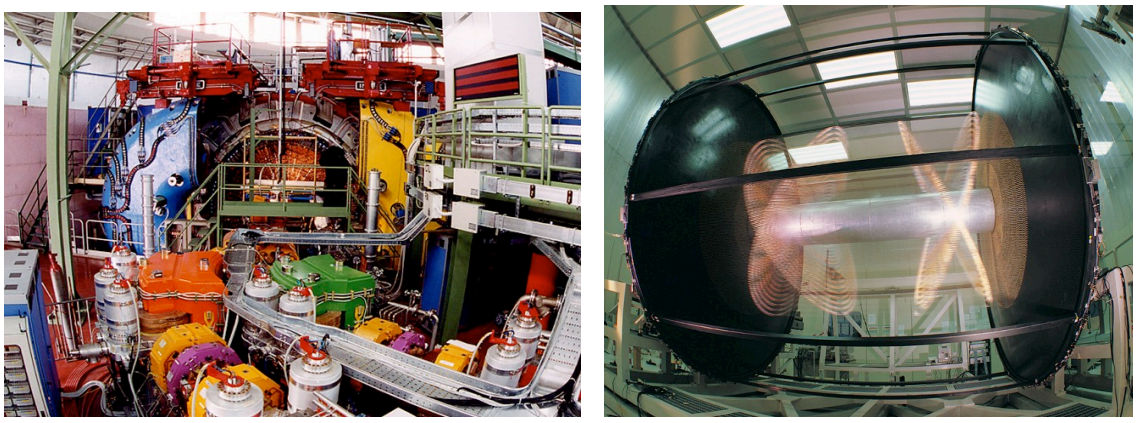


**Figura 1-9: Esperimento KLOE: a sinistra il calorimetro, a destra la camera**

#### 1.3.3.2 FINUDA

L'esperimento FINUDA usa i K, generati dal decadimento della $\Phi$, per produrre un nuovo stato di materia nucleare: l'ipernucleo *lambda*. Nella materia ordinaria, i nuclei atomici consistono di protoni e neutroni, collettivamente conosciuti come nucleoni. Dentro al nucleo, i nucleoni si legano privilegiando i livelli sempre più vicini alla superficie del nucleo. Tutto ciò, seguendo il principio di esclusione di Pauli per il quale due particelle identiche e con gli stessi numeri quantici, non possono occupare lo stesso livello.

Nell'impatto del K con il nucleo, uno dei nucleoni e il K incidente si annichilano, dando luogo alla creazione di una particella $\Lambda$ o barione $\Lambda$. Il nuovo nucleo, chiamato ipernucleo $\Lambda$, conterrà un nucleone in meno e un barione $\Lambda$ in più. Essendo $\Lambda$ differente dai nucleoni, esso può eludere il principio di Pauli, e si lega a livelli più profondi, sondando di conseguenza le parti più profonde e inesplorate del nucleo. L'ipernucleo $\Lambda$, tuttavia, non è stabile: $\Lambda$ decade e il nucleo, espellendo un $\pi$, ritorna ad essere di tipo normale.

Le caratteristiche di questo ritorno alla normalità dipendono dalle proprietà dell'ipernucleo. Misurando l'energia del $\pi$, FINUDA ricostruisce le caratteristiche delle interazioni nucleari verificando i relativi modelli teorici. Il rivelatore FINUDA ha una struttura molto articolata: consiste in un apparato che misura le traiettorie dei K prodotti dalle $\Phi$, quindi in un bersaglio nucleare (un sottile strato di carbonio o di berillio o di litio, etc.), e infine in un sistema di rivelazione che misura il punto di creazione dell'ipernucleo con una precisione 0,1 mm e l'energia del $\pi$ espulso durante il suo decadimento con un errore dello 0,3%.

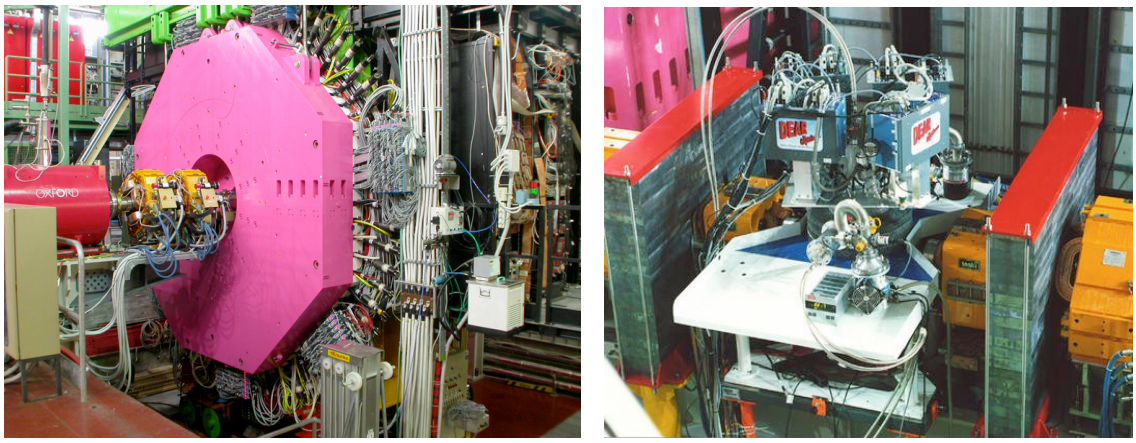


**Figura 1-10: A sinistra l'esperimento FINUDA, a destra l'esperimento DEAR**

#### 1.3.3.3 DEAR

L'obiettivo di DEAR (rif. Figura 1-10) è quello di produrre atomi esotici inesistenti in natura, sostituendo uno degli elettroni di un atomo ordinario con un $K^-$ prodotto dal decadimento di una $\Phi$ generata da DAFNE.

Lo scopo è di studiare le forze nucleari forti tra il nucleo atomico (un protone in questo caso) e il K. In genere questa forza non è apprezzabile, perché il suo effetto si riduce molto velocemente con l'aumentare della distanza. Per cui questa può essere studiata tramite una sonda (il $K^-$) molto vicina al nucleo. Il $K^-$ si avvicina gradualmente al nucleo, che lo cattura formando un atomo esotico. Dato che è instabile, il K precipita attraverso una serie di stati atomici fino a quando è assorbito dal protone a causa della forte interazione. Misurando l'energia dei raggi X emessi da queste transizioni, DEAR ricostruisce le caratteristiche dell'interazione forte tra il K e il protone.

L'apparato di DEAR consiste in un bersaglio di idrogeno gassoso alla temperatura di 25 °K e ad una pressione di 3 atmosfere e un rivelatore CCD (*Charge-Coupled Device*) per la rivelazione dei raggi X. La risoluzione energetica, circa 150eV, l'efficienza nella regione di energia di interesse, circa 6-10 keV, e soprattutto, la capacità di filtrare il rumore di fondo, fanno del CCD il miglior possibile rivelatore per questo tipo di misure.

Grazie all'elevato numero di K generati da DAFNE e alle sue caratteristiche, DEAR registrerà un numero sufficiente di eventi in pochi mesi di operazione.

#### 1.3.3.4 NAUTILUS

La forza gravitazionale è la più debole in natura. La teoria della relatività generale di Einstein descrive la gravità come una deformazione spazio-temporale indotta dalle concentrazioni di materia/energia. L'accelerazione dei corpi produce onde gravitazionali che si propagano alla velocità della luce. È molto difficile rivelare le onde gravitazionali, a causa della loro interazione debole con la materia e non esiste ancora un'evidenza sperimentale della loro esistenza.

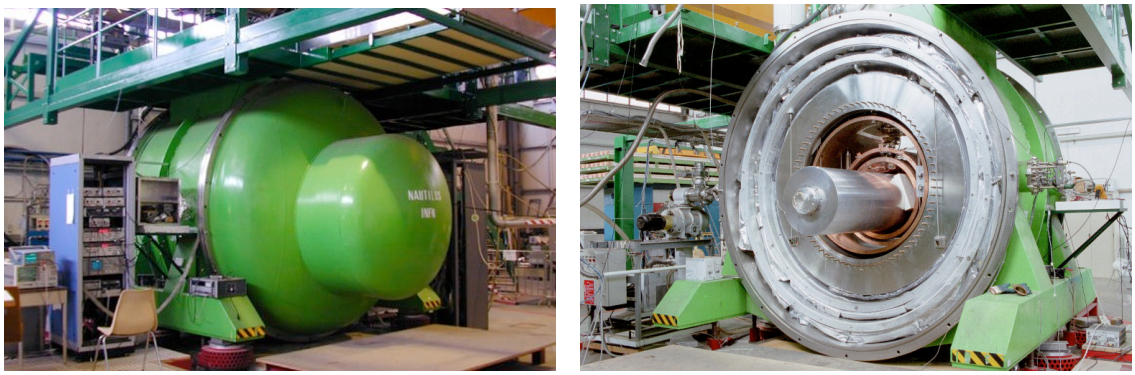


**Figura 1-11: NAUTILUS: a sinistra chiuso, a destra si nota l'antenna**

Inoltre è impossibile produrre onde gravitazionali con sufficiente intensità per la loro rivelazione in un laboratorio. Soltanto sorgenti astrofisiche con grandi masse possono produrre onde rivelabili. Tali sorgenti sono suddivise in tre categorie:

1. eventi catastrofici a breve vita, come l'esplosione di una supernova;
2. sorgenti continue, come le pulsar non a simmetria sferica;
3. sorgenti stocastiche quali quelle associate alla distribuzione galattica delle pulsar o le onde gravitazionali prodotte nei primi istanti di vita dell'universo.

L'arrivo di un'onda gravitazionale modifica la distanza tra due punti nello spazio, quindi un rivelatore di onde gravitazionali ideale consiste di due masse elementari posizionate in un sistema in grado di misurare la variazione della loro distanza nel tempo. Invece di masse elementari, possono essere usati corpi elastici continui e omogenei per studiarne le vibrazioni dei differenti elementi in relazione l'uno con gli altri.

Il rivelatore NAUTILUS consiste in un'antenna cilindrica di alluminio lunga 3 metri di massa pari a 2.300 kg, raffreddata ad una temperatura vicinissima allo zero assoluto (0,1 °K) allo scopo di ridurre il più possibile il rumore generato dall'agitazione termica degli atomi che la compongono (e ciò rappresenta un record mondiale).

L'antenna è in grado di vibrare all'arrivo di onde gravitazionali. Le vibrazioni sono convertite in segnali elettrici e quindi misurate. NAUTILUS può rivelare variazioni di $10^{-18}$ metri nella lunghezza dell'antenna. Questo significa che la sua sensibilità è tra le più alte mai raggiunte con questo tipo di rivelatori.

### 1.3.4 Altre attività

Le attività di ricerca dei Laboratori Nazionali di Frascati non si limitano agli esperimenti realizzati in casa, ma si estendono ad un gran numero di esperimenti operanti, o in via di realizzazione, in altri importanti laboratori sparsi in tutto il mondo.

I Laboratori Nazionali di Frascati partecipano a tutti gli aspetti degli esperimenti: disegno, progettazione, costruzione e realizzazione degli apparati sperimentali, operatività, analisi dei dati raccolti, etc..

Un ruolo particolarmente significativo è svolto nella realizzazione degli apparati sperimentali, grazie alla complessità e alla vastità di infrastrutture esistenti ai LNF.

Come molte altre sedi dell'INFN, anche i Laboratori Nazionali di Frascati sono fortemente coinvolti negli esperimenti di LHC al CERN di Ginevra; in modo particolare ATLAS e LHC-B e ALICE (già presentati nei precedenti paragrafi 1.2.3 e 1.2.4).

Tutte le altre attività di ricerca, non elencate in questo capitolo, non sono meno importanti dal punto di vista scientifico, ma semplicemente hanno minore rilevanza ai fini dei temi che vengono affrontati nel lavoro oggetto di questa tesi.

### 1.3.5 Organizzazione dei LNF

La struttura dei Laboratori Nazionali di Frascati è suddivisa in Divisioni e Unità funzionali. La differenza tra i due tipi di strutture è nelle dimensioni in quanto le Divisioni sono più grandi e complesse, le Unità funzionali più piccole e dedicate. C'è, infatti, un'ulteriore suddivisione delle Divisioni in Servizi. Le strutture di maggiori dimensioni dei LNF sono le tre Divisioni: Acceleratori, Ricerca e Tecnica

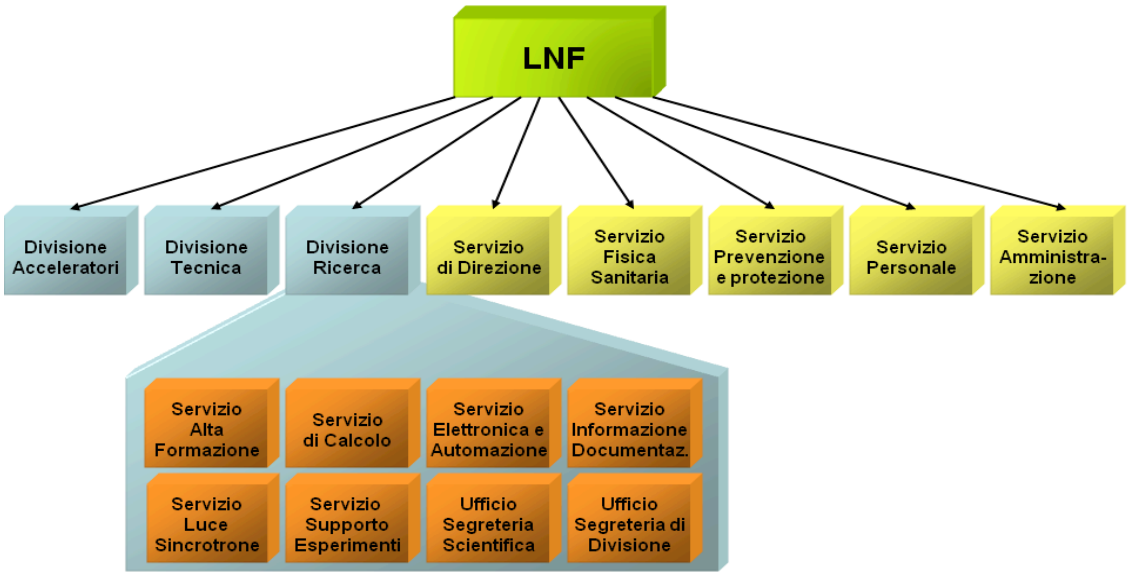



**Figura 1-12: Diagramma dell'organizzazione dei LNF**

Nella figura precedente (rif. Figura 1-12) viene riportato l'organigramma dei LNF. Il lavoro oggetto di questa tesi è stato svolto presso il Servizio di Calcolo della Divisione Ricerca.

Il Servizio di Calcolo e Reti dei LNF si occupa della configurazione e dell'amministrazione della rete di trasmissione dati, delle infrastrutture informatiche e delle risorse di calcolo dei LNF e dell'AC (Amministrazione Centrale). Inoltre svolge un ruolo particolarmente importante per tutto l'INFN in quanto gestisce alcuni importanti servizi informatici che hanno una valenza nazionale, anche se centralizzati ai LNF.

In particolare il Servizio di Calcolo gestisce:

- L'infrastruttura di rete:
  - ✓ Cablaggio strutturato in rame e in fibra ottica
  - ✓ Apparati di rete locale (switch di livello 2 e di livello 3)
  - ✓ Apparati di rete Wireless
  - ✓ La connessione alla rete geografica e il router di accesso
  - ✓ Apparati per la gestione della sicurezza informatica

- I seguenti servizi infrastrutturali per le funzionalità della rete:
  - ✓ DNS, DHCP
  - ✓ SMTP (Mail Relay)
  - ✓ IMAP
  - ✓ WebMail
  - ✓ WWW e WebProxyCache
  - ✓ Database (MySQL e Oracle DB)
  - ✓ Streaming audio e video
  - ✓ Authentication Authorization Infrastructure
  - ✓ Printing
  - ✓ Log recording
  - ✓ Monitoring system
  - ✓ PC Cloning e Deploying system

- La seguente infrastruttura di Storage e risorse di memoria di massa:
  - ✓ Storage Area Network
  - ✓ Andrew File System
  - ✓ Network Attached Storage
  - ✓ Tivoli Storage Manager
  - ✓ Librerie a nastro

- Le seguenti risorse di calcolo:
  - ✓ Computer centrale *axcalc* (IBM AIX)
  - ✓ Computer centrale *dxcalc* (Digital Compaq True 64)
  - ✓ Farm linux
  - ✓ Windows Domain

- Le risorse di calcolo di alcuni esperimenti ed in particolare Il sistema di calcolo basato su una Griglia Computazionale per l'esperimento ATLAS (Tier-2)

- Il Sistema Informativo per l'attività gestionale dell'INFN:
  - ✓ Gestione del Personale e Rilevazione delle Presenze
  - ✓ Gestione dei Documenti e "Workflow" dei Procedimenti

- Il servizio di hosting dei web server dell'INFN, dell'Amministrazione Centrale e dei LNF

In particolare le ultime 3 voci sono caratterizzate dal fatto di avere una valenza nazionale. In aggiunta a quanto sopra esposto, il Servizio di Calcolo fornisce il supporto per:

- Gli esperimenti che hanno una gestione autonoma delle proprie risorse di calcolo

- La configurazione e l'amministrazione di workstation e personal computer utilizzati da dipendenti, associati, dottorandi, laureandi, ospiti, servizi e/o esperimenti INFN

- L'uso delle risorse informatiche esportate e delle periferiche distribuite

### 1.3.6 Obiettivi di questo lavoro

L'obiettivo principale di questo lavoro è quello di realizzare un insieme di servizi informatici che forniscano il maggior numero di strumenti, ai gruppi di ricerca e a tutti gli utenti, per il raggiungimento dei traguardi scientifici primari dei Laboratori Nazionali di Frascati.

Il lavoro consiste nella collaborazione con il valido gruppo di informatici afferenti al Servizio di Calcolo dei LNF, per progettare le strutture informatiche di supporto alla ricerca e definirne le relative strategie implementative, a partire dal livello infrastrutturale, per arrivare ai livelli di servizio e di applicazione, passando per l'infrastruttura di rete e di mass storage; il tutto tenendo in particolare considerazione i seguenti importanti obiettivi:

- *Semplicità d'uso da parte dell'utenza*
- *Alta affidabilità*
- *Alta disponibilità*

# 2. L'infrastruttura di rete dei LNF

La comunità dei fisici si trova sempre ad affrontare problemi di calcolo per i quali occorrono reti di collegamento in grado di fornire prestazioni più avanzate di quelle generalmente disponibili. Anche la prossima generazione di esperimenti che si svolgeranno presso il CERN di Ginevra, grazie alla costruzione del nuovo acceleratore LHC, imporrà di gestire una quantità di dati enorme con cui non ci si è mai confrontati in passato: circa 10 Petabyte all'anno (1 Petabyte = $10^{15}$ byte) per la cui elaborazione sarà necessaria una grandissima capacità di calcolo. Per far fronte a queste formidabili necessità è stato lanciato il progetto GRID (griglia computazionale), il quale mira a costruire un insieme più sofisticato di servizi, anche se pur sempre dotati di un'interfaccia semplice ed intuitiva. Questo non dovrebbe limitarsi a consentire solo lo scambio di testi, immagini o filmati, come avviene sostanzialmente con il web, ma deve permettere di condividere grandi risorse di calcolo e di accedere a banche dati di dimensioni ingenti.

È evidente che per la realizzazione di tali ambiziosi progetti la rete svolge un ruolo di fondamentale importanza, sia per lo scambio dei dati di fisica tra i vari computer che compongono la griglia, sia per la comunicazione tra i fisici che compongono i gruppi di ricerca.

Occorre sottolineare che l'INFN storicamente, è stato pioniere nell'uso delle reti di trasmissione dati e nell'uso delle soluzioni e dei protocolli tecnologicamente più avanzati.

D'altra parte, l'esplosione dell'uso di internet è considerata uno degli eventi tecnologici più rivoluzionari mai avvenuti: internet ha prodotto un tale mutamento nei collegamenti tra le persone, nelle modalità di scambio di documenti e nell'accesso alle informazioni che oggi senza di esso è difficile immaginare di svolgere molte attività.

La rete di trasmissione dati, ai LNF, deve essere considerata un'infrastruttura sulla quale si basano tutti i servizi informatici a disposizione della comunità scientifica; per questa ragione deve soddisfare le esigenze di velocità e di banda necessarie per le attività di ricerca e contemporaneamente avere le caratteristiche di robustezza e affidabilità.

## *2.1 La rete locale*

La rete locale (LAN ovvero Local Area Network) copre praticamente l'intero territorio dei LNF, che è rappresentato in Figura 2-1.

La topologia della rete locale è a stella, o più precisamente ad albero, con centro stella principale, o radice, nell'edificio Calcolo (ed. N. 14)

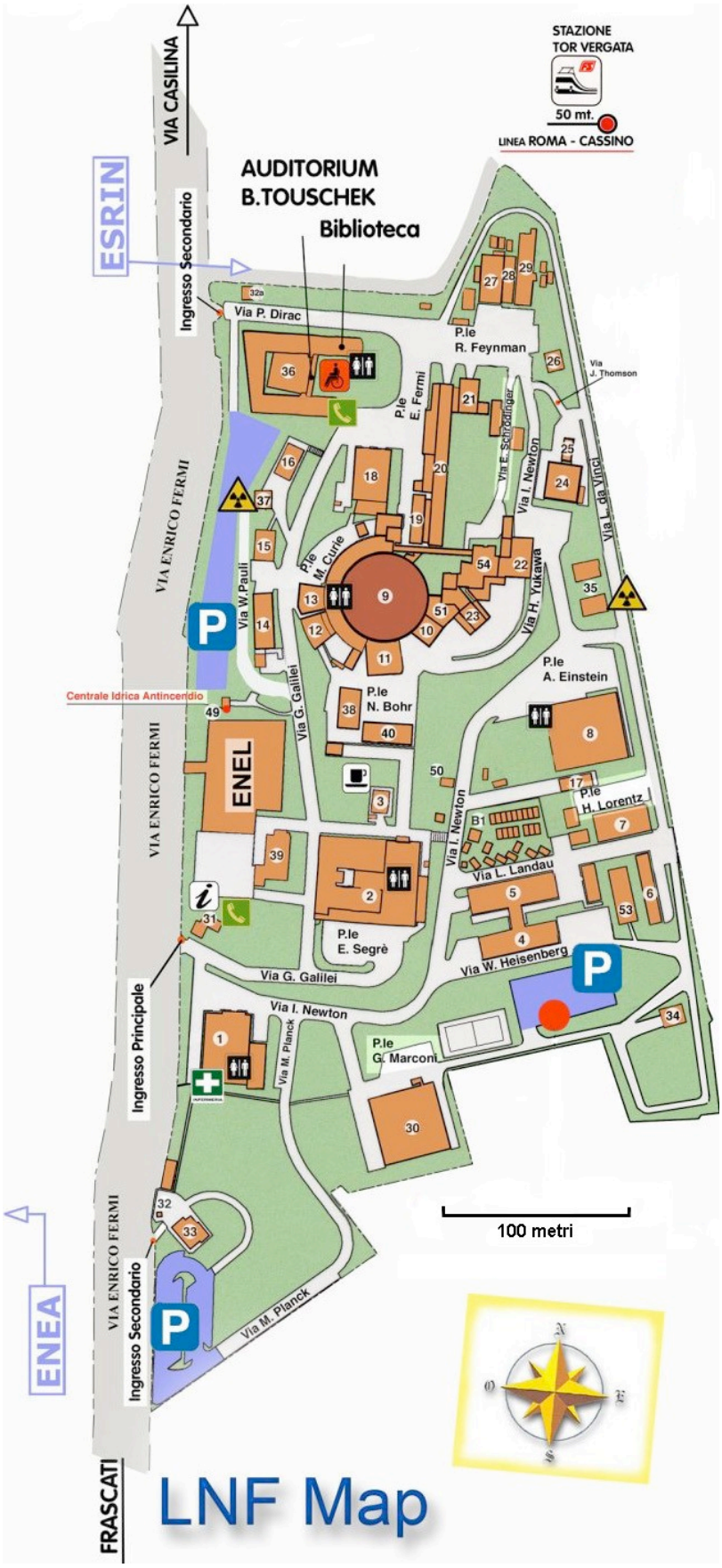



**Figura 2-1: Mappa del territorio dei LNF**

1. Direzione, Amministrazione
2. Divisione Acceleratori
3. Bar
4. Uffici PULS
5. Officine DA
6. Ex LISA
7. Laboratorio Tecnologie
8. Gran Sasso – Nautilus
9. DAFNE
10. Impianto Criogenico
11. Sala Sperimentale KLOE
12. DAFNE Luce
13. DAFNE Luce UV
14. Centro di Calcolo
15. Fisica Sanitaria
16. Servizi Generali, Sicurezze
17. Officina VIRGO – ROG
18. Sala Macchine Alimentatori
19. Laboratorio Impianti a Fluido
20. Sala Modulatori
21. LINAC
22. Fisica Nucleare – DEAR
23. Sala Accumulatore
24. Uffici distaccati AE
25. Lab. Produzione supercondut
26. Foresteria
27. Laboratorio FINUDA
28. Laboratorio LHCB
29. Lab. Divisione Ricerca
30. Amministrazione Centrale
31. Guardiania Principale
32. Guardiania pedonale
33. Foresteria
34. Box archivio A.C.
35. Deposito materiali
36. Edificio Alte Energie
37. Deposito sorgenti radioattive
38. Lab. Misure Magnetiche
39. Stazione Elettrica
40. Sala Pompe DAFNE

- Centrale idrica antincendio
- Box impresa pulizie
- Sala pompe accumulatore

53. Capannone deposito materiali
54. Sala controllo BTF

B1. Box uffici e magazzini

## 2.2 Protocolli di rete

Le caratteristiche principali che vengono richieste ad una rete di calcolatori sono:

1. Che sia di tipo *corporate,* perché una rete di calcolatori deve servire l'intero istituto in tutte le sue funzioni (per esempio alla progettazione, alla ricerca, all'amministrazione e alla gestione, etc.) e in tutte le sue sedi eventualmente distribuite sul territorio. Inoltre, tale rete deve essere collegata efficientemente con le reti di altri istituti di ricerca nazionali ed esteri o che hanno con essa frequenti rapporti interaziendali.
2. Che sia *multiprotocol*, perché è illusorio pensare di riuscire ad imporre all'interno di una azienda un'unica architettura di rete. Infatti occorre considerare che le reti sono nate all'interno delle aziende non con un processo progettuale "top-down", bensì con un'integrazione di tipo "bottom-up" in cui reti diverse, eterogenee, nate per risolvere problemi specifici, sono state a poco a poco integrate per formare una rete aziendale. Tale situazione si complica ulteriormente tutte le volte che si verificano fusioni interaziendali in cui occorre fondere anche sistemi informativi eterogenei. In letteratura tale problema è anche noto con il termine *internetworking*.

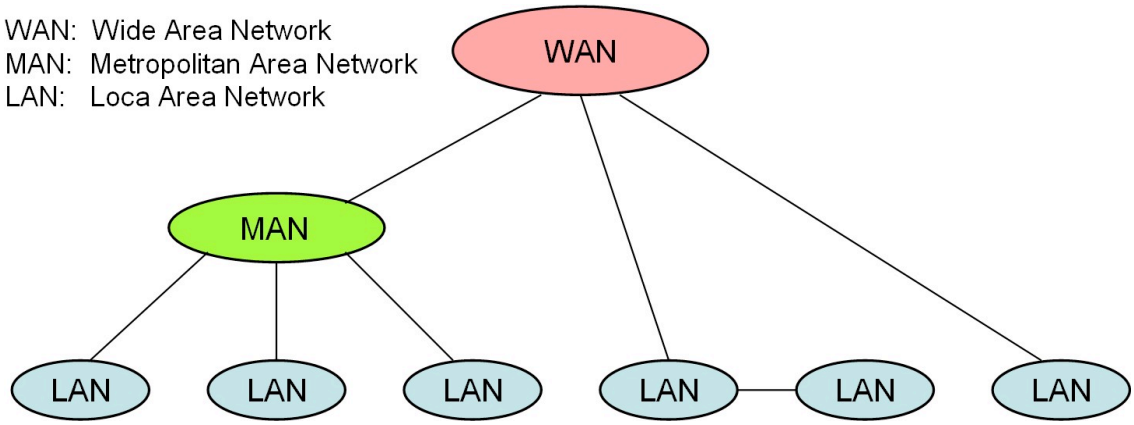



**Figura 2-2: Struttura di una rete di calcolatori**

### 2.2.1 Gli standard

- **CCITT** (*Comitè Consultatif International de Telegraphie et Telephonie*)
  È l'organismo internazionale che emette le specifiche tecniche che dovevano essere adottate dalle PTT.

- **ISO** (*International Standard Organization*)
  è il principale ente di standardizzazione internazionale che si occupa anche di reti di calcolatori.

- **ANSI** (*American National Standards Institute*)
  è il rappresentante USA nell'ISO.

- **UNINFO**

è il rappresentante italiano nell'ISO per le tematiche di reti di calcolatori.

- **IEEE** (*Institute of Electrical and Electronics Engineers*)
  è l'organizzazione professionale mondiale degli ingegneri elettrici ed elettronici con gruppi di standardizzazione sulle reti di calcolatori.

L'**OSI** (*Open System Interconnections*) è un progetto di ampio respiro formulato dall'ISO alla fine degli anni '70 con lo scopo principale di fungere da modello di riferimento per le reti di calcolatori. Per gestire la complessità dei problemi, l'OSI ha adottato un approccio a livelli (*layers*): l'intero problema della comunicazione tra due applicazioni è stato spezzato in un insieme di sette livelli, ciascuno dei quali esegue funzioni ben specifiche (rif. paragrafo A.1 in Appendice).

OSI si prefiggeva di essere molto più di un importante modello di riferimento. Infatti l'ISO ha standardizzato per OSI una serie di protocolli, da inserirsi ai vari livelli del modello, per formare una vera e propria architettura di rete concorrenziale con altre quali SNA, DECnet o TCP/IP.

Nel processo di standardizzazione, OSI è partito dai livelli bassi (quelli più vicini all'hardware) ed è salito verso quelli alti (quelli più vicini all'uomo) ricevendo gradimento ed accettazione differenti.

I livelli 1 (fisico) e 2 (*data link*) di OSI sono oggi assolutamente standard e questo consente l'interoperabilità dei prodotti. Dal livello 3 al livello 7 i protocolli esistono da tempo, ma non sono riusciti ad imporsi per l'alto impatto che la loro adozione ha sul software dei sistemi informativi stessi e sui dispositivi di instradamento (*router*).

Il progetto **IEEE 802,** perfettamente inserito nel modello OSI, riguarda i livelli 1 e 2 limitatamente alle reti locali e metropolitane. Concepito anch'esso tra la fine degli anni '70 e l'inizio degli anni '80, ha portato ad una voluminosa serie di standard noti con sigle del tipo 802.X, oggi anche approvati dall'ISO (rif. paragrafo A.2 in Appendice).

IEEE 802 è nato per razionalizzare i numerosi sforzi presenti in quegli anni per la creazione di nuove reti locali, spesso appositamente concepite – per ragioni commerciali – per essere incompatibili una con l'altra, ed ha ottenuto un notevole successo.

L'ISO non può affrontare autonomamente il problema della standardizzazione dei livelli 1 e 2 del modello OSI per le reti geografiche. A tale scopo si appoggia al **CCITT** che a livello 1 utilizza standard consolidati quali RS-232 (o gli equivalenti V.24 e V.28), V.35 e G.703/704, mentre a livello 2 adotta ad esempio una famiglia di standard derivati dal protocollo SDLC (*Synchronous Data Link Control*) proposto da IBM per la rete SNA.

SDLC stesso non viene riconosciuto come standard, ma alcune sue importanti varianti sì, quali HDLC, LAP-B, LAP-D e LAP-F. Inoltre, una variante di HDLC denominata LLC (*Logical Link Control*) viene adottata dall'IEEE per le reti locali con la sigla 802.2.

Gli anni '90 sono stati caratterizzati dalla comparsa di standard quali l'EIA/TIA 568 e 569 e il successivo ISO/IEC 11801 sul cablaggio strutturato degli edifici che verrà descritto nei paragrafi successivi. Tali standard regolamentano la progettazione e realizzazione degli impianti per il trasporto dei segnali da effettuarsi contestualmente alla costruzione o alla ristrutturazione organica di un edificio.

Il cablaggio strutturato è il veicolo preferenziale per il trasporto dei dati delle reti locali (LAN). Accanto alle due reti locali "storiche" Ethernet e Token Ring si sono aggiunte tutte quelle comprese nel progetto IEEE 802 e altre ancora che sono state standardizzate da altri enti (ad esempio l'ANSI standard FDDI).

Una rete locale è un mezzo di trasporto equamente condiviso tra tutte le stazioni che vi si collegano, ad alta velocità e basso tasso di errore, limitato ad un ambito locale (senza attraversamento di suolo pubblico). Le velocità trasmissive sono comprese nell'intervallo 4 Mb/s - 1000 Mb/s (ed oltre, oggi anche 10Gb/s).

Le reti metropolitane, nate dallo sforzo di standardizzazione congiunto tra ISO e CCITT, sono estensioni delle reti locali in ambito urbano. All'interno di una città, infatti, si può disporre spesso di dorsali in fibra ottica, veloci ed affidabili.

Le reti geografiche si basano sui servizi offerti dal fornitore nazionale di telecomunicazioni. In Italia, ad esempio, la trasmissione dati è nata con i CDA (Canali Diretti Analogici) i quali sono stati sostituiti nel tempo con i CDN (Canali Diretti Numerici) forniti dalla Telecom Italia. Le velocità di tali canali attualmente variano dai 2400 b/s ai 2 Mb/s. Con la liberalizzazione del mercato, oggi anche altri *provider* sono in grado di fornire servizi di rete geografica.

Sono inoltre state realizzate reti pubbliche per la sola trasmissione dei dati quali quelle conformi allo standard X.25 (in Italia ITAPAC), o successivamente Frame-Relay, SMDS (*Switched Multi-Megabit Data Service*) e ATM (*Asynchronous Transfer Mode*), che sono concepite per trasmissione dati a velocità rispettivamente basse (64 Kb/s-2 Mb/s), medie (2 Mb/s-34 Mb/s) e alte (155 Mb/s-622Mb/s e oltre).
Le più recenti soluzioni sono basate su POS (*Packet Over Sonet*) e WDM o DWDM (*Dense Wavelenght Division Multiplexer*) su fibra ottica monomodale (le velocità in quest'ultimo caso possono arrivare a 40 Gb/s).

La struttura di rete *corporate* e *multiprotocol* precedentemente descritta implica problematiche di *internetworking* tutte le volte che ci si trova a collegare due LAN tra di loro, una LAN con una MAN, una LAN con una WAN, etc..

L'*internetworking*, a causa dell'eterogeneità delle architetture di rete già citate, deve essere necessariamente multiprotocollo. Questo significa che la stessa struttura fisica, sia locale sia geografica, può essere utilizzata simultaneamente per il trasporto di più protocolli di rete (ad esempio TCP/IP, NetBeui, AppleTalk, etc.) i quali devono convivere il più armoniosamente possibile.

In generale, è opportuno cercare di minimizzare gli investimenti creando sinergie tra tutte le strutture atte a trasportare informazione. Ad esempio, a livello di rete geografica si cerca sempre più di far convivere sugli stessi mezzi trasmissivi non

solo i vari protocolli delle reti di calcolatori, ma anche le comunicazioni telefoniche e le videoconferenze tra sedi distinte della stessa azienda.

## 2.3 Cablaggio strutturato

La stella è una topologia interessante perché permette di precablare in modo strutturato gli edifici. La topologia stellare implica la presenza di un centro stella che può divenire un punto critico per l'affidabilità della rete, ma d'altro canto semplifica moltissimo la gestione e la manutenzione della rete stessa permettendo l'esclusione di sistemi malfunzionanti.

Molto spesso la stella è in realtà una stella gerarchica e quindi, più propriamente, un albero. Nelle stelle e negli alberi le problematiche di instradamento sono semplici poiché esiste un solo cammino che collega due sistemi.

La stella si realizza impiegando due mezzi trasmissivi punto-punto (doppino o fibra ottica) per interconnettere ogni sistema al centro stella (uno dal centro stella verso il sistema e l'altro in direzione opposta).

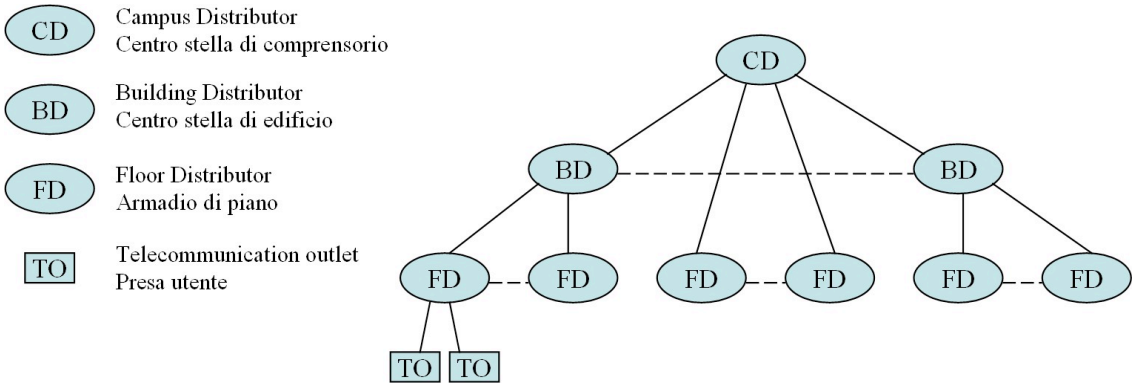



**Figura 2-3: Topologia a stella gerarchica**

### 2.3.1 Le fibre ottiche

Il vetro, se stirato a dimensioni micrometriche, perde la sua caratteristica di fragilità e diventa un filo flessibile e robusto. Una fibra ottica si presenta come un sottile filo di materiale vetroso costituito da due parti: la più interna prende il nome di nucleo (*core*), e l'esterna di mantello (*cladding*).

Il *core* ed il *cladding* hanno indici di rifrazione diversi, ed il primo è più denso del secondo. La differenza negli indici di rifrazione determina la possibilità di mantenere la luce totalmente confinata all'interno del core.

Il grande successo delle fibre ottiche è dovuto a diversi fattori tra cui:

- totale immunità da disturbi elettromagnetici,
- alta capacità trasmissiva: sono operative fibre ottiche a 10 e a 40 Gb/s,
- bassa attenuazione: alcuni decimi di dB/km,
- dimensioni ridottissime e costi contenuti.

Per contro, le fibre ottiche sono unicamente adatte a collegamenti punto-punto, non essendo possibile prelevare o inserire il segnale in un punto intermedio, cosa invece possibile con mezzi trasmissivi elettrici.

Le proprietà e i modi di propagazione dell'energia luminosa in una fibra ottica possono essere studiati mediante la teoria delle guide d'onda. Un'analisi semplificata, ma precisa sino a quando le dimensioni della fibra sono molto maggiori di quelle della lunghezza d'onda, può essere effettuata applicando le leggi dell'ottica geometrica.

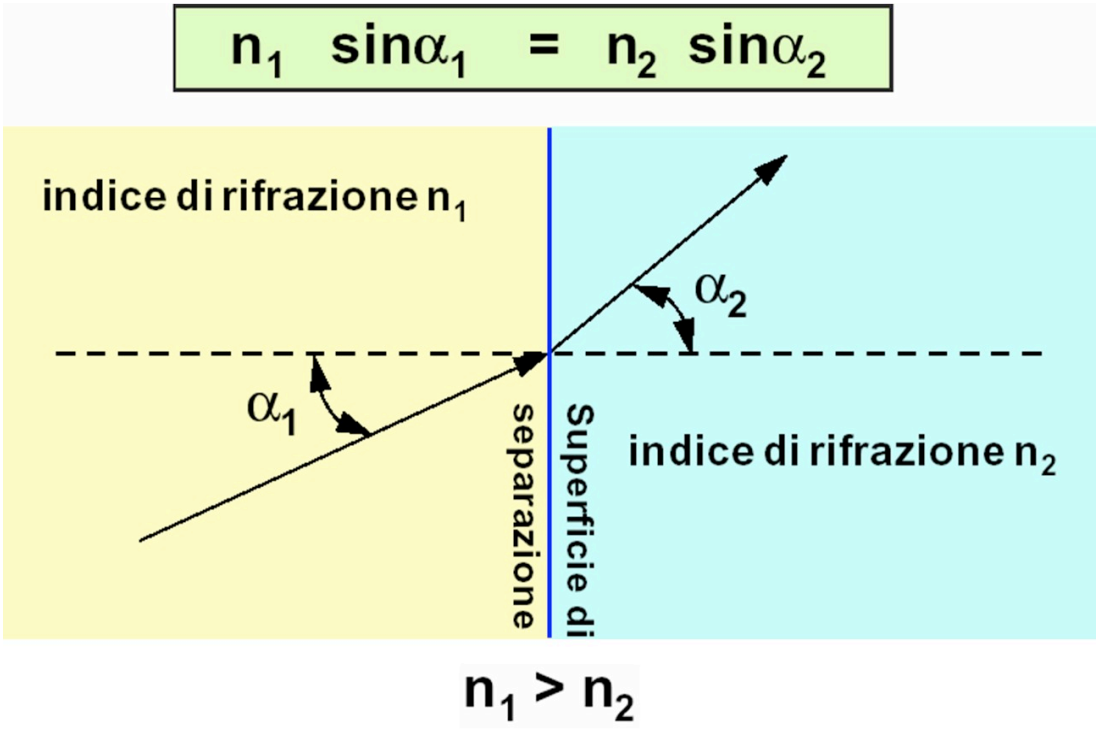



**Figura 2-4: Legge di Snell**

La legge di Snell in particolare, studia la riflessione e la rifrazione di un raggio luminoso incidente sulla superficie di separazione di due materiali (rif. Figura 2-4).

Essa dimostra che per valori dell'angolo di incidenza superiori a

$$\alpha_c = \text{arcsen}(n2/n1)$$

detto angolo critico, si ha riflessione totale.

Nelle fibre ottiche valori tipici per gli indici di rifrazione sono n2=1.475 per il cladding e n1=1.5 per il core. Pertanto, $\alpha_c$ = 79.5 gradi (rif. Figura 2-5).

Affinché fra il core e il cladding avvenga la riflessione totale dei raggi luminosi è necessario che essi siano introdotti ad una estremità ottica entro un certo angolo di accettazione della fibra.

Tanto maggiore sarà l'angolo di accettazione tanto più alta sarà la cosiddetta apertura numerica (NA) della fibra, cioè la quantità di luce che si riesce ad introdurre. Con i valori dell'esempio precedente risulta NA = 0.18.

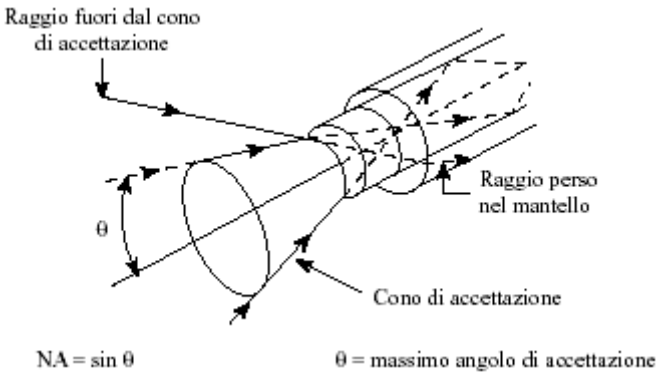



**Figura 2-5: Cono di accettazione**

Un altro parametro delle fibre ottiche estremamente importante è l'attenuazione. Essa può essere espressa in funzione della lunghezza d'onda, ottenendo un grafico simile a quello di Figura 2-6.

Vi si individuano tre minimi di attenuazione in corrispondenza di tre intervalli di lunghezza d'onda, detti finestre.

Le finestre corrispondono a tre tipi di utilizzazioni diverse:

- per la prima si usano solo LED comuni,
- per la seconda LED comuni e laser,
- per la terza solo laser.

Le lunghezze d'onda che interessano le comunicazioni ottiche sono quelle comprese tra i 750 nm ed i 1600 nm, cioè nel vicino infrarosso, in quanto le radiazioni visibili all'occhio umano vanno dai 455 nm (violetto) ai 750 nm (rosso).

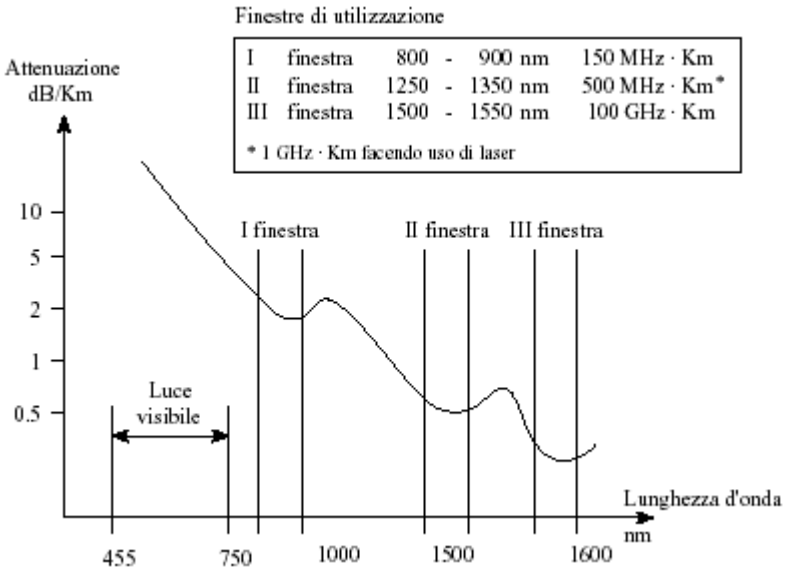



**Figura 2-6: Finestre di utilizzo**

La prima finestra è collocata intorno agli 850 nm ed è stata la prima ad essere usata per la realizzazione di sistemi di trasmissione su F.O. (presente solo nella fibra multimodale).

La seconda finestra è posta a 1300 nm ed essendo caratterizzata da una attenuazione inferiore, è quella attualmente utilizzata per esigenze di bande passanti medie o alte (presente sia nella fibra multimodale sia in quella monomodale). La banda passante varia in funzione del tipo di fibra e del tipo di emettitore/ricevitore utilizzato, e può essere di:

- 500 MHz · km, se si usano i LED su fibra multimodale;
- 1 GHz · km, se si usano i laser su fibra multimodale;
- da decine a centinaia di GHz · km su fibra monomodale, a seconda del laser utilizzato.

La terza finestra si colloca a 1550 nm, dove l'attenuazione è ancora inferiore ed è presente solo nella fibra monomodale.

Ai LNF il *Campus Distributor* (centro stella di comprensorio) si trova nell'edificio N° 14 (edificio Calcolo); da qui, la connessione ai vari *Building Distributor* è realizzata con cavi in fibra ottica.

I cavi in fibra ottica scelti per la realizzazione delle dorsali contengono generalmente 12 fibre ottiche multimodali 62,5/125 μm (6 coppie). Raramente il numero delle fibre è inferiore (8 o 6), nei casi in cui gli edifici da servire sono molto piccoli oppure con poca utenza. Nelle più recenti installazioni i cavi utilizzati contengono fibre ottiche multimodali 50/125 μm.

Dovendo percorrere anche cavidotti interrati nei giardini esterni agli edifici, tutti i cavi in fibra ottica utilizzati sono armati da esterno "*Low smoke and fume*" e "*0-Halogen*", antiroditori e con protezione antiumidità. Inoltre, nei percorsi non protetti, i cavi sono contenuti in apposita tubazione in PVC, con grado di protezione IP55.

L'attestazione delle fibre avviene su appositi pannelli di distribuzione, sempre montati su armadi standard (rack da 19 pollici); il connettore scelto come standard è quello con innesto a baionetta di tipo ST. Al termine della fase di installazione tutte le fibre vengono certificate con apposito strumento (*Optical Time Domain Reflectometer*) in grado di vedere le perdite di segnale luminoso durante il percorso. Nelle installazioni già realizzate il valore massimo di attenuazione totale sulla singola fibra ottica si aggira intorno ai 3 o 4 decibel.

Fino ad oggi non si è sentita l'esigenza di installare fibre ottiche monomodali (9/125 μm). Questo è dovuto al fatto che la distanza tra l'edificio Calcolo e gli edifici periferici non supera mai i 500 metri. Con tali distanze in gioco si riesce ad ottenere un'alta banda trasmissiva (dell'ordine di decine di gigabit/s) anche su fibre ottiche multimodali, utilizzando *transceiver* a tecnologia laser che lavorano in seconda finestra ($\lambda$=1300 nm).

Ciò determina anche un vantaggio in termini economici in quanto la realizzazione di un'infrastruttura ottica basata su fibre ottiche monomodali è notevolmente più costosa.

### 2.3.2 La distribuzione orizzontale (cablaggio in rame)

I mezzi trasmissivi elettrici rappresentano ancora oggi il mezzo più diffuso, e nell'ambito delle reti locali assumono fondamentale importanza soprattutto per la realizzazione di infrastrutture per la trasmissione di segnali all'interno degli edifici. Dovendo trasportare il segnale in forma di energia elettrica, è necessario che le caratteristiche elettriche del mezzo siano tali da rendere massima la trasmissione dell'energia da un estremo all'altro e minima la dissipazione in altre forme e la forma d'onda resti il più possibile inalterata.

La sezione dei conduttori può essere espressa come misura del diametro in millimetri (valori tipici 0.4 - 0.7 mm), ma questa soluzione è poco usata. Molto più diffusa è l'unità di misura detta AWG (*American Wire Gauge*).

L'AWG è una scala a regressione geometrica con 39 valori compresi nell'intervallo da 0 gauge (0.460 pollici di diametro) a 36 gauge (0.005 pollici di diametro); ogni incremento di un gauge corrisponde ad un rapporto tra i diametri di

$$(0.460/0.005)^{1/39} \approx 92^{1/39} \approx 1{,}229322$$

Avere un basso AWG, e quindi diametro elevato, è un parametro di merito, in quanto diminuisce la resistenza e quindi la potenza dissipata sul cavo.
I diametri dei cavi comunemente usati per la trasmissione dati sono compresi tra 26 AWG (doppini per sola telefonia) e 22 AWG (cavo di tipo 1 IBM).

I materiali isolanti usati nella costruzione dei cavi possono essere di due tipi: compatti o espansi. La scelta determina notevoli differenze nella costante dielettrica, che per l'isolante di un cavo è tanto migliore quanto più vicina a quella dell'aria.

Gli isolanti espansi (che contengono aria) sono migliori di quelli compatti, ma presentano due gravi inconvenienti: sono estremamente infiammabili, in quanto contengono sia il combustibile (plastica) che il comburente (aria), e sono più voluminosi, rendendo maggiori le dimensioni dei cavi.

Per queste ragioni ormai per quasi tutti i cavi si usano isolanti compatti, molto più sottili e che presentano, in caso d'incendio, un'emissione di fumi limitata e non tossica. Quest'ultimo aspetto è fondamentale in quanto i cavi per trasmissione dati devono sottostare a normative per la sicurezza in caso di incendio.

Esistono principalmente due tipi di cavi che possono essere utilizzati: non plenum e plenum.

I cavi di tipo plenum (per ora usati solo negli Stati Uniti) hanno la proprietà di resistere ad alte temperature, poiché sia il materiale isolante sia la guaina esterna sono in teflon, non propagano l'incendio e non bruciano, ma nel caso peggiore si carbonizzano emettendo gas tossici.

I cavi di tipo non plenum sono quelli più usati e, a seconda del materiale costituente la guaina esterna, hanno caratteristiche diverse:

- *flame retardant*: ritardano la propagazione della fiamma;
- *low smoke fume (LSF)*: bassa emissione di fumi in caso d'incendio;
- *zero halogen (0H)*: assenza di emissione di gas tossici.

I circuiti elettronici funzionano generalmente controllando correnti o tensioni rispetto ad un unico conduttore di ritorno (per le correnti) o di riferimento (per le tensioni). Le prime tecniche di trasmissione dei segnali digitali erano basate sul medesimo principio: si portava il riferimento di tensione da trasmettitore a ricevitore tramite un conduttore, e il segnale (o i segnali) su un altro conduttore (o su più d'uno). Questa tecnica di trasmissione è detta sbilanciata (*longitudinal mode*).

Poiché il conduttore che trasporta il segnale si comporta da antenna nei confronti dei campi elettromagnetici in cui è immerso e la corrente indotta nel conduttore si somma a quella del segnale, rendendone difficile o impossibile la decodifica, per tale tipo di trasmissione si impiega spesso il cavo coassiale. Il foglio o la calza metallica che avvolge il cavo coassiale svolge la triplice funzione di conduttore per il ritorno della corrente del segnale, riferimento di tensione e gabbia di Faraday, cioè schermatura, per il conduttore interno.

La tecnica alternativa alla trasmissione sbilanciata è la trasmissione bilanciata (*differential mode*). Essa consiste nell'utilizzare due conduttori perfettamente simmetrici (detti "coppia"), sui quali viene inviato lo stesso segnale elettrico, ma in opposizione di fase.

Il vantaggio rispetto alla trasmissione sbilanciata consiste nell'assenza della tensione di riferimento che deve essere identica per ricevitore e trasmettitore; il segnale è ricostruito per differenza delle tensioni presenti sui due conduttori della coppia.

Presupposto fondamentale per la trasmissione bilanciata è che i due conduttori siano perfettamente simmetrici rispetto a qualsiasi punto dello spazio, in modo da annullare sia l'emissione che la sensibilità ai disturbi elettromagnetici. La perfetta simmetria potrebbe essere raggiunta soltanto se i due conduttori coincidessero, cosa irrealizzabile ma approssimabile ritorcendo i due conduttori. Si realizza così il "doppino ritorto" (*twisted pair*, TP – rif. Figura 2-7).

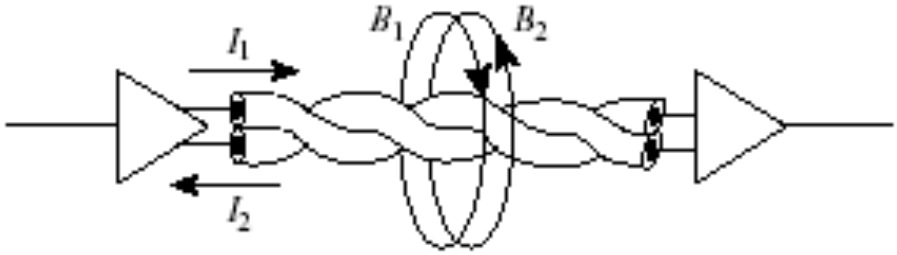



**Figura 2-7: Trasmissione bilanciata su doppino**

La trasmissione bilanciata su TP riduce le emissioni di disturbi elettromagnetici in quanto le correnti che attraversano i due conduttori sono di uguale intensità e verso opposto, e generano campi magnetici opposti che tendono ad annullarsi.

È in continua crescita l'attenzione al problema dei disturbi elettromagnetici, dei quali le reti locali sono al contempo vittime e sorgenti. Con la presenza di schermi e con una corretta messa a terra si possono ridurre drasticamente la sensibilità e l'emissione di disturbi elettromagnetici.

Esistono numerosi tipi di schermo, tra i quali i più utilizzati nelle LAN sono:

- foglio (*foil*): si tratta normalmente di un foglio di alluminio, molto sottile (da 0.05 mm a 0.2 mm) che avvolge il cavo immediatamente sotto alla guaina di protezione esterna. Poiché l'alluminio presenta elevata resistenza elettrica rispetto al rame e una notevole fragilità, lungo il foglio scorre un filo di rame nudo, detto *drain*.
- calza (*braid*): si tratta di una trecciola di fili di rame che avvolgono il cavo in due direzioni opposte. Presenta una conducibilità molto migliore del foglio di alluminio, ma la copertura non è completa, in quanto in corrispondenza degli intrecci rimangono inevitabilmente dei fori nello schermo. Inoltre, l'ossidazione dei fili di rame e la loro deformazione in fase di posa del cavo possono alterare l'efficacia della schermatura.

I migliori risultati si ottengono dalla combinazione di più schermi diversi, come foglio più calza, foglio più calza più foglio, e così via.

**Caratteristiche elettriche**

I parametri meccanici finora descritti determinano i parametri elettrici del cavo stesso:

- L'**impedenza** è il parametro elettrico più importante per un cavo usato ad alte frequenze. L'impedenza, normalmente indicata con il simbolo Z, è espressa in ohm (Ω) ed è la somma di due componenti ($Z = R + jI$) in quanto sintetizza in un solo valore resistenze, capacità ed induttanze presenti sul cavo. Ciò che interessa analizzare non è tanto il valore nominale di impedenza ad una data frequenza, ma il variare di tale valore al variare della frequenza. Più l'impedenza è stabile al variare della frequenza, migliore è il cavo, e la presenza di schermi normalmente migliora tale aspetto. Oggi si certifica l'impedenza dei cavi nell'intervallo da 100 KHz a 350 MHz.
- Si definisce **velocità** di propagazione $\mathbf{v_p}$ la percentuale della velocità della luce nel vuoto (circa $3 \cdot 10^8$ m/s) alla quale si propaga un segnale elettrico sul cavo. Per i cavi in rame $\mathbf{v_p}$ varia tra il 55% e il 75%. Questo implica una velocità di propagazione dell'informazione di circa 200.000 km/s. Anche se sembra una velocità elevata, va considerato che ad una velocità di trasmissione di 10 Mb/s, al termine del tempo dedicato alla trasmissione di un bit, il bit stesso ha percorso soltanto 20 m.
- Altro importante parametro elettrico è l'**attenuazione**, che per i mezzi elettrici è definita come rapporto, in dB, della tensione del segnale in ingresso al cavo e la tensione misurabile all'altra estremità. L'attenuazione così misurata cresce linearmente con la lunghezza del cavo e con la radice quadrata della frequenza.
- La **diafonia** (*cross-talk*) è invece la misura in dB di quanto un cavo disturba un altro cavo vicino. Spesso viene data come attenuazione di diafonia e quindi come parametro di merito (quanto è attenuato il segnale indotto da un cavo nel cavo vicino). In linea di principio esistono due modi diversi per misurare la diafonia: se la misura del segnale indotto nel cavo vicino è effettuata dalla stessa

parte del trasmettitore si parla di paradiafonia o NEXT (*Near End Cross-Talk*, se è effettuata all'estremo opposto si parla di telediafonia o FEXT (*Far End Cross-Talk*).

- Ai fini di una corretta ricezione non interessano tanto l'attenuazione assoluta del cavo o il suo valore di diafonia, quanto la combinazione di questi due parametri. Infatti, se si considera trascurabile il rumore indotto dall'esterno, è tale combinazione che determina il rapporto segnale/rumore in ingresso al ricevitore, e quindi l'integrità del segnale. Esiste un parametro che rappresenta le due grandezze in modo combinato: l'**ACR** (*Attenuation to Cross-talk Ratio*), che esprime il rapporto tra il segnale attenuato presente su una coppia ed il segnale indotto dalla coppia vicina. Esso varia in funzione della frequenza e della lunghezza del cavo.

Nella Figura 2-8 l'ACR è rappresentato dalla differenza (ovvero la distanza) tra i valori di attenuazione e di diafonia. Quando questa distanza è troppo ridotta non è

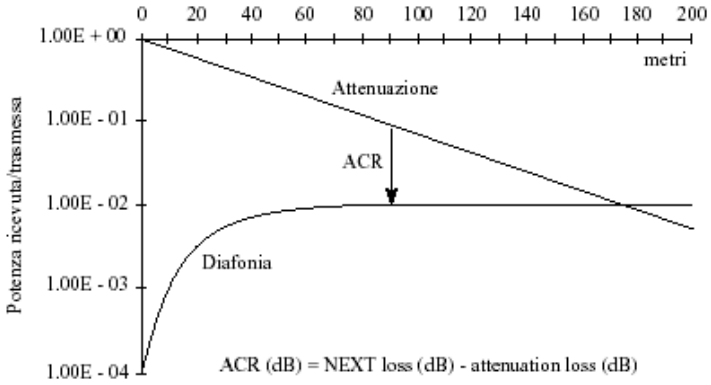



**Figura 2-8: ACR, attenuazione e diafonia**

più possibile trasmettere sul cavo in modo affidabile in quanto il segnale è troppo debole rispetto al rumore e quindi possono verificarsi troppi errori di trasmissione.

Il doppino è il mezzo trasmissivo classico della telefonia e consiste in due fili di rame ricoperti da una guaina isolante e ritorti (*twisted*) detti comunemente "coppia" (*pair*, da cui twisted pair o TP). Il tipo di doppino più usato attualmente ha un diametro di 24 AWG e un'impedenza di 100 Ohm.

Normalmente si utilizzano cavi con 4 coppie ed è allora necessario adottare passi di binatura differenziati da coppia a coppia per ridurre la diafonia tra le coppie. Infatti, se i passi di binatura fossero uguali, ogni conduttore di una coppia si troverebbe sistematicamente affiancato, ad ogni spira, con uno dei due conduttori dell'altra coppia, e quindi verrebbe a cadere l'ipotesi di perfetta simmetria della trasmissione bilanciata. I campi elettromagnetici generati dalle due coppie interferirebbero reciprocamente con un considerevole peggioramento della diafonia.

Esistono varie versioni di doppino:

- STP (*Shielded Twisted Pair*), versione con uno schermo per ogni coppia più uno schermo globale;
- FTP (*Foiled Twisted Pair*), versione con un unico schermo (normalmente in foglio di alluminio) per tutto il cavo;

- UTP (*Unshielded Twisted Pair*), versione non schermata.

I parametri elettrici di qualsiasi cavo variano al variare della frequenza. Occorre pertanto chiedersi, per una data applicazione, a quale frequenza sia opportuno analizzare i parametri per decidere se un cavo sia adeguato all'applicazione stessa.

Dovendo realizzare un'infrastruttura di trasmissione di segnali, e quindi installare cavi adatti a più applicazioni, sarebbe necessario considerare un elevato numero di valori dei parametri elettrici, a tutte le frequenze interessate dalle possibili applicazioni.

Per evitare questa operazione si è ricorso ad una classificazione dei cavi di uso più comune, cioè dei doppini. Tale classificazione, secondo lo standard Europeo ISO/IEC 11801 prevede 5 e più categorie, dette classi, in base alle applicazioni per le quali i cavi sono idonei:

- classe A adatta per applicazioni fino a 100 kHz;
- classe B adatta per applicazioni fino a 1 MHz;
- classe C adatta per applicazioni fino a 16 MHz;
- classe D adatta per applicazioni fino a 100 MHz.
- classe E adatta per applicazioni fino a 270 MHz.

La classe di connessione definisce le caratteristiche elettriche più importanti quali attenuazione, diafonia, ACR, riferite all'insieme di tutti i componenti passivi interposti tra due apparati attivi di telecomunicazione.

Ai LNF la distribuzione di piano (dai vari *Floor Distributor)* è realizzata con cablaggio strutturato basato sul sistema AMP ACO. Le prime realizzazioni sono state fatte utilizzando il cavo 24 AWG di classe D schermato con foglio di alluminio (FTP). Già da alcuni anni si sta utilizzando il cavo di classe E, avente schermatura distinta per singolo doppino.

I cavi utilizzati nelle realizzazioni di cablaggio strutturato contengono 4 doppini e generalmente ad ogni cavo corrisponde una connessione di rete (tipicamente Ethernet 10/100 Mbit/s). Tuttavia il protocollo Ethernet 10/100 Mbit/s prevede l'uso di due soli doppini, per cui due doppini per ciascun cavo rimangono inutilizzati.

Il sistema AMP ACO tende a massimizzare l'uso dei doppini per ciascun cavo: nel caso di funzionalità di tipo Ethernet, ad esempio, esso sfrutta tutti e 4 i doppini contenuti nel cavo, trasportando due connessioni di tipo Ethernet 10/100 Mbit/s dal pannello di distribuzione in rame verso le postazioni utente. Il sistema ACO infatti definisce una serie di connettori binati (Ethernet, FDDI, ISDN, etc.) che consentono di risparmiare nel numero di cavi da installare negli edifici. Tale sistema ha l'ulteriore vantaggio di essere molto versatile in quanto permette l'attestazione dei cavi su appositi connettori di terminazione (detti *edge connector*), sui quali si innestano molto semplicemente i connettori necessari per realizzare le funzionalità e protocolli che si vogliono distribuire. Inoltre, grazie alla facilità con cui è possibile sostituire o intercambiare i connettori, è molto semplice adattare il cablaggio alle reali necessità dell'utente finale.

Le scelte realizzative fatte ai LNF, prevedono il cablaggio di due cavi per ogni postazione utente (*telecommunication outlet*), che consentirebbero di portare fino a 4 connessioni Ethernet 10/100 Mbit/s per ogni postazione. Tuttavia la tipica presa utente prevede un connettore binato RJ45 per realizzare due connessioni Ethernet 10/100 Mbit/s e un connettore singolo, sempre RJ45, per realizzare una connessione Ethernet 10/100 Mbit/s o Gigabit Ethernet (quest'ultima prevede l'uso di tutti e 4 i doppini di un cavo). In tal modo si ottengono complessivamente 3 connessioni Ethernet RJ45 per ogni postazione utente (di cui una fino ad 1 Gigabit/s).

L'attestazione dei cavi in rame avviene su appositi pannelli di distribuzione, sempre montati su armadi standard (rack da 19 pollici). Al termine della fase di installazione tutte i cavi vengono certificati con apposito strumento (*Time Domain*

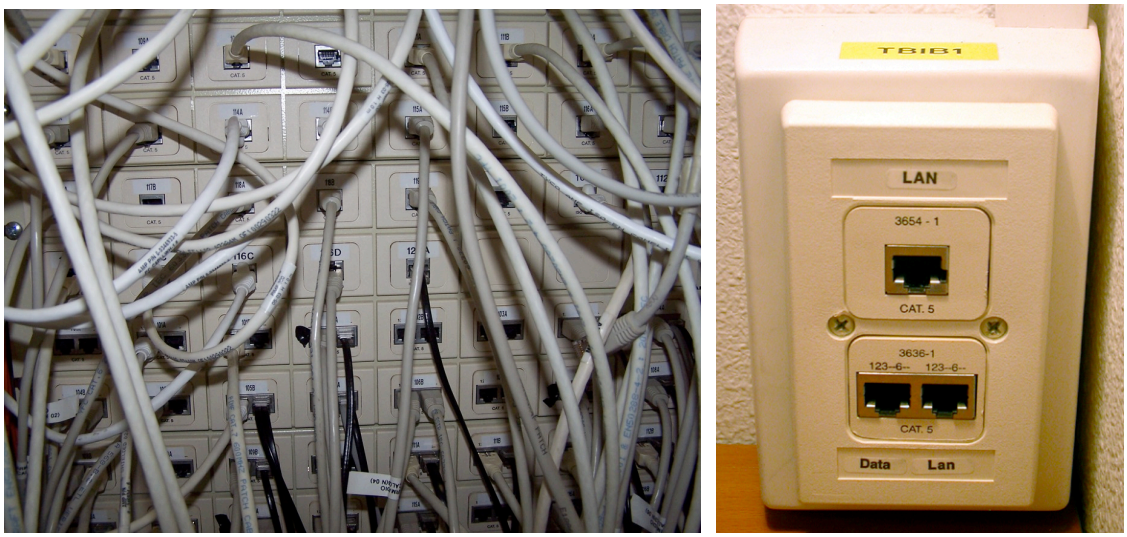



**Figura 2-9: A sinistra: pannello di distribuzione; a destra: presa utente**

*Reflectometer*) in grado di verificare la rispondenza dei cavi ai requisiti previsti dallo standard definito dalla classe di appartenenza del cavo stesso (rif. Figura 2-9).

Tutte le installazioni di cablaggio strutturato realizzate, ovvero il materiale fornito e le lavorazioni svolte, sia per quanto riguarda le dorsali in fibra ottica, sia per quanto riguarda la distribuzione di piano in rame, sono state realizzate in accordo con l'*AMP Warranty Program* che prevede un servizio di assistenza in garanzia per la durata di 15 anni dalla data d'installazione.

Ai LNF attualmente sono installate complessivamente quasi 1500 prese utente; dedicate ai client d'accesso al network. In tale numero non sono incluse le utenze dalle varie sale calcolo, in cui i server sono direttamente connessi agli switch con opportune bretelle ottiche o patch cord in rame sufficientemente lunghe.

## 2.4 Apparati attivi di rete

Già nel 1983, ai LNF, fu realizzata una delle prime reti locali del territorio italiano, basata sul protocollo Ethernet (su cavo coassiale, standard 10Base5). Da allora la rete si è evoluta molto, adeguandosi via via alle esigenze dei suoi utilizzatori e dei grandi esperimenti di fisica.

La storia dell'attuale generazione di apparati attivi inizia nel 2000, quando fu avviata una gara per l'acquisto di 3 apparati di *core* che avrebbero svolto funzioni di *switching*, in edifici particolarmente critici quali il Calcolo, DAFNE e Direzione.

Il capitolato tecnico prevedeva una serie funzionalità di cui le più importanti erano:

- Montaggio su rack standard da 19 pollici
- Alimentatori (AC 220V/50Hz) ridondabili e sostituibili a caldo
- Moduli di switching e di controllo ridondati e sostituibili a caldo
- Interfacce supportate (sostituibili a caldo):
  - Gigabit Ethernet 1000 BaseSX e 1000 BaseLX
  - Ethernet/FastEthernet 10/100/1000 BaseTX autosense
- Capacità complessiva e aggregata:
  - Backplane e switching fabric aventi capacità di switch $\geq$ 128Gb/s
  - Switching e routing throughput $\geq 10^8$ pacchetti/s
- Supporto di aggregazione EtherChannel verso altri switch o verso server:
  - $\geq$ 4 canali FastEthernet
  - $\geq$ 4 canali Gigabit Ethernet
- Gestione delle Virtual LAN basate su politica di porta o di nodo
- Supporto di Multilayer Switching (per il protocollo IP unicast e multicast)
- Gestione del protocollo di spanning tree
- Gestione dei protocolli di routing (RIPv1, RIPv2, OSPF, BGP4)
- Gestione del Policy-based routing
- Gestione delle politiche di security (ACL e filtri di livello 2 e superiori)
- Gestione del Quality of Service (standard IEEE 802.1Q, 802.1p)
- Gestione del Network Address Translation e del Server Load Balancing
- Gestione e supporto delle Jumbo-Frame
- Gestione dei protocolli per il management (SNMP agent e RMON agent)

Alla gara parteciparono le ditte Cisco, Enterasys e Nortel e vinse la ditta Cisco per aver presentato l'offerta economicamente più vantaggiosa. Fornì all'istituto 3 Catalyst 6500 con le ridondanze e la densità delle porte richieste nel capitolato.

Da allora e fino ad oggi, il Servizio di Calcolo ha continuato ad espandere la nuova infrastruttura di rete, acquistando molti altri apparati, ma per scelta strategica, tutti di marca Cisco.

Sono stati acquistati Catalyst delle famiglie 6000, 4000, 3500 e 2900, tutti facilmente integrabili nella nuova infrastruttura di rete che si andava delineando.

La scelta di acquistare tutti apparati della stessa marca e famiglia, presenta una serie di notevoli vantaggi, tra i quali:

- Maggiore facilità di integrazione e di interoperabilità
- Gestione attraverso Sistemi software unici e centralizzati
- Possibilità di utilizzo di protocolli proprietari (non definiti negli standard)

Quest'ultimo punto rappresenta forse il vantaggio più grande. Infatti gli standard vengono definiti per rispondere all'esigenze dell'utenza, ma spesso l'obiettivo non viene centrato, o almeno non pienamente; inoltre frequentemente arrivano troppo tardi, sicuramente dopo delle soluzioni proprietarie che il mercato è in grado di fornire in risposta a dette esigenze.

Ovviamente, l'uso di soluzioni proprietarie, come già anticipato, vincola all'acquisto di apparati omogenei, ovvero tutti della stessa casa costruttrice. E questo si potrebbe pagare in termini economici.

Dal 2000 ad oggi sono stati installati oltre 130 apparati Cisco, di cui circa 90 sono apparati di switching per la distribuzione dei servizi di rete attraverso il cablaggio strutturato a copertura dell'intero territorio dei LNF, mentre più di 40 sono apparati di tipo *Wireless*, installati prevalentemente nelle aule adibite alle conferenze o alle riunioni, per garantire l'accesso a tutti i partecipanti, inclusi i visitatori occasionali.

Lo schema di interconnessione degli apparati di switching è ad albero (rif. Figura 2-10), con radice nell'edificio calcolo. Ogni switch è montato in apposito armadio standard, lo stesso dove sono attestati i pannelli di distribuzione ottica e i pannelli di distribuzione in rame descritti nei paragrafi precedenti. In tal modo con semplici bretelle ottiche di permutazione è possibile realizzare l'interconnessione tra i vari switch, mentre con semplici cavi di permutazione in rame è possibile distribuire i servizi di rete fino alle postazioni utente.

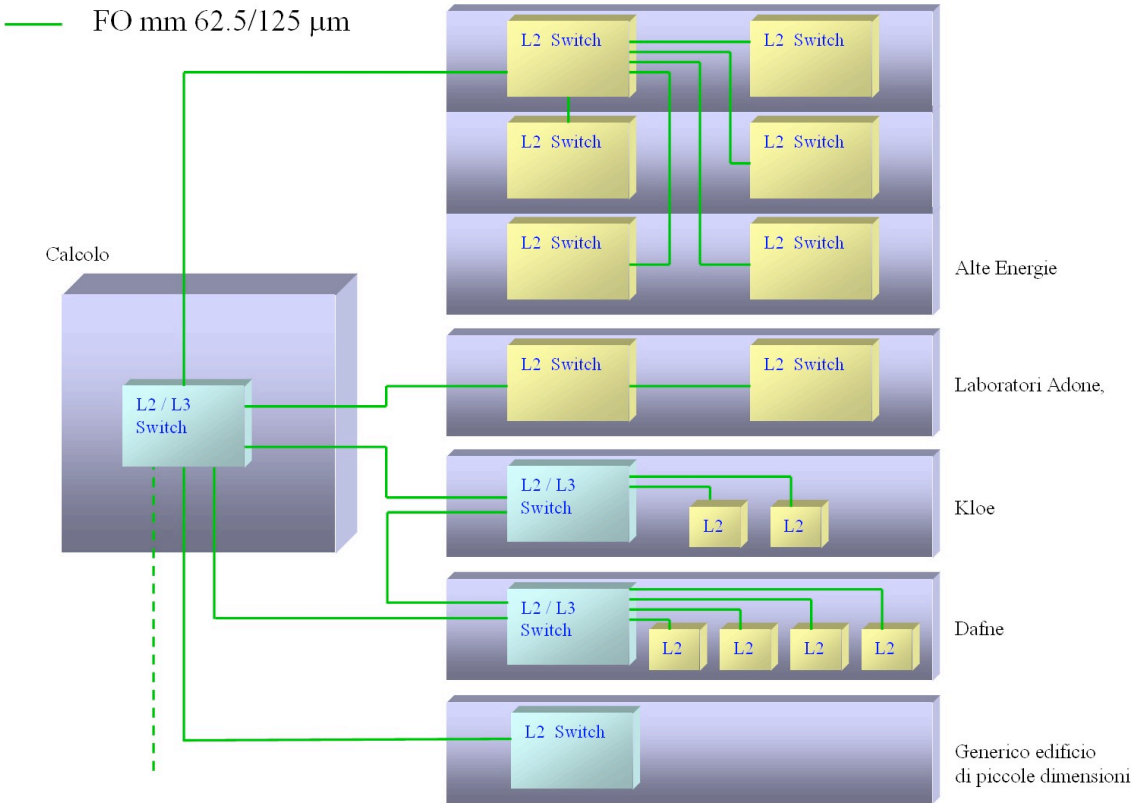



**Figura 2-10: Schema di interconnessione degli apparati di switching**

Oltre 90 apparati di switching complessivamente servono circa 4000 connessioni di tipo FastEthernet o Gigabit Ethernet. Contando sia i server che i client d'accesso alle risorse informatiche, alla rete locale sono stabilmente connessi circa 2000 nodi di rete (*end node*).

## 2.5 Soluzioni tecniche implementate ai LNF

Il grande numero di *end node* connessi stabilmente alla rete locale dei LNF, impone di avere un adeguato range di indirizzamento di terzo livello (ovvero IP). I range di indirizzi internet pubblici attribuiti ai LNF sono rappresentati in tabella:

| Range | Aggregabilità | Numero di indirizzi |
|---|---|---|
| 192.84.128.0 – 192.84.131.255 | 192.84.128.0/22 | 1024 |
| 193.206.80.0 – 193.206.87.255 | 193.206.80.0/21 | 2048 |
| 192.135.25.0 – 192.135.25.255 | 192.135.25.0/24 | 256 |
| 192.135.26.0 – 192.135.26.255 | 192.135.26.0/24 | 256 |
| 193.205.228.0 – 193.205.228.255 | 193.205.228.0/24 | 256 |

**Tabella 2-1: Range di indirizzi internet pubblici attribuiti ai LNF**

Tali indirizzi sono pubblicamente assegnati dall'autorità mondiale IANA (*Internet Assigned Numbers Authority*) e sono unici al mondo, per cui rendono univocamente determinata a livello mondiale la raggiungibilità su Internet degli *end-node* a cui vengono attribuiti.

Inoltre ai LNF si fa largo uso di range di indirizzamento privati (o nascosti) per un'ampia serie di *end-nodes* che comunicano all'interno della LAN e non hanno la necessità di raggiungere il mondo esterno (ad esempio gli switch di rete, le stampanti, gli strumenti connessi alla rete, alcune macchine di acquisizione dati, etc.).

Tuttavia non è conveniente che l'elevato numero di *end-nodes* dei LNF comunichino sulla stessa LAN a livello di *data link*, in quanto si verrebbero a trovare in un unico dominio di *broadcast*. Ogni nodo connesso alla rete, genera frequentemente dei pacchetti di tipo *broadcast*. Ciascun pacchetto *broadcast* viene catturato dal livello più basso nello stack di protocollo generando un interrupt alla CPU del nodo che lo riceve. Per questo motivo si tende a limitare il numero dei nodi che insistono sullo stesso dominio di *broadcast*. L'esperienza dimostra che un numero accettabile è circa 200-250 nodi. Questo numero rappresenta il miglior compromesso tra semplicità infrastrutturale della rete e degrado delle prestazioni.

### 2.5.1 Reti Locali Virtuali (VLAN)

La tecnologia delle reti locali virtuali (Virtual LAN o VLAN) fa riferimento alla capacità offerta dagli switch e dai router di configurare più reti logiche sopra un'unica rete locale fisica. Ogni Virtual LAN è costituita da un insieme di *switch-port* dello stesso apparato o anche di diversi apparati. Le stazioni appartenenti ad una VLAN sono logicamente interconnesse a livello *Data Link*. Operando unicamente a livello di centro

di gestione della rete è possibile creare più domini di *broadcast*, cioè più reti locali virtuali, su un'infrastruttura trasmissiva comune senza alcun intervento a livello Fisico.

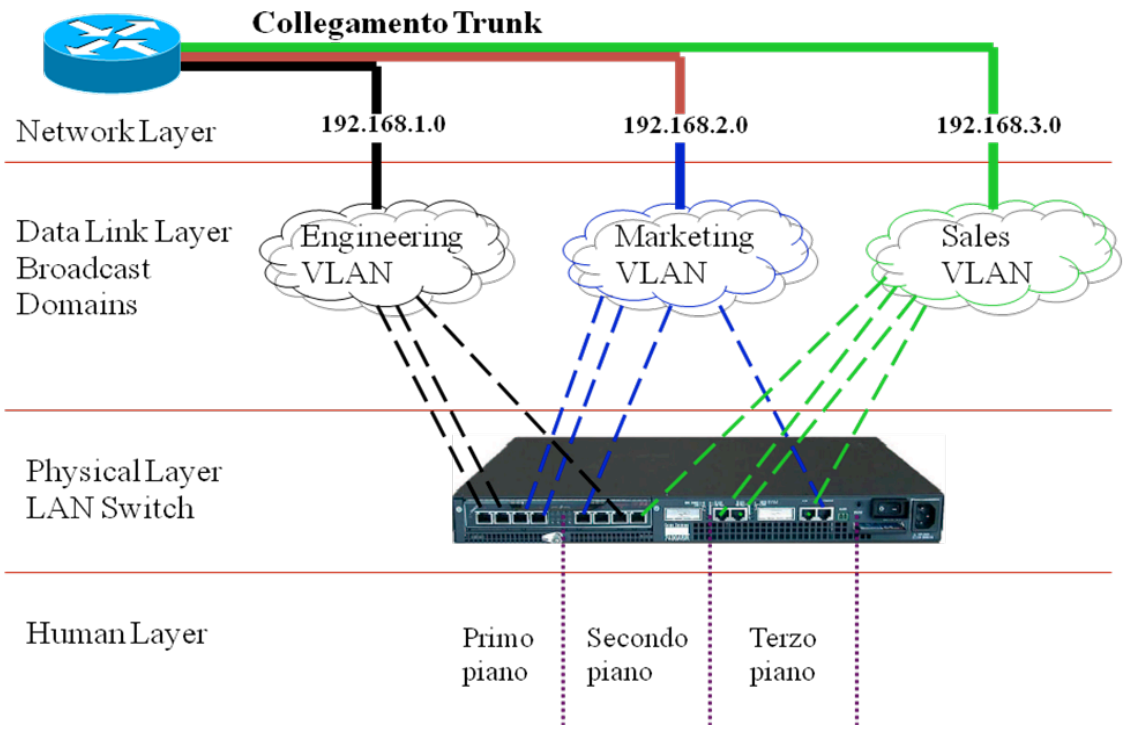



**Figura 2-11: Schema di suddivisione in VLAN**

La possibilità di creare reti locali virtuali da assegnare ai vari gruppi di lavoro permette un'elevata flessibilità in quanto non è necessario che i componenti di un gruppo occupino spazi fisicamente contigui. I vantaggi principali che si ottengono da tale assegnazione derivano dall'isolamento del traffico dei vari gruppi di lavoro al livello *Data Link*. Questo non solo è importante per ragioni di sicurezza e riservatezza dei dati, ma anche perché consente di mantenere separato il traffico di *multicast/broadcast* delle diverse reti virtuali.

L'interoperabilità tra le reti virtuali è garantita da una unità di *internetworking* esterna, normalmente un *router* (rif. Figura 2-11).

La tecnica di suddivisione della rete fisica in VLAN prevede che, a livello centrale, vengano inizialmente definite le VLAN e successivamente che vengano fisicamente mappate le singole porte fisiche di ciascuno *switch* sulle varie VLAN definite. Il singolo nodo che partecipa ad una determinata VLAN si troverà necessariamente attribuite le opportune impostazioni di rete (indirizzo, *netmask*, *gateway*, etc.) in funzione della VLAN stessa a cui appartiene, quindi della porta fisica su cui è connesso.

Questa tecnica, sebbene standard, ha il grosso limite di rappresentare una condizione di particolare staticità della rete e dei suoi *end-nodes*. Infatti, con la mappatura statica sopra descritta, se un *end-node* dovesse spostarsi fisicamente, andrebbe a connettersi su un'altra porta fisica (dello stesso o di altro *switch*) e avrebbe

moltissime probabilità di trovarsi connesso su un'altra VLAN. In tale evenienza, purtroppo molto frequente, rimarrebbero due soluzioni tecniche alternative da adottare:

1) mappare staticamente la nuova porta, su cui si è connesso l'*end-node*, sulla stessa VLAN su cui era originariamente
2) cambiare le impostazioni di rete all'*end-node*

La seconda soluzione è la meno indicata in quanto non è trasparente nei confronti dell'utilizzatore e potrebbe procurare anche alcuni disservizi sull'*end-node*. Inoltre, in entrambi i casi, è richiesto un discreto lavoro di gestione che assume dimensioni considerevoli nel caso di una rete estesa come quella dei LNF.

Grazie alla scelta di avere apparati della stessa marca e famiglia, si è potuto utilizzare una soluzione proprietaria che risolve brillantemente il problema raggiungendo il duplice obiettivo di minimizzare la complessità di gestione della rete nel caso di spostamenti degli *end-nodes* e di favorire la mobilità dei ricercatori e degli utilizzatori dei servizi di rete (VMPS).

### 2.5.2 VMPS

VMPS è l'acronimo di *VLAN Membership Policy Server*, ovvero server di gestione di appartenenza alle *VLAN*. Tramite questo protocollo, proprietario di Cisco, si può definire l'appartenenza di un nodo ad una specifica VLAN in funzione del suo *MAC address* (indirizzo fisico della scheda di rete) e non in funzione della porta dello *switch* al quale esso viene connesso.

È evidente che questo meccanismo permette una completa mobilità degli utilizzatori della rete nell'ambito della LAN. Infatti gli *end-nodes* (ad esempio i Personal Computer) verranno mappati sempre sulla stessa VLAN, indipendentemente da quali siano le porte dello switch di rete a cui vengono connessi. Ciò permette anche di garantire ai singoli nodi di rete, sempre le stesse impostazioni di rete, indipendentemente dal loro posizionamento fisico.

Affinché il sistema funzioni, è necessario preventivamente costruire un database che metta in relazione i *MAC address* con le relative *VLAN*. Questo database risiede su alcuni *switch* centrali (che svolgono il ruolo di *VMPS server* per la *LAN*). L'aggiornamento del database avviene su un semplice file testuale che viene caricato dagli switch centrali tramite un *TFTP server* automaticamente al *boot,* o manualmente su richiesta del gestore di rete. Inoltre, su tutti gli switch della *LAN*, va definita ciascuna porta in modalità d'accesso *dynamic*, che significa che non viene assegnata staticamente ad una specifica *VLAN*, ma è pronta ad impostarsi sulla *VLAN* designata, in funzione del *MAC address* del nodo che ad essa viene connesso.

Questo meccanismo è molto utile per realizzare una *LAN* sicura e al tempo stesso non troppo 'blindata': se una stazione di lavoro appartiene a un dipendente o a una persona che lavora stabilmente all'interno dei LNF il suo *MAC address* viene inserito all'interno del database *VMPS* e, di conseguenza, il nodo stesso sarà inserito in una *VLAN* opportuna.

Se invece qualcuno collega alla rete un PC "sconosciuto" (cioè il cui *MAC address* non è presente nel database *VMPS*) esso viene comunque inserito in una *VLAN*

particolare (detta di *fallback*), in modo che l'utente possa lavorare solo se è in possesso di credenziali d'accesso ai sistemi di autenticazione dei LNF (rif. paragrafo 2.6).

### 2.5.3 Partizionamento della LAN dei LNF

Le diversificate necessità dei vari utenti dei Laboratori hanno indotto ad un partizionamento della rete locale in varie VLAN. Seguono due tabelle riepilogative, la prima rappresentante le network ad indirizzamento pubblico (raggiungibili dall'esterno), la seconda rappresentante le network ad indirizzamento privato, il cui traffico è confinato all'interno della LAN.

In particolare le *network* elencate nella Tabella 2-2, sono utilizzate per tutti quei nodi che hanno la necessità di raggiungere anche la rete Internet esterna:

- le VLAN 2 e 26 sono dedicate all'uso dell'esperimento Kloe, la 2 per le macchine dedicate al calcolo *offline*, e la 26 per le macchine di acquisizione;
- le VLAN 228 e 230 sono dedicate al Sistema Informativo e Gestionale dell'INFN, la 228 per i *database server* e per gli *application server*, la 230 per i *client* degli sviluppatori;
- la VLAN 84 è dedicata ai server del Servizio di Calcolo;
- la VLAN 87 è dedicata ai PC portatili che non hanno una collocazione fissa e stabile nei Laboratori
- la VLAN 131 è dedicata all'uso dei portatili connessi in *wireless* (rete senza fili) e integrati nel progetto TRIP (rif. paragrafo 2.6)
- Tutte le altre sono dedicate ai nodi (*server* e *client*) che hanno una connessione stabile alla LAN dei LNF. In particolare la 129 è dedicata ai nodi dell'Amministrazione Centrale e la 3 alla grigia computazionale (rif. capitolo 5).

| N. VLAN | Nome VLAN | Network IP | Netmask | Gateway |
|---|---|---|---|---|
| 2 | Kloe-25 | 192.135.25.0 | 255.255.255.0 | 192.135.25.11 |
| 26 | Kloe-26 | 192.135.26.0 | 255.255.255.0 | 192.135.26.11 |
| 80 | Public-80 | 193.206.80.0 | 255.255.255.0 | 193.206.80.11 |
| 81 | Public-81 | 193.206.81.0 | 255.255.255.0 | 193.206.81.11 |
| 82 | Public-82 | 193.206.82.0 | 255.255.255.0 | 193.206.82.11 |
| 83 | Public-83 | 193.206.83.0 | 255.255.255.0 | 193.206.83.11 |
| 84 | Public-84 | 193.206.84.0 | 255.255.255.0 | 193.206.84.11 |
| 85 | Public-85 | 193.206.85.0 | 255.255.255.0 | 193.206.85.11 |
| 86 | Public-86 | 193.206.86.0 | 255.255.255.0 | 193.206.86.11 |
| 87 | Dynamic | 193.206.87.0 | 255.255.255.0 | 193.206.87.11 |
| 3 | 84-128 | 192.84.128.0 | 255.255.255.0 | 192.84.128.11 |
| 129 | AC | 192.84.129.0 | 255.255.255.0 | 192.84.129.11 |
| 4 | 84-130 | 192.84.130.0 | 255.255.255.0 | 192.84.130.11 |
| 131 | LANesterna | 192.84.131.0 | 255.255.255.0 | 192.84.131.11 |
| 228 | sisinfo | 193.205.228.0 | 255.255.255.192 | 193.205.228.1 |
| 230 | sisinfodev | 193.205.228.64 | 255.255.255.224 | 193.205.228.65 |

**Tabella 2-2: VLAN e network ad indirizzamento pubblico**

| N. VLAN | Nome VLAN | Network IP | Netmask | Gateway |
|---|---|---|---|---|
| 1 | default | 172.16.0.0 | 255.255.0.0 | 17.16.14.1 |
| 17 | LoadBalancing | 172.17.0.0 | 255.255.0.0 | 172.17.1.1 |
| 128 | HiddenPrinters | 192.168.128.0 | 255.255.255.0 | Non Instradata |
| 130 | ACprinters | 192.168.130.0 | 255.255.255.0 | 192.168.130.1 |
| 132 | LNFprinters | 192.168.132.0 | 255.255.255.0 | 192.168.132.1 |
| 138 | Fisa | 192.168.138.0 | 255.255.255.0 | 192.168.138.1 |
| 140 | Strumenti | 192.168.140.0 | 255.255.255.0 | 192.168.140.1 |
| 160 | PCMaster | 192.168.160.0 | 255.255.255.0 | 192.168.160.1 |
| 161 | LabMaster | 192.168.161.0 | 255.255.255.0 | 192.168.161.1 |
| 180 | Teorici | 192.168.180.0 | 255.255.255.0 | 192.168.180.1 |
| 192 | Dante | 192.168.192.0 | 255.255.255.0 | 192.168.192.1 |
| 193 | Sunray1 | 192.168.193.0 | 255.255.255.0 | Non Instradata |
| 194 | Sunray2 | 192.168.194.0 | 255.255.255.0 | Non Instradata |
| 195 | Dafne | 192.168.195.0 | 255.255.255.0 | 192.168.195.1 |
| 196 | Apple | 192.168.196.0 | 255.255.255.0 | Non Instradata |
| 197 | Sparc | 192.168.197.0 | 255.255.255.0 | 192.168.197.1 |
| 198 | Dante-data | 192.168.198.0 | 255.255.255.0 | Non Instradata |
| 199 | Finuda | 192.168.199.0 | 255.255.255.0 | 192.168.199.1 |
| 200 | Guests | 192.168.200.0 | 255.255.255.0 | 192.168.200.1 |
| 204 | Leale | 192.168.204.0 | 255.255.255.0 | Non Instradata |
| 208 | Vpnlnf | 192.168.208.0 | 255.255.255.0 | 192.168.208.1 |
| 218 | Sanstk | 192.168.218.0 | 255.255.255.0 | Non Instradata |
| 219 | Sanemc2 | 192.168.219.0 | 255.255.255.0 | Non Instradata |
| 220 | Console | 192.168.220.0 | 255.255.255.0 | 192.168.220.1 |
| 221 | Nfs | 192.168.221.0 | 255.255.255.0 | Non Instradata |
| 229 | HiddenSisinfo | 192.168.229.0 | 255.255.255.0 | Non Instradata |

**Tabella 2-3: VLAN e network ad indirizzamento privato**

Le *network* elencate nella Tabella 2-3, sono invece utilizzate per tutti quei nodi che hanno necessità di connettività limitata alla rete locale dei LNF. In alcuni casi le network non sono instradate, il ché implica che il traffico è confinato all'interno della stessa VLAN; le VLAN con sfondo giallo sono invece dedicate alla gestione di servizi critici del Servizio di Calcolo:

- le VLAN 130 e 132 sono dedicate alle stampanti personali (o di gruppo) dei LNF e di AC rispettivamente;
- la VLAN 128 è dedicata alle stampanti di volume e/o di piano gestite dal Servizio di Calcolo; la network non è instradata e le stampanti possono essere raggiunte esclusivamente attraverso un *printer-Server*;
- la VLAN 138 è dedicata a nodi e a strumenti di acquisizione dei dati dosimetrici della Fisica Sanitaria dei LNF;
- la VLAN 140 è dedicata a strumenti di acquisizione e/o di controllo di apparati elettronici (PLC, oscilloscopi, etc.);
- le VLAN 160 e 161 sono dedicate a nodi (*client* e *server*) prevalentemente usati a scopo didattico; inizialmente usati per la realizzazione di alcuni Master post-universitari in Fisica Nucleare e in *Information Technology*;
- la VLAN 180 è dedicata ad una network di servizio per una farm di calcolo del gruppo dei Fisici Teorici dei LNF;

- le VLAN 192, 193, 194, 195, 197 e 198 sono dedicate alla Divisione Acceleratori per la gestione della Macchina Acceleratrice; in particolare la VLAN 192 accoglie i *server* e i *client* di applicazioni per la gestione dello *Slow Control* di Dafne; le VLAN 193 e 194 sono dedicate a particolari client d'accesso (*SunRays*) ai suddetti *server*; le VLAN 195 e 198 sono destinate a network di servizio;
- la Vlan 199 è dedicata all'uso dell'esperimento Finuda, in particolare per le macchine di acquisizione dati;
- la VLAN 200 è dedicata ai nodi sconosciuti (non registrati nel database VMPS) e ai nodi Wireless; la connettività di tali nodi verso l'esterno è condizionata all'autenticazione degli utilizzatori su un apposito *Web Captive Portal* (rif. paragrafo 2.6);
- la VLAN 204 è dedicata ad una network di servizio per una farm di calcolo del Gruppo III dei Fisici Nucleari dei LNF;
- la VLAN 208 è dedicata all'accesso dall'esterno in Virtual Private Network alle risorse della LAN dei LNF (rif. paragrafo 6.5.3.1);
- le VLAN 1, 17, 196, 218, 219, 220, 221 e 229 sono dedicate al Servizio di Calcolo per il traffico di servizio e di gestione di importanti dispositivi informatici, di rete, di calcolo e di storage, quali le *console* dei *server*, le *console* degli *switch Fibre-Channel* e degli *Storage Controller*, il *Load Balancing* di alcuni importanti servizi all'utenza, etc..

### 2.5.4 Connessione tra il Calcolo e gli altri edifici

Nel centro stella della rete, situato nella sala macchine dell'edificio Calcolo, sono installati 2 switch Catalyst 6500, che rappresentano gli apparati di *core* della rete locale. Il primo è configurato con 50 porte Gigabit Ethernet ottiche e 48 porte Gigabit Ethernet in rame. Il secondo è configurato con 18 porte Gigabit Ethernet ottiche e 144 porte Gigabit Ethernet in rame.

Tutti gli edifici periferici sono connessi ad uno dei due *switch* del Calcolo (rif. Figura 2-12). I collegamenti di interconnessione sono realizzati in Gigabit Ethernet su fibra ottica e sono tutti definiti come *trunk* in modo da poter permettere il traffico di più VLAN sullo stesso collegamento fisico, secondo lo standard IEEE 802.1q.

Gli edifici rappresentati nella precedente figura sono destinati all'utenza, quindi ospitano prevalentemente *client* di accesso alla rete ed alle risorse informatiche. Nei paragrafi successivi saranno viste in modo più dettagliato le reti dedicate ad attività scientifiche specifiche, quali Macchine Acceleratrici (Sparc e Dafne), Esperimenti (Kloe e Finuda), attività di Fisica Nucleare (Gruppo III) e Servizio di Calcolo centrale.

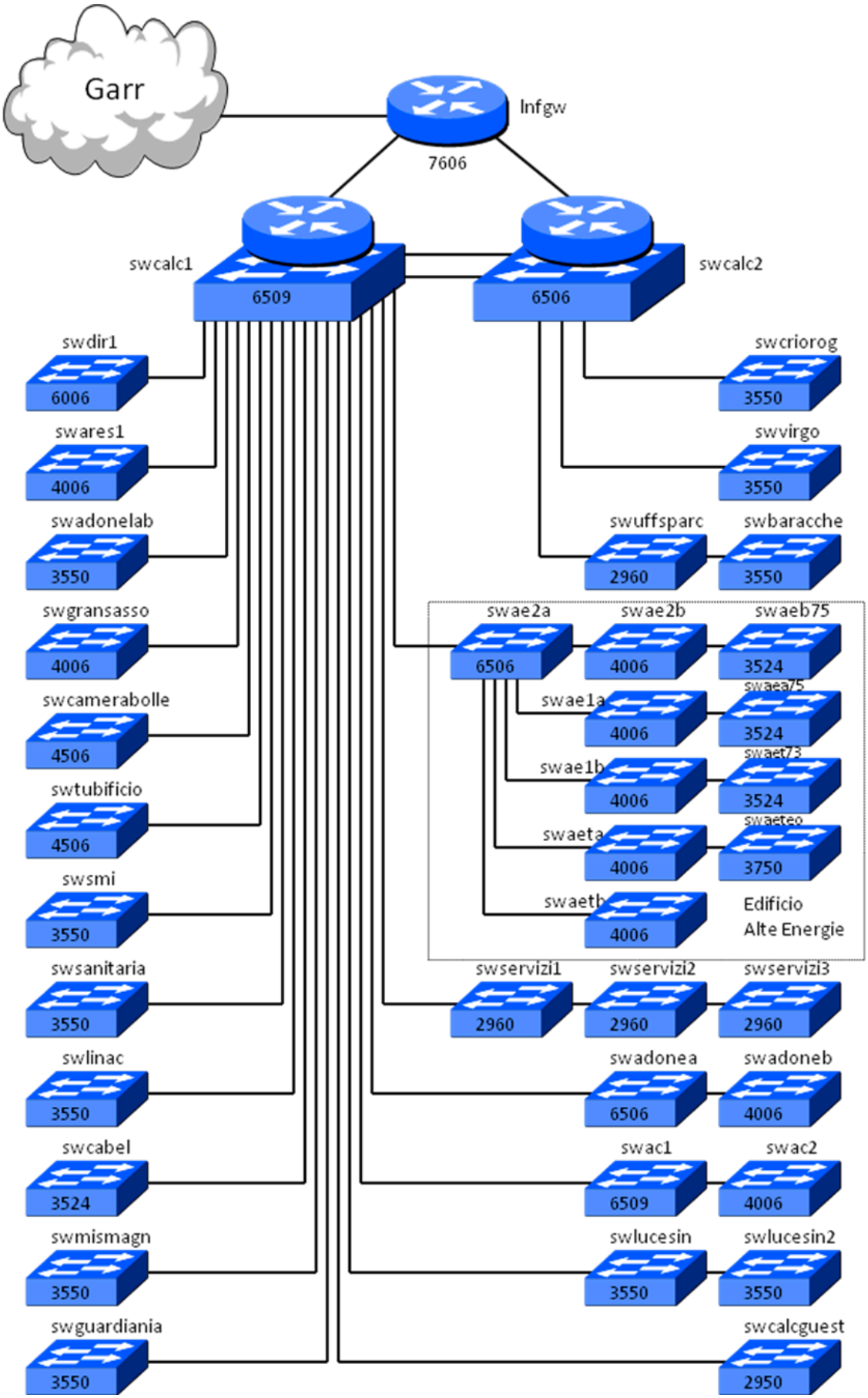



**Figura 2-12: Connessione tra il Calcolo e gli altri edifici**

### 2.5.5 La rete di Sparc

La sezione di rete realizzata per l'attività scientifica di Sparc (Sorgente Pulsata Autoamplificata di Radiazione Coerente) è rappresentata nella Figura 2-13.

Gli switch sono distribuiti geograficamente nelle 3 grandi sale dedicate alla macchina e ai controlli, posizionati in modo da favorire la connessione dei *server*, degli strumenti di controllo e dei sistemi di *monitoring*.

Per motivi prestazionali, è realizzata con *switch* che supportano porte in rame a 10/100/1000 Mbit/s, interconnessi al centro stella (*switch swsparcge1*) attraverso canali ottici *Gigabit EtherChannel* definiti come *trunk* per il trasporto del traffico di più VLAN.

*Gigabit EtherChannel* è un protocollo standard di aggregazione di *link* omogenei. Lo standard prevede l'aggregazione di 2, 4 o 8 *link*. Nel caso di Sparc sono stati realizzati canali *Gigabit EtherChannel* che aggregano 2 *link*. Tale soluzione garantisce il raggiungimento di 2 obiettivi principali: i canali sono in grado di raddoppiare il *throughput* complessivo (2 Gigabit/s *full duplex*); inoltre aumenta l'affidabilità del singolo canale, in quanto la raggiungibilità degli *switch* interconnessi tra loro in *Gigabit EtherChannel* è garantita anche a seguito di interruzioni di un singolo *link*.

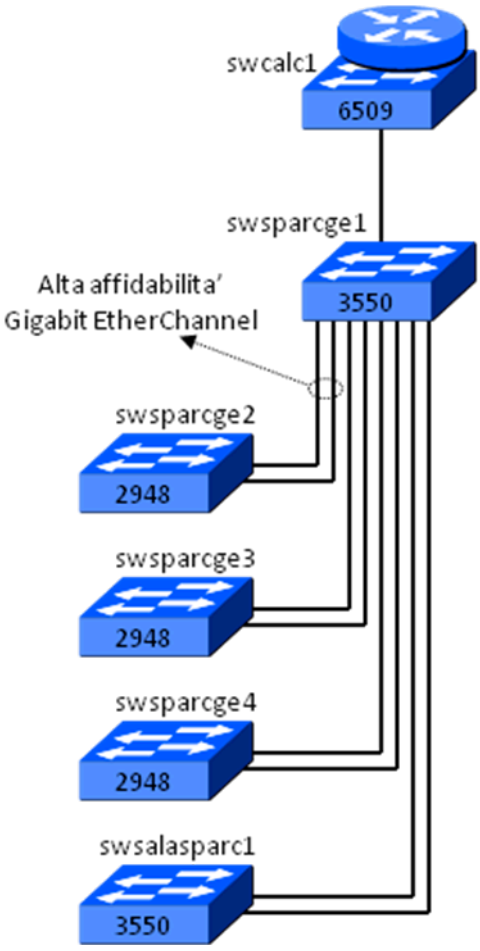



**Figura 2-13: La rete di Sparc**

Il centro stella della rete di Sparc, *swsparcge1*, funge contemporaneamente da nodo di concentrazione ottico e da nodo di distribuzione di rete in rame. Gli altri *switch* svolgono esclusivamente la funzione di distribuzione di rete in rame. Lo *switch swsparcge1* a sua volta è direttamente connesso allo *switch* principale del Calcolo *swcalc1* con un link ottico in *Gigabit Ethernet*, definito come *trunk*, per la connessione al resto della rete (LAN e WAN).

### 2.5.6 La rete di Dafne

La sezione di rete dedicata alla Macchina Acceleratrice Dafne è rappresentata nella Figura 2-14.

Il sistema di controllo della Macchina Acceleratrice è particolarmente complesso e sofisticato. Molti apparati elettronici svolgono il compito di controllare ed eventualmente correggere il funzionamento dei componenti principali attivi della macchina, quali, ad esempio, gli alimentatori, i magneti, i modulatori, i sistemi a radiofrequenza, i sistemi di criogenia, i sistemi di iniezione, i sistemi di rilevazione dei parametri di intensità di corrente, di energia, di luminosità, etc..

Tutti questi apparati elettronici sono a loro volta controllati da sistemi informatici di elevate prestazioni e di elevata affidabilità. Sistemi *unix* dedicati a tale scopo su cui vengono eseguite applicazioni appositamente progettate con l'ausilio di strumenti s/w specifici (*Labview*).

La rete è un'infrastruttura necessaria per la comunicazione tra i sistemi e gli apparati elettronici, dalla cui funzionalità dipende il controllo della Macchina Acceleratrice.

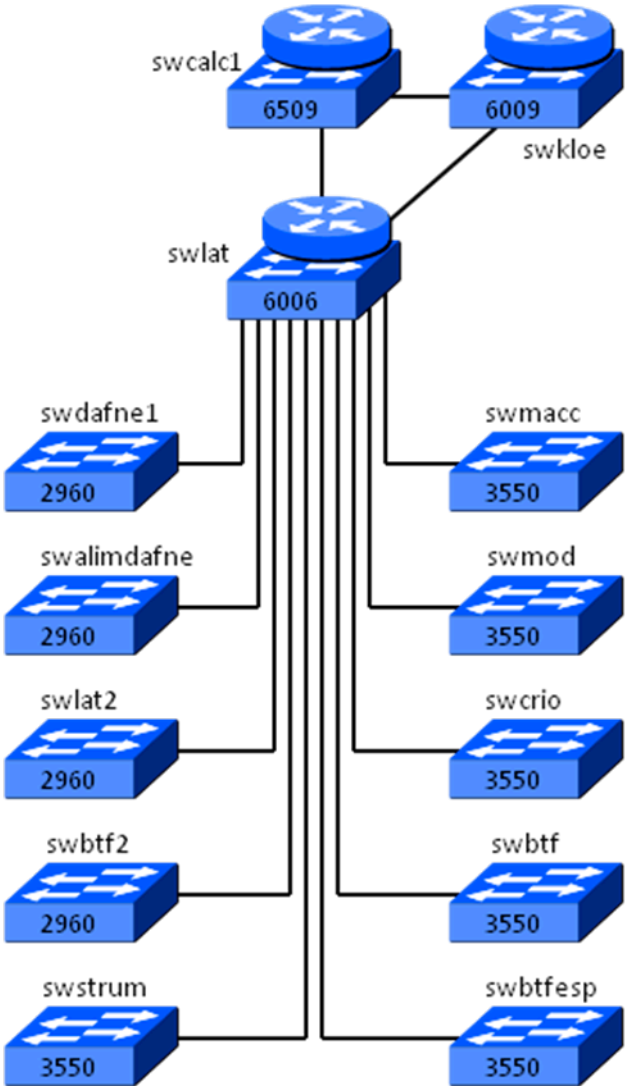



**Figura 2-14: La rete di Dafne**

Il centro stella della rete è lo *switch swlat*; tale *switch* funge contemporaneamente da nodo di concentrazione ottico e da nodo di distribuzione di rete in rame, per la connessione dei *server* principali, dei sistemi *client* e dei sistemi di *monitoring* situati in Sala Controllo Dafne. *Swlat* svolge sia funzioni di *switching* di livello 2, sia funzioni di *switching* di livello 3 (*routing*). Gli altri *switch* svolgono esclusivamente la funzione di distribuzione di rete in rame e sono connessi a *swlat* tramite *link* ottici (necessari per evitare disturbi elettromagnetici), definiti come *trunk* per il trasporto del traffico di più VLAN. Lo *switch swlat* a sua volta è direttamente connesso allo *switch* principale del Calcolo *swcalc1* e allo *switch* dedicato all'esperimento kloe *swkloe* con link ottici in *Gigabit Ethernet* (anch'essi definiti come *trunk*), per la connessione al resto della rete (LAN e WAN).

### 2.5.7 La rete dell'esperimento Kloe

La sezione di rete dedicata all'esperimento Finuda è rappresentata nella Figura 2-15.

Gli *switch* sono distribuiti geograficamente in 4 aree, una dedicata alla sala calcolo Kloe, una dedicata alla robotica a nastri (nell'edificio Calcolo), una vicino alla gestione del gas del detector, e una proprio sul detector.

La parte più critica è rappresentata dallo switch centrale *swkloe*, a cui sono connesse, in *Gigabit Ethernet* in rame, tutte le macchine della farm di calcolo dell'Esperimento.

Gli altri *switch* connettono strumenti di controllo e sistemi di *monitoring*, *server* per l'acquisizione dati e sistemi per la gestione della robotica e delle librerie a nastri (queste ultime localizzate nell'edificio Calcolo).

Lo *switch swkloe* svolge la funzione di centro stella della rete; tale *switch* funge contemporaneamente da nodo di concentrazione ottico e da nodo di distribuzione di rete in rame. Gli altri *switch* svolgono esclusivamente la funzione di distribuzione di rete in rame e sono connessi a *swkloe* tramite *link* ottici, definiti come *trunk* per il trasporto del traffico di più VLAN.

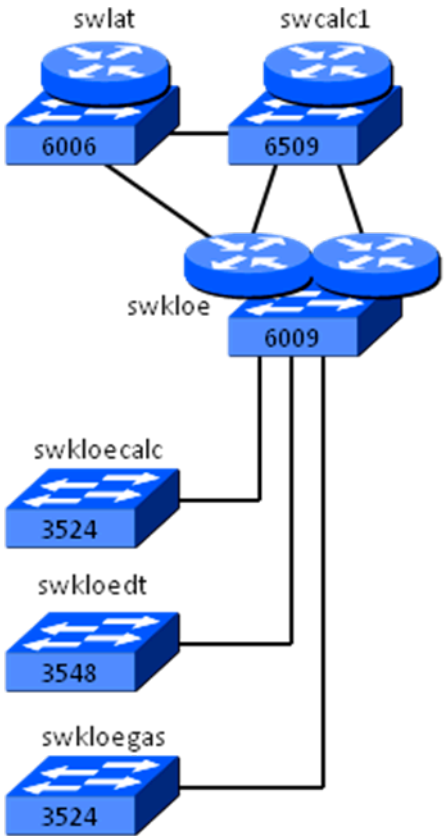



**Figura 2-15: La rete di Kloe**

Lo *switch swkloe* ospita due moduli di processing per il *multylayer switching* (che svolgono sia funzioni di *switching* di livello 2, sia funzioni di *switching* di livello 3 o *routing*), ognuno dei quali è direttamente connesso allo *switch* principale del Calcolo *swcalc1*, tramite *link* ottici definiti come *trunk* per il trasporto del traffico di più VLAN.

Allo scopo di aumentare l'affidabilità complessiva dell'infrastruttura di rete, oltre alla doppia connessione allo switch centrale del Calcolo *swcalc1*, gestita con il protocollo *HSRP* (Hot Stand-by Routing Protocol, rif. paragrafo 2.5.14), è realizzata anche una maglia triangolare con lo switch *swlat*, atta a garantire la comunicazione tra l'Esperimento e la Macchina Acceleratrice Dafne, anche in caso di down del Calcolo.

### 2.5.8 La rete dell'esperimento Finuda

La sezione di rete dedicata all'esperimento Finuda è rappresentata nella Figura 2-16.

Gli *switch* sono distribuiti geograficamente in 2 aree, una dedicata al *Pit* dell'esperimento vicino al Calorimetro e l'altra dedicata alla Sala Controllo, posizionati in modo da favorire la connessione dei *server*, degli strumenti di controllo e dei sistemi di *monitoring*.

Per motivi prestazionali, è realizzata con *switch* che supportano porte in rame a 10/100/1000 Mbit/s, quasi tutti interconnessi al centro stella (*switch swfinudage*) attraverso canali ottici *Giga-bit EtherChannel* definiti come *trunk* per il trasporto del traffico di più VLAN.

Unica eccezione è rappresentata dallo *switch swfinudasala* che è connesso con un *link* ottico *Gigabit Ethernet* al centrostella e con un *link* ottico *Gigabit Ethernet* allo *switch swfinudamac* in modo tale da formare una maglia triangolare atta a garantire una maggiore affidabilità complessiva della rete.

Il centro stella della rete di Finuda, *swfinudage*, funge da nodo di concentrazione ottico. Gli altri *switch* svolgono esclusivamente la funzione di distribuzione di rete in rame. Lo *switch swfinudage* a sua volta è direttamente connesso allo *switch* principale del Calcolo *swcalc1* con un link ottico in *Gigabit Ethernet*, definito come *trunk*, per la connessione al resto della rete (LAN e WAN).

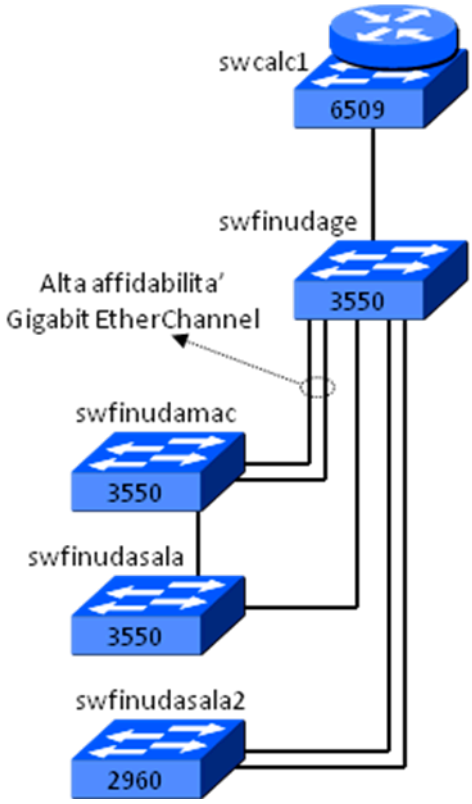



**Figura 2-16: La rete di Finuda**

### 2.5.9 La rete di Gruppo III

La sezione di rete dedicata all'esperimento Finuda è rappresentata nella Figura 2-17.

Gli *switch* sono distribuiti geograficamente in 2 aree, dedicate alle farm di calcolo, posizionati in modo da favorire la connessione dei *server*.

Per motivi prestazionali, è realizzata con *switch* che supportano porte in rame a 10/100/1000 Mbit/s, quasi tutti interconnessi al centro stella (*switch swleale1*) attraverso canali ottici *Gigabit EtherChannel* definiti come *trunk* per il trasporto del traffico di più VLAN.

Il centro stella della rete di Gruppo III, *swleale1*, funge sia da nodo di concentrazione ottico, che da nodo di distribuzione di rete in rame, per la connessione dei *server* principali, dei sistemi *client*, e degli uffici presenti nell'edificio LEALE.

Gli altri *switch* svolgono esclusivamente la funzione di distribuzione di rete in rame.

Lo *switch swleale1*, a sua volta, è direttamente connesso allo *switch* principale del Calcolo *swcalc1* con un link ottico in *Gigabit Ethernet*, definito come *trunk*, per la connessione al resto della rete (LAN e WAN).

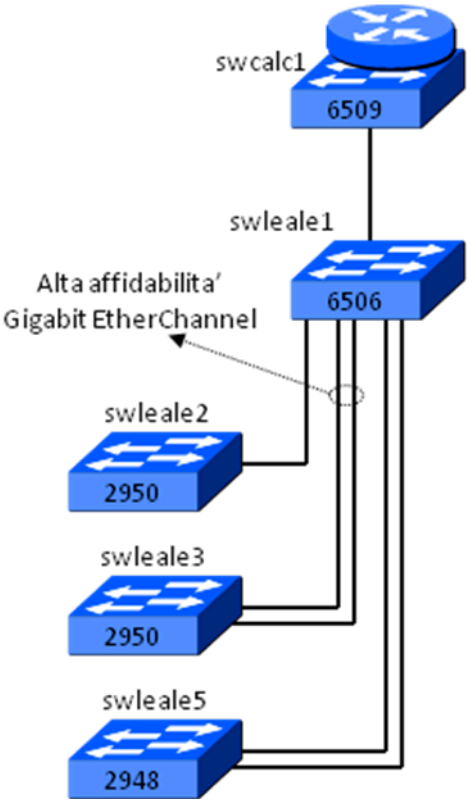



**Figura 2-17: La rete di Gruppo III**

## 2.5.10 La rete del Calcolo

La sezione di rete dedicata al Calcolo è rappresentata nella **Figura 2-18** Figura 2-18. Tale infrastruttura di rete è tutta concentrata nella sala macchine dedicata al Servizio di Calcolo e rappresenta il nucleo centrale della rete dei LNF. Infatti tutte le sezioni di rete trattate nei precedenti paragrafi, sono connesse ad uno dei due *switch* centrali di tale infrastruttura (*swcalc1* e *swcalc2*).

Nella sala macchine del Servizio di Calcolo sono installati un centinaio di server che svolgono le funzioni più disparate e, in molti casi, forniscono servizi di estrema criticità per il corretto funzionamento dell'intera infrastruttura informatica, o per la fruibilità di applicazioni e servizi di fondamentale importanza per l'utenza. Molti di questi sono descritti nei capitoli successivi (rif. capitoli 4 e 5).

Data la criticità dei ruoli svolti da tali server, la rete a cui essi sono connessi è stata costruita in modo da garantire una affidabilità di altissimo livello, per scongiurare le interruzioni di servizio in conseguenza di guasti hardware e software, sia degli apparati di *switch*, che dei server stessi.

I server più critici sono generalmente ridondati (ovvero 2 o più server forniscono indistintamente lo stesso servizio) e sono connessi a 2 diversi *edge switch* (ad esempio *swcalcrack* o *swcalcfarm1*). Gli edge switch, a loro volta, sono connessi ad entrambi gli switch di centrali.

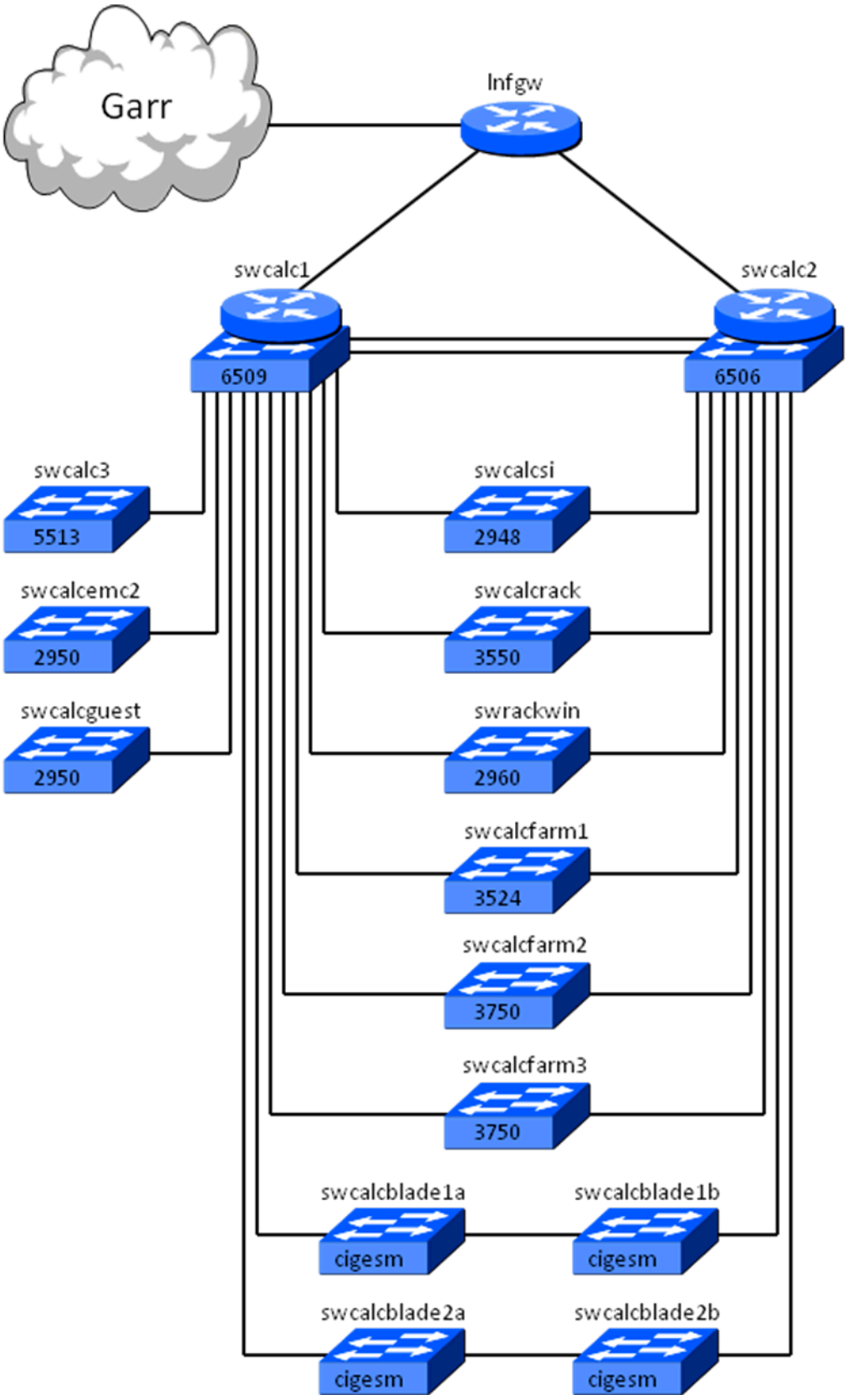



**Figura 2-18: La rete del Servizio di Calcolo**

La funzionalità dello *switching* di livello 2 (Ethernet), che pretende una struttura ad albero e non prevede la possibilità di magliature nella rete, attraverso un protocollo standard denominato *Spanning Tree,* definisce un persorso prioritario (o primario) e mette l'altro in *hot stand-by*. Ovvero uno dei due link di connessione agli *switch* di *core* viene disabilitato al traffico, ma rimane pronto a rientrare in funzione nel caso di interruzione del link prioritario o dello *switch* di *core* a cui esso è connesso.

Questo metodo consente di garantire sempre un percorso possibile tra *l'edge switch* e uno dei due *switch* di *core*. Nel caso di fallimento dell'*edge switch*, la funzionalità del servizio viene garantita dal server ridondato (gemello) connesso all'altro *edge switch*.

In sala macchine sono installate anche alcune soluzioni di calcolo basate su sistemi *blade*, ovvero macchine ad alta densità incluse in un unico chassis, che già alloggia coppie di moduli per la connettività verso l'esterno, come *switch Fibre Channel* e *switch Gigabit Ethernet*. Ad esempio, gli *switch swcalcblade2a* e *swcalcblade2b* (anch'essi Cisco) sono inclusi in uno chassis IBM in grado di alloggiare fino a 14 macchine biprocessori *quad-core* (complessivamente 8 core per macchina); ogni macchina è connessa tramite 2 canali *Gigabit Ethernet* a entrambi gli *switch* interni allo *Chassis* e i 2 *switch* a loro volta sono connessi ai 2 *switch* di *core* (secondo lo schema in figura), a formare una maglia quadrangolare.

La funzionalità di *switching* di livello 3 (*routing*) è garantita sempre dagli *switch* di *core*. Anche in questo caso l'alta affidabilità è garantita da un apposito protocollo standard denominato HSRP (*Hot Stand-by Routing Protocol*), descritto nel paragrafo 2.5.14.

### 2.5.11 VTP

Il management dei *trunk* è generalmente particolarmente complesso. Infatti su ogni *trunk* devono essere definite singolarmente tutte le VLAN il cui traffico è abilitato a transitare. Nel caso della rete dei LNF, in virtù delle scelte effettuate, significherebbe dover definire praticamente tutte le VLAN su ciascun *trunk*.

Ogni volta che si installa un nuovo apparato di *switch*, oppure ogni volta che si definisce un nuovo trunk occorre quindi abilitare il passaggio di tutte le VLAN definite. Ancor peggio, quando si crea una nuova VLAN, questa va abilitata sugli innumerevoli *trunk* già realizzati. Ciò richiederebbe una serie di comandi ripetuti 2 volte per ogni *trunk*, ovvero complessivamente centinaia di volte.

Per risolvere questo problema, si è utilizzato il protocollo VTP (*VLAN Trunk Protocol*). Tramite questo protocollo, proprietario di Cisco, si possono definire tutte le VLAN a livello centrale, su uno *switch* che funge da VTP *server*; sarà tale switch a propagare l'informazione della tabella delle VLAN agli altri switch, che fungono da VTP *client,* e fare in modo che siano abilitate su ciascun trunk.

Il protocollo è particolarmente evoluto, in quanto permette di evitare l'inutile traffico dei *broadcast* di una determinata VLAN, sui trunk che interconnettono apparati che non presentano nodi di quella VLAN (VTP *pruning*).

### 2.5.12 Instradamento del traffico tra diverse VLAN

I nodi connessi ad una determinata VLAN sono in grado di comunicare direttamente solo con gli altri nodi della stessa VLAN. Affinché i nodi su una data VLAN possano comunicare con i nodi di un'altra VLAN è necessario che il traffico passi per un opportuno dispositivo che è connesso ad entrambe le VLAN (canonicamete detto *router* o *gateway*) .

Sulla LAN dei LNF l'instradamento del traffico (*routing*) è una funzionalità insita in alcuni apparati di *switching* di livello 2. Ovvero alcuni *switch* centrali sono in grado di effettuare anche l'instradamento del traffico di livello 3 (*Internet Protocol*) tra le diverse VLAN. Questi apparati svolgono lo stesso lavoro di un *router*, tuttavia sono in grado di instradare i pacchetti IP con la stessa latenza con cui effettuano il lavoro di switching di livello 2. Per questo vengono più comunemente chiamati *Layer 3 switch*. È evidente che con tali apparati il traffico tra VLAN diverse non è penalizzato rispetto al traffico che rimane confinato nella stessa VLAN.

Sulla rete dei LNF, sei apparati svolgono la funzione di *Layer 3 switch* per l'instradamento del traffico locale, mentre uno svolge la funzione di *router e firewall* verso la rete esterna. La rete fisica di interconnessione tra i router (rif. Figura 2-19) è realizzata con alcune magliature per garantire una maggiore affidabilità. In particolare è realizzata una maglia triangolare tra Calcolo, Kloe e Dafne per garantire la raggiungibilità tra Kloe e Dafne anche nel caso di *down* del Calcolo o nel caso di interruzione di una qualunque fibra della maglia. Un'altra maglia è realizzata tra due router principali del Calcolo e il router di frontiera per l'accesso al GARR.

Nello *switch* di Kloe, vi sono 2 moduli di *routing* in grado di svolgere il lavoro indipendentemente; questi sono stati configurati per effettuare l'instradamento in alta affidabilità attraverso il protocollo HSRP (*Hot Stand-by Routing Protocol*).

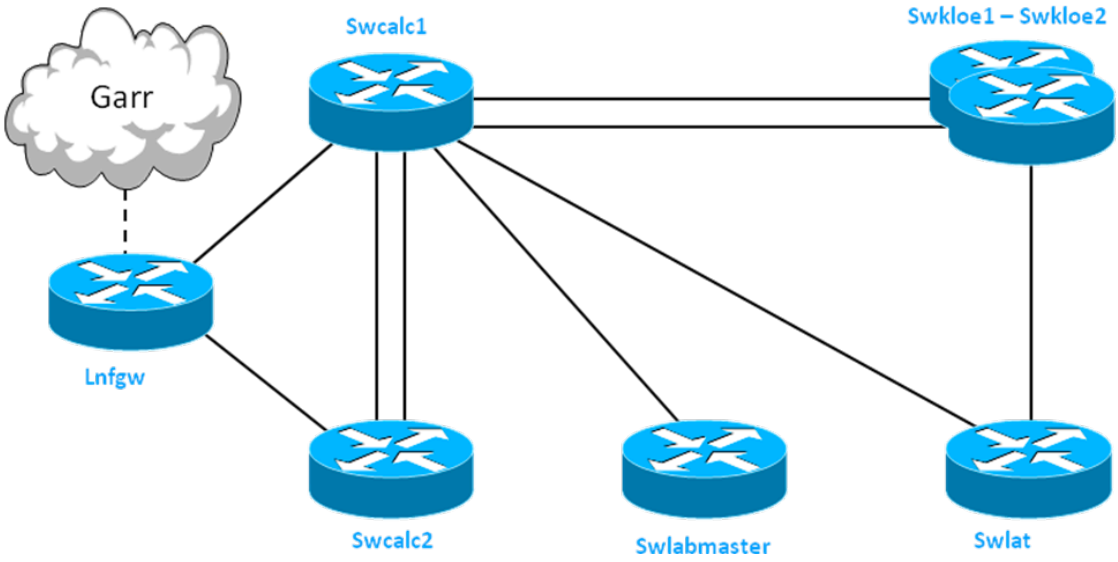



**Figura 2-19: Rete fisica di interconnessione tra i router**

Anche i due switch indipendenti del Calcolo sono configurati in HSRP per aumentare l'affidabilità complessiva della rete.

L'instradamento delle network, a livello logico, è rappresentato nella Figura 2-20. Ogni router ha un compito specifico e instrada le network di propria competenza. La topologia della rete è tale per cui da ogni router a qualsiasi altro, la raggiungibilità a livello 3 è garantita da due possibili strade.

Su tutti router è configurato il protocollo di routing OSPF (Open Shortest Path First, rif. paragrafo successivo), per cui le rotte di instradamento dei router sono dinamiche in funzione del cambiamento della topologia della rete dovuta a guasti o a nuove installazioni.

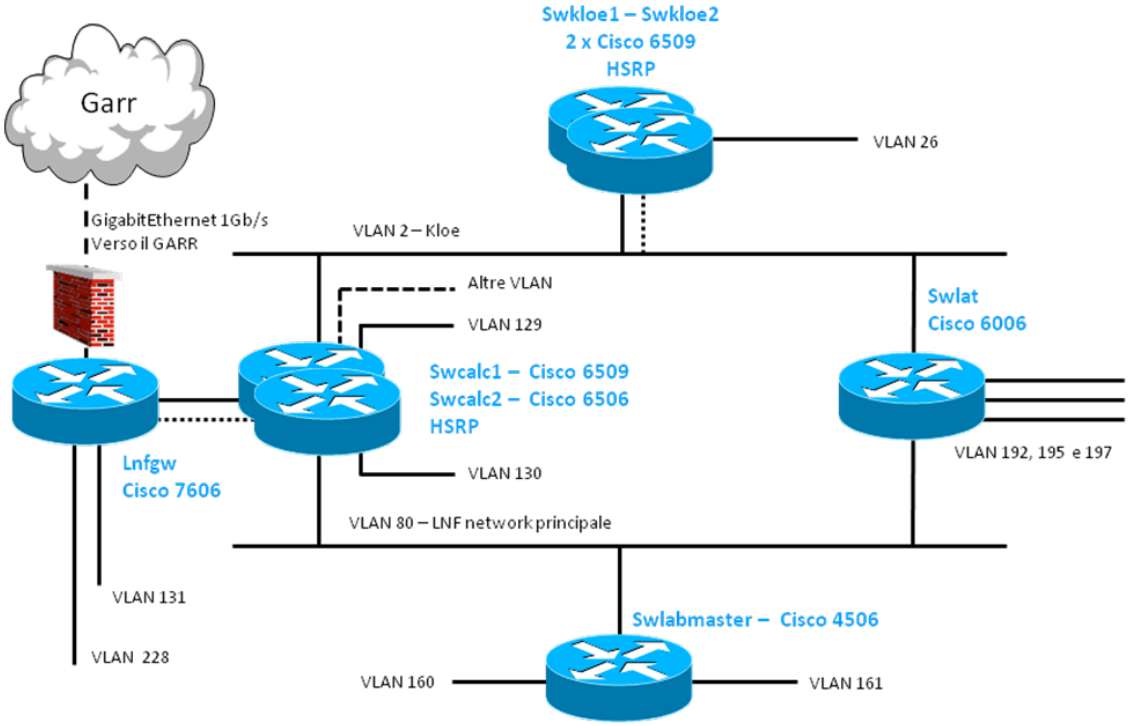



**Figura 2-20: Rete logica di instradamento delle network**

## 2.5.13 OSPF

L'*Open Shortest Path First* è un protocollo di routing sviluppato appositamente per il TCP/IP dall'IETF (*Internet Engineering Task Force*). Il gruppo di lavoro è stato costituito nel 1988 con lo scopo di realizzare un protocollo di tipo *link state packet* per TCP/IP.

OSPF è stato definito dalla RFC 1247 nel 1991 e ridefinito dalla RFC 1583 nel 1994. OSPF ha il concetto di gerarchia. La radice della gerarchia è l'AS (Autonomous System) che può essere suddiviso in aree, ciascuna delle quali contiene un gruppo di reti contigue.

Il routing all'interno di un'area è detto intra-area, quello tra aree diverse inter-area. Ogni AS ha un'area detta di *backbone* ed identificata con 0.0.0.0 o più semplicemente 0. La backbone area (detta più semplicemente nel seguito backbone) può essere anche non contigua; in tal caso occorre configurare dei *virtual links* per garantire la coesione del backbone.

I router OSPF sono classificati secondo quattro categorie non mutuamente esclusive:

- *Internal router*. Un router in cui tutte le network direttamente connesse appartengono alla stessa area. Questi router utilizzano una sola copia dell'algoritmo OSPF. I router che hanno solo interfacce sul backbone appartengono a questa categoria.
- *Area border router*. Un router che collega più aree. Questi router utilizzano più copie dell'algoritmo OSPF: una copia per ogni area direttamente connessa e una copia per il backbone. Gli area border router condensano le informazioni delle aree a loro collegate e le ridistribuiscono sul backbone. Il backbone ridistribuisce a sua volta queste informazioni alle altre aree.
- *Backbone router*. Un router che ha una interfaccia sul backbone. Questo include tutti i router che si collegano a più di un'area (area border router). I backbone router che hanno tutte le interfacce sul backbone sono considerati internal router
- *AS boundary router*. Un router che scambia informazioni di routing con altri router appartenenti ad altri AS. Questa classificazione è ortogonale alle altre precedenti: un AS boundary router può essere un internal o area border router.

Il protocollo OSPF prevede che, In caso di guasto di un link, i router interconnessi tramite quel link, mandino gli avvertimenti di tale evento su tutte le loro interfacce (LSA, Link State Advertisement). Dopo 5 secondi tutti i router dell'AS ricalcolano, secondo sofisticati algoritmi, le loro tabelle di routing in funzione del nuovo stato della rete. In pochi secondi tutti i router convergono verso una nuova topologia della rete con l'aggiornamento delle relative tabelle di routing.

### 2.5.14 HSRP

L'HSRP (*acronimo di Hot Stand-by Router Protocol*) è un protocollo proprietario di Cisco che serve a garantire la *fault-tolerance* tra più router Cisco nella scelta di un default gateway. Il protocollo è descritto nella RFC 2281.

Il protocollo, se usato in associazione a protocolli di routing a rapida convergenza, quale ad esempio l'OSPF, consente di far fronte al malfunzionamento del router definito come *default gateway*: il protocollo, che opera al "livello applicazione" della pila OSI, consente infatti di organizzare più router Cisco in gruppi, ciascuno dei quali identificato da un numero (*id*); per ogni gruppo viene scelto, in base ad un apposito meccanismo di elezione, un router primario ed uno secondario, che diventa primario in caso di malfunzionamento del primo.

Operativamente l'HSRP invia il proprio messaggio di *hello* in *multicast* (ovvero a tutti i router sulla rete), per contattare gli altri router abilitati all'HSRP e definire le

priorità tra i router: il router primario con la priorità configurata più alta agirà da router virtuale (utilizzando il proprio indirizzo IP e MAC address), che gli hosts sul segmento di rete utilizzeranno come gateway; in caso di malfunzionamento di tale apparato di rete il router con la seconda più alta priorità configurata sarà eletto dall'HSRP come default gateway, minimizzando così i problemi di connettività.

La caratteristica peculiare del protocollo HSRP è quella di assegnare ad ogni gruppo un indirizzo IP virtuale, riferito sempre al router primario, che comunque conserva il suo indirizzo IP fisico. In questo modo, l'IP virtuale resta sempre costante, anche in caso di malfunzionamento del router primario: il router secondario che subentra è identificato sempre dal medesimo indirizzo IP virtuale, pur avendo indirizzo IP fisico diverso.

## *2.6 Rete Wireless – Progetto TRIP e sua implementazione ai LNF*

La rete *wireless* o senza fili (*standard 802.11g*) è una rete condivisa con una banda massima nominale di 54Mbit/s. È evidetemente meno efficiente e meno sicura della rete *wired*, per cui quest'ultima è certamente preferibile nel caso di postazioni fisse o poco mobili.

Tuttavia il proliferare dei PC portatili, e il relativo impiego nelle riunioni di lavoro, ha indotto all'installazione di *Access Point* di rete *wireless* sulla LAN dei LNF. Per tali specifiche esigenze, la politica di implementazione della rete *Wireless* ai LNF prevede che gli *Access Point* vengano installati prevalentemente nelle aule riunioni di ciascun edificio. Dato il numero di aule riunioni nei Laboratori, come conseguenza, una buona parte del territorio dei LNF è raggiunta dal servizio di rete *wireless*.

Data la difficoltà di confinare geograficamente l'estensione del campo elettromagnetico di copertura del servizio di rete Wireless, è evidentemente altrettanto difficile impedire l'accesso a clienti che non ne hanno il diritto.

Inoltre, ad aggravare la situazione, Il 16 agosto 2005, il Ministero dell'Interno ha emanato un Decreto che specifica le misure per contrastare il terrorismo internazionale, focalizzandosi sull'identificazione degli utenti che accedono alle reti da postazioni telematiche non vigilate oppure ai quali viene offerta la possibilità di connettersi alla rete Internet attraverso una tecnologia wireless.

Questo decreto si rivolge principalmente ai titolari ed ai gestori di esercizi pubblici o di circoli privati, nei quali sono posti a disposizione del pubblico, dei clienti oppure dei soci, apparecchi terminali utilizzabili per le comunicazioni telematiche. Individua, inoltre, degli obblighi di condotta per i soggetti che forniscono apparecchi terminali ai frequentatori di centri di ricerca, università ed altri istituti di istruzione, all'interno di tali strutture.

Per quanto riguarda l'INFN, il provvedimento introduce l'obbligo di identificare gli studenti o gli ospiti, cui sono resi disponibili apparecchi terminali oppure la possibilità di connettersi alla rete, disponendo la raccolta e la conservazione dei dati anagrafici acquisiti con modalità informatiche. Naturalmente questo provvedimento

non riguarda coloro che, essendo dipendenti della struttura attraverso un contratto od una borsa di studio, sono già stati identificati e possono usufruire dei servizi di rete.

Il problema introdotto da questo decreto, è stato risolto attraverso l'attuazione del progetto nazionale TRIP (The Roaming INFN Physicist), la realizzazione del quale è descritta nei paragrafi successivi.

### 2.6.1 Infrastruttura di rete Wireless

La rete *wireless* dei LNF è configurata nel rispetto del progetto nazionale TRIP dell'INFN, che prevede la classificazione degli utenti in 2 categorie:

1. dipendenti e associati INFN;
2. visitatori occasionali.

La prima categoria è costituita dall'insieme dei dipendenti, associati e collaboratori di qualunque sede dell'INFN, la cui procedura di identificazione ufficiale è già avvenuta presso i servizi di Direzione della sede di appartenenza e, di conseguenza, già dotati di credenziali di accesso alle strutture informatiche della propria sede. Questa categoria di utenti potrebbe aver bisogno di connettersi alla rete per accedere ai servizi informatici locali, messi a disposizione dal Servizio di Calcolo (login interattivo, calcolo scientifico, mailbox, servizi in intranet, stampanti, etc.), ivi comprese le stazioni di rete personali cui fanno regolarmente accesso per il lavoro quotidiano.

La seconda categoria è costituita da persone che non lavorano all'interno della struttura, ma la frequentano per assistere alle conferenze tenute su argomenti relativi alla fisica, la visitano come studenti o vi accedono per altri motivi (docenza, corsi, seminari, collaborazioni di breve durata, etc.). Questa categoria di utenti ha bisogno tipicamente di connettersi alla rete internet; talvolta può aver bisogno del servizio locale di stampa e di un servizio di SMTP relay (mail server).

Il progetto TRIP, per queste 2 categorie, prevede la diversificazione delle autorizzazioni e dei permessi di accesso alle varie risorse informatiche, con il triplice obiettivo di consentire gli accessi alle strutture informatiche secondo le diverse esigenze sopra evidenziate, di mantenere più elevato possibile il livello di sicurezza informatica e contemporaneamente di garantire l'accesso alla rete internet secondo il rispetto della legge.

L'architettura del progetto e i suoi elementi costitutivi sono rappresentati nella successiva Figura 2-21. L'infrastruttura di rete wireless è realizzata con circa 40 apparati Cisco Aironet modello 1230, in grado di gestire SSID multipli. Sono configurati in modo da annunciare due diversi SSID (*Service Set IDentifier*), ognuno dei quali corrisponde ad una determinata VLAN:

| SSID | VLAN id | VLAN name |
|---|---|---|
| INFN-dot1x | 131 | LANEsterna |
| INFN-Web | 200 | guests |

**Tabella 2-4: mappatura SSID-VLAN**

Naturalmente affinché le 2 VLAN vengano trasportate sulla rete wired, i link di interconnessione tra gli Aironet e gli switch sono definiti come trunk.

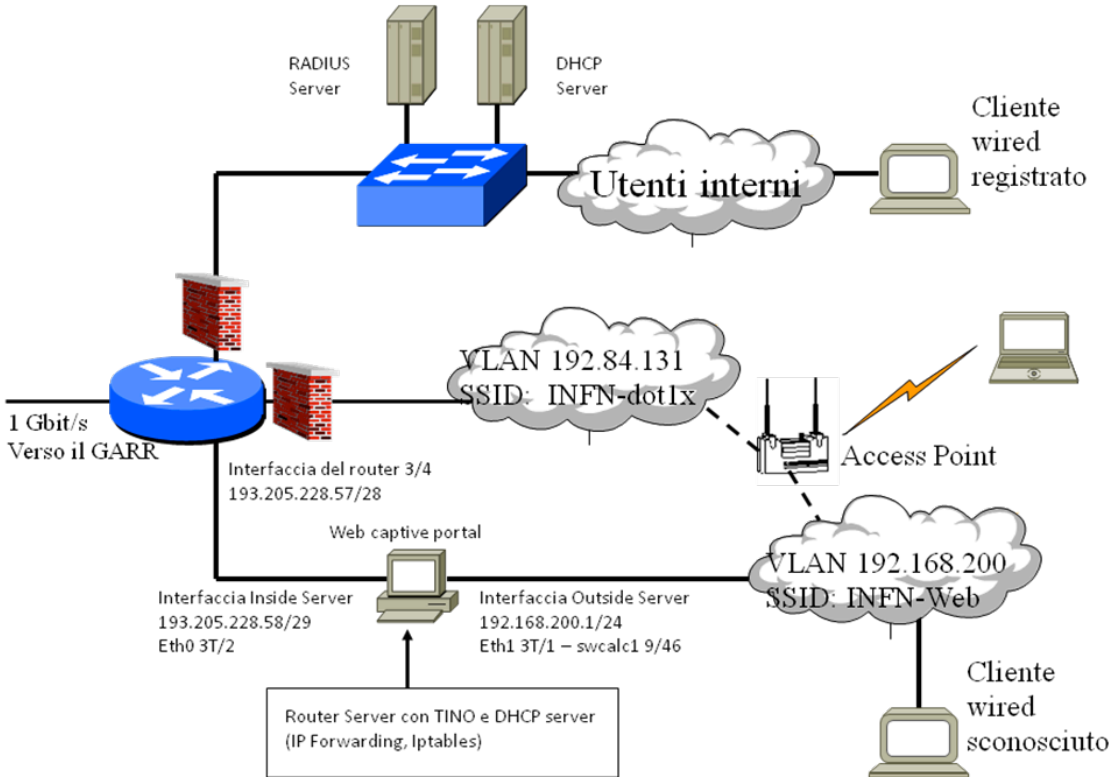



**Figura 2-21: Architettura di rete wired e wireless**

I due SSID sono configurati in modo diverso, il primo (dedicato alla prima delle categorie di utenti viste sopra) per garantire un livello di sicurezza più elevato, secondo la seguente tabella:

| SSID | Authentication | Key Management | Encryption |
|---|---|---|---|
| INFN-dot1x | EAP | WPA o WPA2 | TKIP |
| INFN-Web | Open | None | None |

**Tabella 2-5: Autenticazione e cifratura dei SSID**

Nel caso del SSID *INFN-dot1x*, l'autenticazione viene demandata attraverso un protocollo cifrato e sicuro (EAP-TTLS-PAP) ad un apposito server esterno RADIUS (*Remote Authentication Dial-In User Service*); questo è un server di autenticazione standard che a sua volta è in grado di demandare l'autenticazione al sistema di Autenticazione e Autorizzazione centrale dell'INFN.

Nel caso del SSID *INFN-Web*, il client si trova a valle di un *Captive Portal*, che autorizza il client ad accedere alla rete esterna previa autenticazione demandata dallo stesso Captive Portal al RADIUS server.

### 2.6.2 Metodi di accesso alla rete

Secondo quanto descritto in precedenza, l'accesso alle risorse di rete della struttura può avvenire secondo i seguenti metodi:

1. Accesso alla rete *wired* con un client il cui MAC Address è conosciuto, ovvero registrato nel database VMPS.
2. Accesso alla rete *wired* con un client il cui MAC Address è sconosciuto, ovvero non registrato nel database VMPS.
3. Accesso alla rete wireless protetto da autenticazione 802.1x (connessione alla VLAN 131 con protocollo cifrato).
4. Accesso alla rete wireless protetto da autenticazione tramite web captive portal (connessione alla VLAN 200 con protocollo non cifrato).

***1. Accesso alla rete wired con un client il cui MAC Address è conosciuto***

Questo caso è stato ampiamente illustrato nei paragrafi precedenti. Questo tipo di connessione è riservata ai dipendenti e agli associati dell'INFN (persone già identificate dal Servizio di Direzione) che lavorano stabilmente nei LNF.

Tali utenti, tramite un'opportuna form web, fanno richiesta di registrazione dei loro client previa comunicazione della tipologia e del relativo MAC address.

Il Servizio di Calcolo registra il MAC address nel database VMPS, mappandolo su specifica VLAN in funzione anche della tipologia del client, e attribuisce i settaggi IP impostandoli sul DHCP server (rif. paragrafo 4.2.1).

Da questo momento il client, se impostato per ricevere automaticamente i settaggi IP, è in grado di comunicare sulla rete wired.

***2. Accesso alla rete wired con un client il cui MAC Address è sconosciuto***

In questo caso il client viene mappato sulla VLAN di *fallback* che coincide con quella dei visitatori occasionali, cioè con la VLAN 200 di nome guests.

Questo tipo di connessione può essere richiesta o da visitatori occasionali che vogliono connettere alla rete wired il loro PC portatile, oppure da utenti dei LNF che hanno appena acquistato un nuovo client.

La procedura di accesso ai servizi di rete è descritta al successivo punto ***4.***

***3. Accesso alla rete wireless protetto da autenticazione 802.1x***

Questo metodo di accesso è riservato a dipendenti e associati INFN, in possesso di credenziali di accesso alle strutture informatiche della propria sede.

Come prerequisiti, il client deve essere dotato di interfaccia wireless fornita dei moduli software per l'implementazione del protocollo WPA-TKIP (rif. paragrafo 2.6.3). Inoltre deve essere dotato di modulo software per l'implementazione del protocollo EAP-TTLS (rif. paragrafi 2.6.4 e 2.6.5).

Il client fa richiesta di connessione al SSID *INFN-dot1x* instaurando una connessione wireless cifrata WPA-TKIP. Appena instaurata la connessione viene iniziato uno scambio di informazioni tra il supplicant e il RADIUS server tramite un tunnel cifrato EAP-TTLS, passante per l'access point wireless (rif. Figura 2-22).

Dentro questo tunnel cifrato vengono richieste le credenziali all'utente. Tali credenziali arriveranno al server. Il server RADIUS analizzerà le credenziali (o demanderà ad altri server di autenticazione per l'analisi – rif. paragrafo 4.2.3) e, in caso affermativo, comunica all'access point che l'autenticazione ha avuto successo e distrugge il tunnel.

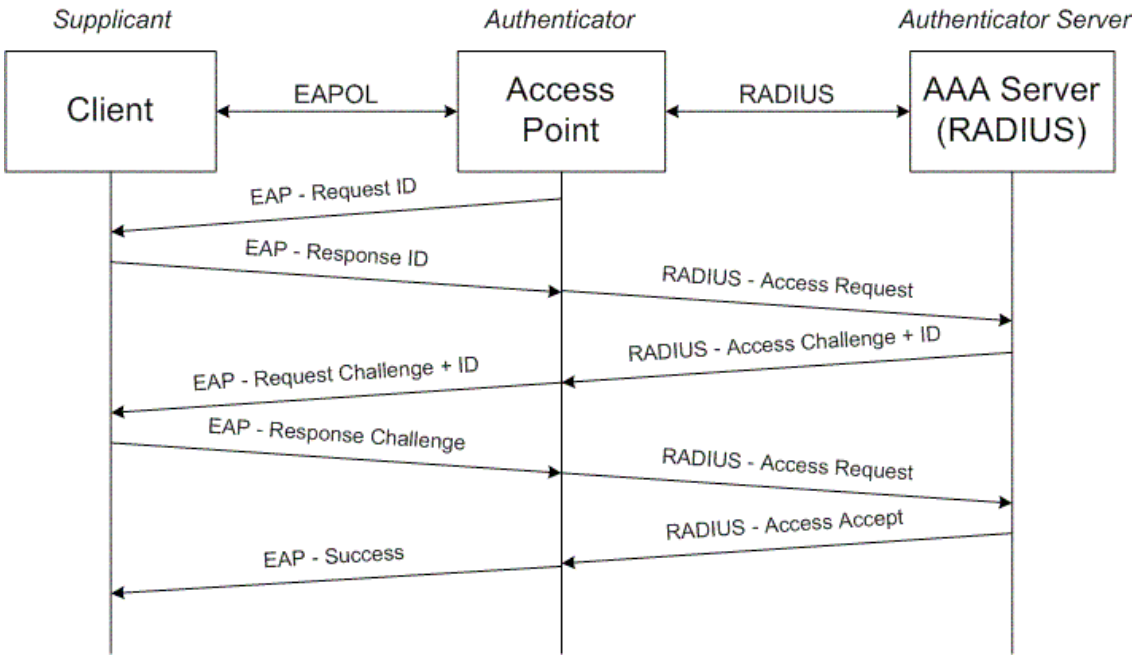



**Figura 2-22: Protocollo di autenticazione 802.1x**

Da questo momento l'access point abilita il client a comunicare sulla rete (sempre in modalità cifrata WPA-TKIP fino all'access point), in particolare sulla VLAN 131; il server DHCP comunicherà al client le relative impostazioni IP (sulla network pubblica 192.84.131.0/24) e il client è finalmente in grado di comunicare anche a livello IP.

Dalla Vlan 131 i firewall consentono l'accesso a tutti i servizi informatici residenti sulla rete dei LNF e anche verso la rete Internet.

### *4. Accesso alla rete wireless protetto da autenticazione tramite web captive portal*

Questo metodo di accesso è riservato ai visitatori occasionali. È un metodo facilmente fruibile in quanto non servono prerequisiti software sul client; è sufficiente un *browser*. D'altra parte è meno sicuro per via della comunicazione in chiaro di tutto il traffico IP (fatta eccezione per l'autenticazione iniziale).

Il metodo si basa sul lavoro di un Web Captive Portal. Per i nostri scopi è stato scelto il captive portal Tino, un software free ed open source, che può essere utilizzato su un server con sistema operativo Linux.

Il client fa richiesta di connessione al SSID *INFN-Web* instaurando una connessione wireless in chiaro. Appena instaurata la connessione, il client viene mappato sulla VLAN 200 e riceve i settaggi IP (sulla network privata 192.168.200.0/24) dal servizio DHCP installato e configurato sullo stesso Captive Portal. In tali settaggi verrà indicato lo stesso Captive Portal come *default-gateway*.

Tuttavia la VLAN 200, ovvero la network 192.168.200.0/24 è completamente isolata dal resto della rete proprio ad opera del Captive Portal che funge anche da router-firewall, e quindi il client in questo momento è in grado di comunicare esclusivamente con gli altri client mappati sulla stessa VLAN 200.

Per ottenere l'accesso alla rete Internet, l'utente dovrà aprire un browser sul client e puntare qualunque URL di Internet. I pacchetti IP di richiesta di connessione http passeranno necessariamente per il default-gateway, ovvero per il Captive Portal. Quest'ultimo li catturerà e li ridirigerà verso il suo servizio httpd interno (con protocollo cifrato http/ssl), al fine di presentare una pagina web con la richiesta di autenticazione tramite certificati digitali a chiave pubblica X.509 oppure tramite l'insermento di credenziali d'accesso (rif. Figura 2-23):

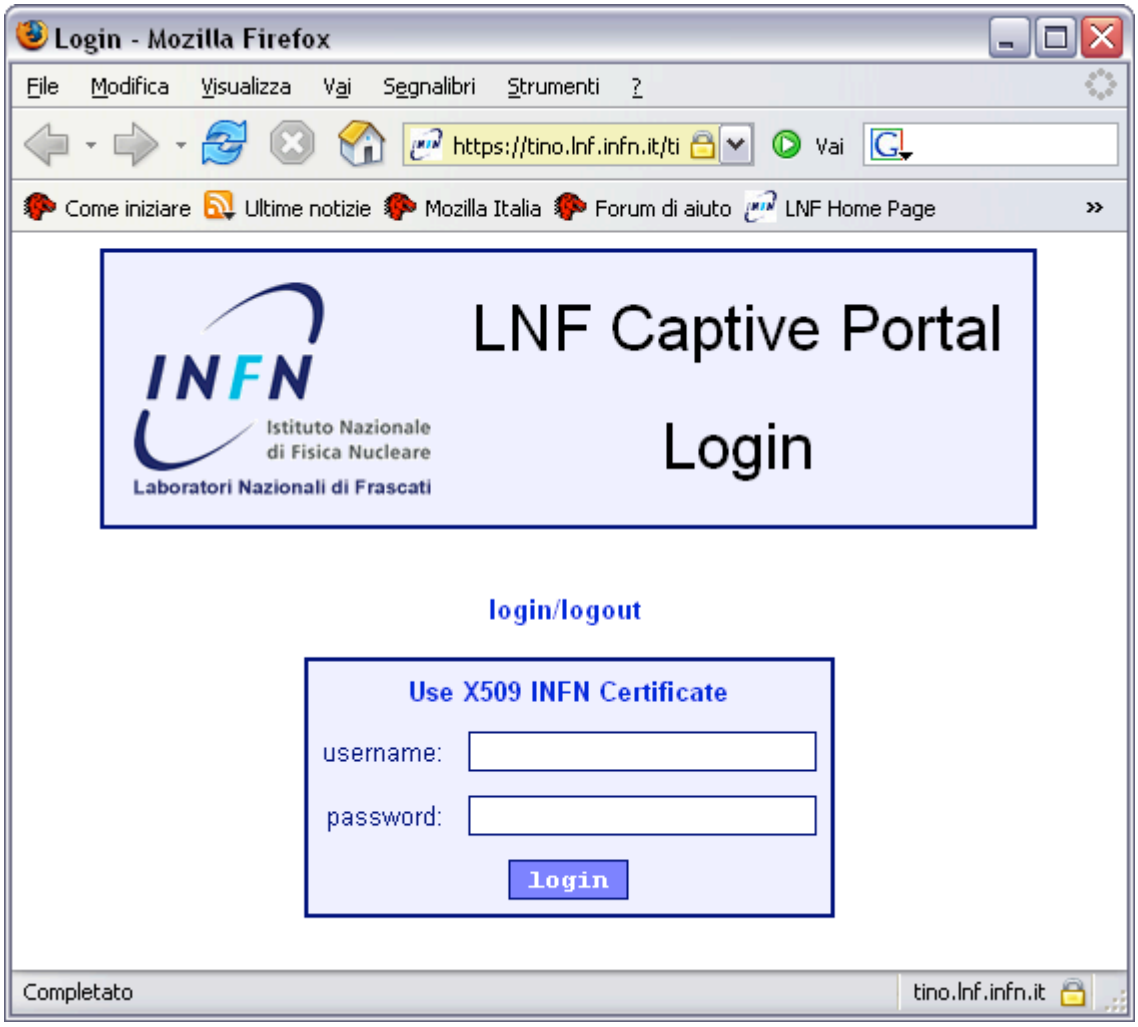



**Figura 2-23: La login page di Tino**

- nel caso in cui l'utente sia dotato di certificto digitale X.509 rilasciato dalla Certification Authority dell'INFN, l'autenticazione viene effettuata direttamente dal Captive Portal;
- nel caso in cui l'utente inserisca delle credenziali di accesso (coppia username e password), l'autenticazione verrà demandata al RADIUS server.

Se l'utente inserisce le giuste credenziali, il RADIUS server restituisce al Captive Portal l'informazione che l'autenticazione è avvenuta con successo e il Captive Portal cambia i filtri sul firewall interno, al fine di garantire l'accesso al mondo esterno per l'indirizzo IP di quel client specifico.

In realtà l'indirizzo del client è su network privata e quindi non è in grado di accedere all'Internet pubblica. Affinché ciò possa avvenire, l'indirizzo sorgente dovrà essere traslato in un indirizzo pubblico attraverso la funzionalità NAT (*Network Address Translation*) svolta dal router di frontiera.

I visitatori occasionali possono ottenere le credenziali di accesso richiedendole ai servizi di segreteria dei LNF. Le segreterie hanno a disposizione un opportuno tool per la gestione del database dei visitatori e dei relativi accessi (rif. paragrafo 4.2.4.1).

### 2.6.3 WPA e TKIP

Wi-Fi Protected Access (WPA) è un protocollo per la sicurezza delle reti senza filo Wi-Fi creato per tamponare i problemi di scarsa sicurezza del precedente protocollo di sicurezza, il WEP. Studi sul WEP avevano individuato delle falle nella sicurezza talmente gravi da renderlo quasi inutile. Il WPA implementa parte del protocollo IEEE 802.11i e rappresenta un passaggio intermedio per il raggiungimento della piena sicurezza. Questa verrà raggiunta quando i dispositivi implementeranno in toto lo standard IEEE 802.11i. Le certificazioni per il WPA sono iniziate nell'aprile del 2003 mentre per l'IEEE 802.11i si è dovuto attendere il giugno del 2004.

WPA è progettato per utilizzare lo standard IEEE 802.1x per gestire l'autenticazione dei client e dei server e la distribuzione di differenti chiavi per ogni utente, sebbene per questioni di compatibilità supporti la precedente gestione a chiave condivisa (PSK). I dati sono cifrati con l'algoritmo di cifratura a stream RC4 con chiave a 128 bit e vettore di inizializzazione a 48 bit.

Una delle modifiche che introducono maggiore robustezza all'algoritmo è la definizione del Temporal Key Integrity Protocol (TKIP). Questo protocollo dinamicamente cambia la chiave in uso e questo combinato con il vettore di inizializzazione di dimensione doppia rispetto al WEP rende inefficaci i metodi di attacco utilizzati contro il WEP.

In aggiunta all'autenticazione e alla cifratura, il WPA introduce notevoli miglioramenti nella gestione dell'integrità. Il CRC utilizzato dal WEP non era sicuro, era possibile modificare il messaggio mantenendo coerente il CRC, anche senza conoscere la chiave. Per evitare il problema, il WPA utilizza un nuovo metodo per verificare l'integrità dei messaggi chiamato "Michael". Questo include un contatore

associato al messaggio per impedire all'attaccante di ritrasmettere un messaggio che sia già stato trasmesso nella rete.

In sostanza il WPA aumenta la dimensione della chiave, il numero delle chiavi in uso, include un sistema per verificare l'autenticità dei messaggi migliore e quindi incrementa la sicurezza della WLAN, rendendola effettivamente analoga a quella di una rete su cavo. La Wi-Fi Alliance (il gruppo che gestisce lo standard Wi Fi) ha dichiarato che utilizzerà il termine WPA2 per identificare i dispositivi che avranno il pieno supporto dello standard IEEE 802.11i.

Wi-Fi Alliance ha introdotto i termini WPA(2)-Personal e WPA(2)-Enterprise per differenziare le due classi di sicurezza fornite dai prodotti. i WPA(2)-Personal utilizzeranno il metodo PSK a chiave condivisa mentre i WPA(2)-Enterprise utilizzeranno un server di autenticazione.

### 2.6.4 EAP-TTLS

Extensible Authentication Protocol (EAP) è un protocollo di autenticazione utilizzato spesso sugli access point e nelle connessioni PPP. Usando EAP, non è l'access point che si preoccupa di autenticare il client, bensì l'autenticazione viene rediretta ad uno specifico server configurato per questo scopo (ad esempio RADIUS).

Definito nella Request for Comments (RFC) 2284 e aggiornato nella RFC 3748 e nella RFC 4017, è uno standard altamente flessibile che può essere implementato in numerose differenti modalità; e 802.1x ha ereditato tale flessibilità per raggiungere svariati obiettivi di sicurezza.

Lo standard 802.1x racchiude un range di metodi di autenticazione EAP, inclusi MD5, TLS, TTLS, LEAP, PEAP, SecurID, SIM e AKA. Ciascuno di questi metodi EAP ha vantaggi e svantaggi a seconda dell'ambiente.

Il particolare il TTLS (Tunnelled Transport Layer Security) è un estensione del TLS ed è stato sviluppato per superare la necessità, generata dal TLS, di certificati lato client (sono invece richiesti certificati lato server).

Il TTLS è un metodo a due passaggi:

1. Nel primo, un algoritmo asimmetrico basato sulle chiavi del server è utilizzato per verificare l'identità del server e per creare il tunnel di crittazione simmetrica.
2. Il secondo passaggio riguarda la verifica dell'identità del client utilizzando un secondo metodo di autenticazione tramite il tunnel di crittazione simmetrica per l'attuale negoziazione dell'autenticazione.

Questo secondo metodo di autenticazione utilizzato con il tunnel può essere un tipo di EAP (spesso MD5) o un metodo di vecchio tipo come PAP (come nel caso di TRIP), CHAP, MS-CHAP, o MS-CHAP V2. Il tunnel a crittazione simmetrica del TTLS è utilizzato solo per proteggere il metodo di autenticazione del client. Una volta verificato, il tunnel collassa.

### 2.6.5 Requisiti del client

Per l'accesso alla rete wireless protetto da autenticazione 802.1x il client deve essere dotato di interfaccia wireless fornita dei moduli software per l'implementazione del protocollo WPA-TKIP.

Questo requisito non rappresenta un problema in quanto, oggi, tutte le interfacce wireless sono dotate degli opportuni moduli software per l'implementazione del protocollo WPA-TKIP o anche WPA2-TKIP. Si può incontrare qualche limitazione nelle interfacce di vecchia generazione, costruite da società minori che non hanno avuto grande penetrazione nel mercato.

Il client, inoltre, deve essere dotato di modulo software per l'implementazione del protocollo EAP-TTLS.

Questo requisito rappresenta un problema per i client di tipo MS Windows. Infatti, il protocollo EAP-TTLS è implementato in maniera nativa nei client di tipo Apple Macintosh (Sistema Operativo MAC OS X), e anche nei client Linux (in particolare nella distribuzione Ubuntu, orientata alla facilità d'uso del desktop); tuttavia la Microsoft ha scelto di implementare nativamente solo il protocollo EAP-TLS che si basa esclusivamente sull'utilizzo dei certificati digitali a chiave pubblica X.509, ignorando completamente il protocollo EAP-TTLS.

Per sopperire a questa mancanza, si è scelto di utilizzare un software free di nome SecureW2 dell'Alfa-Ariss. Questo software permette di integrare i metodi di autenticazione previsti dalla distribuzione Microsoft con il metodo scelto dal progetto TRIP. D'altra parte EAP-TLS è stato scartato dal progetto perché nella distribuzione MS Windows veniva utilizzato senza la possibilità di un'adeguata protezione del certificato digitale personale sul client.

Tuttavia questa scelta ha un impatto non trascurabile sul Servizio di Calcolo se si considera che i client di accesso alla rete sono prevalentemente Windows.

# 3. L'infrastruttura di storage dei LNF

In ambito informatico con il termine *Storage* si identificano i dispositivi hardware, i supporti per la memorizzazione, le infrastrutture ed i software dedicati alla memorizzazione non volatile di grandi quantità di informazioni in formato elettronico.

Il mercato dello storage è quel settore di mercato ICT che indirizza le esigenze di memorizzazione di grandi quantità di dati. Esso si può dividere nei seguenti ambiti applicativi:

- *file sharing*, ossia tutte le esigenze di condivisione di informazioni tra diversi server e tra i server e i personal computer;
- *data backup*, ossia tutte le esigenze di creazione di copie delle informazioni da riutilizzare nel caso la versione originale venga danneggiata o persa.

Con il termine memorizzazione non volatile si intende la possibilità di immagazzinare delle informazioni in maniera persistente con una buona probabilità che l'informazione rimanga inalterata per un ragionevole lasso di tempo.

Per poter memorizzare in maniera persistente le informazioni in formato digitale binario (sequenza di bit) è necessario avere un supporto fisico con le seguenti caratteristiche:

- sul supporto possono essere scritte almeno una volta sequenze di bit;
- le sequenze di bit scritte rimangono inalterate a meno di una specifica operazione di modifica;
- le sequenze di bit possono essere lette un numero elevato di volte senza alterarle.

## *3.1 Dispositivi hardware di memorizzazione*

Esistono molti supporti fisici che rispondono a queste caratteristiche, ma nella storia dell'informatica solo alcuni si sono affermati e sono tuttora utilizzati:

- Dischi magnetici (*hard disk* e *floppy disk*);
- Nastri magnetici;
- Dischi ottici;

### 3.1.1 Dischi magnetici (*hard disk*)

Il disco rigido o hard disk (anche chiamato disco fisso) è un dispositivo utilizzato per la memorizzazione a lungo termine dei dati in un computer. È costituito fondamentalmente da uno o più dischi in alluminio o vetro, rivestiti di materiale ferromagnetico in rapida rotazione e da due testine per ogni disco (una per lato), le quali, durante il funzionamento "volano" alla distanza di poche decine di nanometri dalla superficie del disco leggendo e scrivendo i dati. La testina è tenuta sollevata

dall'aria mossa dalla rotazione stessa dei dischi che può superare i 15.000 giri al minuto; attualmente i valori standard di rotazione sono 5.400, 7.200, 10.000 e 15.000 giri al minuto.

I dischi rigidi moderni hanno capacità e prestazioni enormemente superiori a quelle dei primi modelli, ma poiché nel frattempo la velocità e le prestazioni delle memorie ad accesso casuale (RAM e ROM) sono aumentate molto di più, la loro velocità nella lettura e scrittura dei dati restano comunque di diversi ordini di grandezza al di sotto delle prestazioni della RAM e della componentistica a stato solido che equipaggia un computer. Per questo motivo il disco rigido è spesso la causa principale del rallentamento di un computer soprattutto quando, a causa di una memoria RAM inferiore alla memoria virtuale richiesta dai programmi in esecuzione, il sistema operativo è costretto ad effettuare un gran numero di operazioni di *swap* tra il disco e la memoria centrale.

Le caratteristiche principali di un disco rigido sono:

- la capacità
- il tempo di accesso
- la velocità di trasferimento

La capacità è in genere espressa in gigabyte (GB). I produttori usano i gigabyte metrici, invece delle approssimazioni per potenze di due usate per la memoria. Questo significa che la capacità di un disco rigido è in realtà un poco più piccola di quella di un modulo di memoria con la stessa capacità, e lo scarto aumenta all'aumentare delle dimensioni. Quando la capacità è espressa in GB, il fattore di correzione è di $(1000/1024)^3$, pari a circa 0,93, per cui un disco rigido da 320 GB ha una capacità effettiva di circa 298 GiB. Attualmente i dischi rigidi si trovano in vendita con capacità comprese tra 160 GB e 1,5 TB. La capacità può essere aumentata incrementando la densità con cui le informazioni vengono memorizzate sui dischi, oppure usando dischi più grandi o impiegandone un numero maggiore.

Il tempo di accesso è la variabile più importante nel determinare le prestazioni di un disco rigido. Conoscendo il modello, si può risalire facilmente ai dati tecnici dell'unità, compreso il tempo di accesso; purtroppo molti produttori di computer non menzionano questo dato, e a volte nemmeno la marca, né il modello. Si tratta del tempo medio necessario perché un dato, residente in un punto casuale del disco, possa essere reperito. Il tempo impiegato dipende dalla velocità della testina a spostarsi sulla traccia dove risiede il dato e dalla velocità di rotazione del disco; maggiore è la velocità e più breve è il tempo impiegato dal dato a ripassare sotto la testina nel caso questa non fosse arrivata in tempo sul dato, durante la rotazione precedente (latenza rotazionale). I produttori cercano perciò di realizzare testine sempre più leggere (che possono spostarsi più in fretta perché dotate di minore inerzia) e dischi che girano più velocemente. Il tempo di accesso tipico per un disco rigido di media qualità è attorno ai 10 millisecondi. Per uno ad alte prestazioni (15.000 giri) è di 3 o 4 millisecondi.

La velocità di trasferimento è la quantità di dati che il computer è teoricamente in grado di leggere o scrivere sul disco in un determinato tempo (in genere si prende un secondo come riferimento). Usare dischi che ruotino più velocemente o incrementare la densità di memorizzazione porta ad un miglioramento diretto della velocità di trasferimento. C'è da dire che, a parte casi particolari, la velocità di trasferimento teorica viene raramente raggiunta e il tempo di accesso è quello che maggiormente influenza le prestazioni di un disco rigido.

Oltre alle tre viste sopra, altre caratteristiche influenzano in misura minore le prestazioni di un disco rigido. Tra queste:

- il buffer di memoria
- la velocità dell'interfaccia

Il buffer è una piccola memoria cache (in genere di alcuni megabyte) posta a bordo del disco rigido, che ha il compito di memorizzare gli ultimi dati letti o scritti dal disco. Nel caso in cui un programma legga ripetutamente le stesse informazioni, queste possono essere reperite nel buffer invece che sul disco. Essendo il buffer un componente elettronico e non meccanico, la velocità di trasferimento è molto maggiore, nel tempo, la capacità di questa memoria è andata sempre aumentando, attualmente 32 MB sono una dimensione abbastanza usuale.

L'interfaccia di collegamento tra il disco rigido e la scheda madre (o, più specificatamente, il controller) può influenzare le prestazioni perché specifica la velocità massima alla quale le informazioni possono essere trasferite da o per il disco. Le moderne interfacce tipo ATA133, Serial ATA, SCSI, SSA o Fibre Channel possono trasferire centinaia di megabyte per secondo, molto più di quanto qualunque singolo disco fisso possa fare, e quindi l'interfaccia non è in genere un fattore limitante. Il discorso può cambiare nell'utilizzo di più dischi in configurazione RAID, nel qual caso è importante utilizzare l'interfaccia più veloce possibile, come per esempio Fibre Channel da 4 Gb/s o da 8Gb/s.

***Tempo di accesso a disco***

Il tempo di accesso a disco è influenzato da quattro fattori:

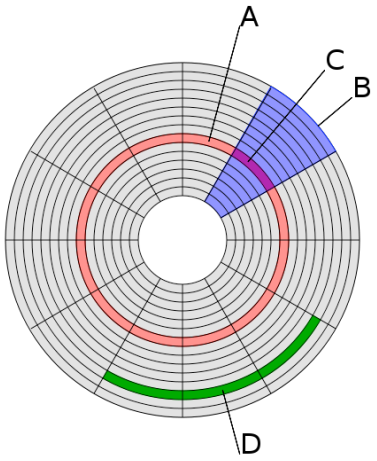
- *Controller Overhead*: è il tempo necessario alla gestione dei dati e l'invio dell'opportuno *interrupt*; è il tempo in assoluto minore;
- *Seek time* (tempo di ricerca): è il tempo necessario a spostare la testina sulla traccia; è il fattore più critico poiché si tratta di un movimento meccanico e non di un impulso elettrico; questo fa sì che non si possa scendere al di sotto di qualche decina di millisecondo;
- *Assessment time* (tempo di assestamento): è il tempo necessario all'assestamento della testina sulla traccia dopo lo spostamento; spesso viene inglobato nel Seek time;
- *Latency time* (tempo di latenza): (anche rotational latency) è il tempo necessario perché, a causa della rotazione del disco, l'inizio del settore desiderato arrivi a trovarsi sotto la testina; ovviamente dipende dalla velocità di rotazione; per esempio con una velocità (tipica) di 5400 giri/min. il tempo di latenza massimo è di circa 11 millisecondi;

- *Transfer time* (tempo di trasferimento): è il tempo necessario al settore per passare sotto la testina, tempo durante il quale il settore viene letto o scritto.

Tempo di accesso: ControllerOverhead + SeekTime + Latency + TransferTime

***Organizzazione fisica della memorizzazione dei dati***

I dati sono generalmente memorizzati su disco seguendo uno schema di allocazione fisica ben definito in base al quale si può raggiungere la zona dove leggere/scrivere i dati sul disco. Uno dei più diffusi è il cosiddetto CHS acronimo per il termine inglese *Cylinder/Head/Sector* (Cilindro/Testina/Settore);




Struttura della superficie di un piatto:
A) Traccia
B) Settore
C) Settore di una traccia
D) Cluster, insieme di settori contigui

in questa struttura i dati sono memorizzati avendo come indirizzo fisico un numero per ciascuna delle seguenti entità fisiche:

***Piatto***
un disco rigido si compone di uno o più dischi paralleli, di cui ogni superficie, detta "piatto" e identificata da un numero univoco, è destinata alla memorizzazione dei dati.

***Traccia***
ogni piatto si compone di numerosi anelli concentrici numerati, detti tracce ciascuna identificata da un numero univoco.

***Cilindro***
l'insieme di tracce alla stessa distanza dal centro presenti su tutti i dischi è detto cilindro. Corrisponde a tutte le tracce aventi il medesimo numero, ma diverso piatto.

***Settore***
ogni piatto è suddiviso in settori circolari, ovvero in "spicchi" radiali uguali ciascuno identificato da un numero univoco.

***Blocco***
L'insieme di settori posti nella stessa posizione in tutti i piatti.

***Testina***
Su ogni piatto è presente una testina per accedere in scrittura o in lettura ai dati memorizzati sul piatto; la posizione di tale testina è solidale con tutte le altre sugli altri

piatti. In altre parole, se una testina è posizionata sopra una traccia, tutte le testine saranno posizionate nel cilindro a cui la traccia appartiene.

### 3.1.2 Dispositivi a nastro magnetico

Il nastro magnetico è un supporto destinato alla memorizzazione di dati che consiste in una sottile striscia in materiale plastico, rivestita di un materiale magnetizzabile.

I sottosistemi a nastro più moderni utilizzano bobine di film dalle dimensioni molto contenute (dal momento che la densità di registrazione è aumentata) e tali bobine sono collocate dentro una *cartridge* (cartuccia) per proteggere il nastro e facilitarne la manipolazione. Esempi di formati delle cartridge più diffusi sono: DAT, Exabyte ed LTO.


**Figura 3-1: Cartridge (da sinistra): Dat, Exabyte, LTO**

Nei formati più utilizzati, i dati sono scritti sul nastro a blocchi spaziati tra loro, ciascun blocco viene scritto in una singola operazione, mentre il nastro si muove uniformemente durante la scrittura.

Tuttavia, dal momento che la velocità alla quale il nastro viene letto o scritto non è deterministica, un'unità a nastro è progettata per far fronte alla differenza tra la velocità alla quale i dati vengono letti o scritti e quella dei dati inviati o richiesti dal sistema che lo controlla.

Sono molti i metodi che possono essere utilizzati, singolarmente o combinandoli tra loro per garantire il funzionamento ottimale annullando l'effetto di tali differenze nel trasferimento dei dati. Un ampio buffer di memoria, spesso un vero e proprio *spool*, come pure un controllo meccanico: il *drive* può essere arrestato, tornare leggermente indietro e riavviato. Inoltre il sistema che controlla il drive può comandare l'impiego di una diversa dimensione del blocco che viene scritto o letto sul nastro per ciascuna singola operazione.

La ricerca di un compromesso tra la dimensione del blocco, l'ampiezza del buffer dei dati, la percentuale di nastro perso nella spaziatura tra i blocchi e la velocità incidono in maniera consistente sulla quantità dei dati complessivi che vengono letti e scritti sul nastro.

Molte unità a nastro in uso negli anni più recenti includono il supporto per un qualche tipo di compressione dei dati. Vengono utilizzati a questo scopo diversi algoritmi, che forniscono risultati piuttosto simili: LZ (Most), IDRC (Exabyte), ALDC (IBM, QIC) e DLZ1 (DLT). Gli algoritmi utilizzati non sono in realtà tra i più efficienti disponibili oggi, per questo motivo la soluzione migliore per l'utilizzo per le applicazioni di backup è ottenibile disabilitando la compressione disponibile nell'unità a nastro e utilizzando invece un software di compressione.

La recente diminuzione del costo dei dischi fissi e le migliorie costruttive che ne hanno determinato un generale aumento di affidabilità hanno via via diminuito il ricorso al nastro magnetico. Questo tuttavia rimane in uso in molti centri di elaborazione dati, soprattutto per ragioni di gestione di archivi già precostituiti e per il costo per bit piuttosto basso.

Infatti i nastri magnetici hanno il vantaggio di essere il supporto più economico a parità di quantità di informazione memorizzata, ma hanno lo svantaggio di permettere solo l'accesso sequenziale. Queste due caratteristiche li rendono lo strumento ideale per il backup e l'archiviazione di grandi quantità di dati che richiedono di essere utilizzate raramente.

La gamma di prodotti è molto ampia e spazia dalle unità nastro a singola cassetta che possono essere montate all'interno di server o sono collegabili agli stessi tramite porta USB, fino alle tape library che possono gestire anche centinaia di cassette e che vengono, normalmente, collegate ai server tramite SAN.

Una grande *tape library* è costituita da migliaia di cassette. Un braccio meccanico (detto *piker*) dotato di lettore ottico è in grado di riconoscere le cassette tramite i codice a barre.

**Figura 3-2: Librerie a nastro (da sinistra): IBM 3494, StorageTek L1400**

### 3.1.3 Optical juke box

Gli *optical juke box* sono dispositivi di memorizzazioni basati su dischi ottici. Dal momento che i dischi ottici non sono facilemnte riscrivibili, tali apparati vengono normalmente utilizzati in modalità *write once read many* (WORM), ossia vengono utilizzati per memorizzare informazioni che non dovranno più essere modificate. Inoltre, dato che i dischi ottici possono mantenere inalterati i dati per un periodo di

tempo più lungo rispetto ai nastri (30-50 anni contro i 10-15 di un nastro), gli optical juke box vengono utilizzati per memorizzare i dati che per obblighi di legge devono essere mantenuti per un periodo di tempo molto lungo.

Un altro campo specifico di utilizzo è quello medico: gli optical juke box possono essere utilizzati per memorizzare le cartelle cliniche dei pazienti con tutti i file multimediali associati (lastre, ecografie, scansioni, etc.).

### 3.1.4 Disk array

I dischi magnetici sono il supporto ideale per tutte le applicazioni di *file sharing* dove è richiesto di poter accedere on-line alle informazioni memorizzate.

Le soluzione più semplice basata sui dischi è il *Direct Attached Storage* (DAS): ogni server è "direttamente collegato" con unità a disco dedicate che contegono sia le applicazioni che i dati. Questo tipo di soluzione necessita di allocare spazio disco in maniera dedicata al singolo server; al crescere del numero di server possono insorgere due problemi:

- **efficienza:** ogni server può utilizzare solo il proprio spazio disco con difficoltà di ottimizzazione d'uso dello spazio disco complessivo (il rischio che i dischi di un server siano utilizzati al limite delle capacità degli stessi mentre quelli di un altro server siano sotto-utilizzati).
- **complessità:** un elevato numero di dischi comporta maggiori costi di gestione.

Quando il numero di server aumenta è più conveniente mettere a fattor comune lo storage utilizzando apparati dedicati alla memorizzazione che siano accessibili a tutti i server: questi apparati, composti da una serie di dischi (*disk array*), hanno la capacità di condividere i dischi tra i vari server, che li utilizzano come se fossero dischi propri, ma danno la possibilità di allocarli in maniera flessibile a seconda delle esigenze. Inoltre è possibile fare una copia di backup dell'intero *disck array*.

I *disk array* alloggiano al loro interno un controllore dotato di una certa intelligenza (*Disk Array Controller*) in grado di gestire i dischi e di servirli ai sistemi esterni. Nei casi migliori il *disk array controller* è duale e ridondato per aumentare l'affidabilità e la disponibilità del *disk array*.

***Disk Array Controller***

Il *disk array controller* è un apparato che gestisce i dischi fisici e li presenta ai computer come *logical units*. Esso implementa sempre il *RAID hardware*, per cui viene spesso definito *RAID controller*. Inoltre fornisce una *cache* di memoria addizionale per velocizzare le operazioni di lettura e scrittura.

Il *disk array controller* dispone di interfaccia di *front-end* e di *back-end*.

- L'interfaccia di *back-end* comunica con i dischi. I protocolli usati sono ATA, SATA, SCSI, SAS, FC (Fibre Channel)
- L'interfaccia di *front-end* comunica con i computer (dotati di opportuno *Host Adapter*) e usa più comunemente i protocolli SCSI, SAS e FC

Un singolo controller può usare protocolli di comunicazione differenti per il *back-end* e per il *front-end*. I più comunemente usati a livello *enterprise* sono *Fibre Channel* per il *front-end* e SATA o Fibre Channel per il *back-end*.

### *RAID*

Un *Redundant Array of Independent Disks* ("insieme ridondante di dischi indipendenti", RAID) è un sistema informatico che usa un insieme di dischi rigidi per condividere o replicare le informazioni. I benefici del RAID sono di aumentare l'integrità dei dati, le prestazioni e la tolleranza ai guasti, rispetto all'uso di un disco singolo.

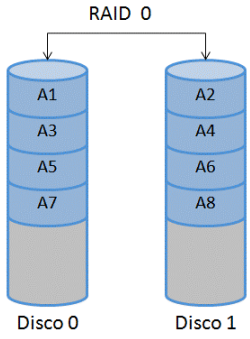
Nel suo livello più semplice, il sistema RAID permette di combinare un insieme di dischi in una sola unità logica (*logical unit*). In questo modo il sistema operativo, invece di vedere differenti dischi, ne vede solamente uno. Il RAID è tipicamente usato nei server, e di solito è implementato con dischi di identica capacità. Con il calo del costo dei dischi rigidi e con il diffondersi della tecnologia RAID nei chipset delle schede madri, il RAID è spesso offerto come opzione sia sui computer di fascia alta sia su quelli usati da utenti domestici, specialmente se dedicati a compiti che richiedono un grande immagazzinamento di dati. Il RAID è sempre usato nei *disk array*, implementato in hardware dal *controller*.

Le specifiche originali suggerivano un diverso numero di "livelli di RAID", o combinazioni di dischi. Ogni combinazione aveva dei vantaggi e degli svantaggi. Con il passare degli anni, sono nate diverse implementazioni del concetto di RAID. La maggior parte differiscono sostanzialmente nell'implementazione dei livelli RAID ideati inizialmente. Questo può portare spesso a confusione, poiché un'implementazione RAID-5 può essere molto diversa da un'altra. RAID-3 e RAID-4 sono spesso confusi o scambiati tra loro.

La vera definizione di RAID è stata oggetto di dibattito nel corso degli anni. L'uso del termine "ridondante" porta a molte discussioni se il RAID-0 sia "vero" RAID. Si considera RAID ogni sistema che sviluppa il concetto base di RAID di ricombinare lo spazio fisico di dischi diversi per lo scopo di aumentare l'affidabilità o le prestazioni del sistema nel suo complesso.

### *RAID 0 (STRIPING)*

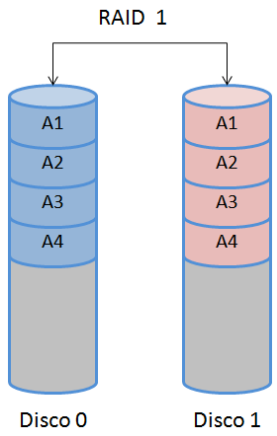
Il sistema RAID 0 divide i dati equamente tra due o più dischi con nessuna informazione di parità o ridondanza (operazione detta di striping). Bisogna notare che il RAID-0 non era presente tra i livelli RAID originari, e che non è ridondante. RAID-0 è usato generalmente per aumentare le prestazioni di un sistema, anche se è molto utile per creare un piccolo numero di grandi dischi virtuali da un grande numero di piccoli dischi fisici.



Sebbene il RAID-0 non sia indicato tra i livelli RAID originari, in un sistema ideale di tipo RAID-0 le operazioni di Input/Output (I/O) si dividerebbero in blocchi di dimensioni uguali e si applicherebbero equamente su tutti i dischi. Le implementazioni di sistemi RAID-0 su più di due dischi sono possibili, ma l'affidabilità di un dato sistema RAID-0 è uguale all'affidabilità media dei dischi diviso per il numero di dischi presenti. Quindi l'affidabilità, misurata come tempo medio tra due guasti (MTBF) è inversamente proporzionale al numero degli elementi; cioè un sistema di due dischi è affidabile la metà di un disco solo. La ragione per la quale questo succede è che il file system è diviso tra tutti i dischi. Quando un drive si guasta, il file system non può gestire una perdita di dati così grande visto che i dati sono divisi tra tutti i dischi. I dati possono essere spesso recuperati con qualche strumento, anche se saranno sicuramente incompleti e corrotti.

### *RAID 1 (MIRRORING)*

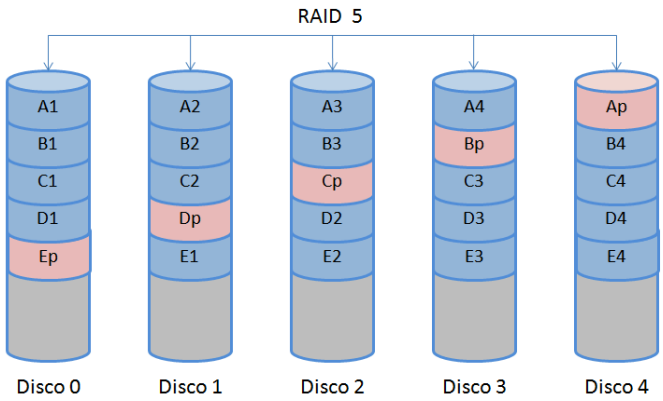
Il sistema RAID-1 crea una copia esatta (*mirror*) di tutti i dati su due o più dischi. È utile nei casi in cui la ridondanza è più importante che usare tutti i dischi alla loro massima capacità: infatti il sistema può avere una capacità massima pari a quella del disco più piccolo. In un sistema tipico, formato da due dischi, l'affidabilità aumenta di molto rispetto al sistema a disco singolo, ed è possibile anche avere più di una copia dei dischi. Un'implemetazione accorta del RAID-1 aumenta anche le prestazioni, potendo parallelizzare molte operazioni di Input/Output.




Una pratica comune è di creare un mirror extra di un disco (detto anche "Business Continuance Volume" o "BCV") che può essere diviso dal sistema RAID originario ed essere usato in maniera indipendente. In alcune implementazioni, questi mirror aggiuntivi possono essere divisi e aggiunti in maniera incrementale, invece di richiedere una ricostruzione completa del RAID.

### *RAID 5*

Un sistema RAID-5 usa una divisione dei dati a livello di blocco con i dati di parità distribuiti tra tutti i dischi appartenenti al RAID. Questa è una delle implementazioni più popolari, sia in hardware che in software. Potenzialmente ogni sistema di storage permette il RAID-5 tra le sue opzioni.



Nell'esempio soprastante, una richiesta al blocco "A1" potrebbe essere evasa dal disco 1. Una simultanea richiesta per il blocco B1 dovrebbe aspettare, ma una richiesta simultanea per il blocco B2 potrebbe essere evasa in contemporanea.
A1, B2, etc. rappresentano ognuno un blocco di dati.

Ogni volta che un blocco di dati (chiamato a volte *chunk*) deve essere scritto nel sistema di dischi, un blocco di parità viene generato all'interno della *stripe*. (Un blocco è spesso composto da molti settori di disco. Una serie di blocchi consecutivi è chiamato *stripe*). Se un altro blocco, o qualche porzione dello stesso blocco, è scritta nella stessa *stripe*, il blocco di parità viene ricalcolato e riscritto. Il disco usato per memorizzare le parità viene modificato tra una stripe e la successiva; in questo modo si riescono a distribuire i blocchi di parità.

Bisogna notare che il blocco di parità non viene letto quando si leggono i dati da disco, visto che rappresenterebbe un sovraccarico non necessario e diminuirebbe le performance. Il blocco di parità è letto, invece, quando la lettura di un settore dà un errore CRC. In questo caso, il settore nella stessa posizione relativa nei blocchi di dati rimanenti della stripe, insieme al blocco di parità, vengono usati per ricostruire il blocco mancante. In questo modo l'errore di CRC viene nascosto al computer chiamante. Nella stessa maniera, se un disco dovesse danneggiarsi all'interno del sistema, i blocchi di parità dei dischi rimanenti sono combinati matematicamente "al volo" con i blocchi dati rimasti per ricostruire i dati del disco guasto.

Questa procedura viene chiamata di solito *Interim Data Recovery Mode*. Il computer principale non è messo al corrente che un disco si è danneggiato. Le letture e scritture verso il sistema di dischi avvengono tranquillamente come prima, sebbene con qualche calo di prestazioni.

In un sistema RAID-5 che ha un solo blocco di parità per stripe, la rottura di un secondo disco comporta la perdita di tutti i dati presenti nel sistema.

Il numero minimo di dischi in un volume RAID-5 è 3, mentre il numero massimo di dischi è teoricamente illimitato, ma una pratica comune è di mantenere il numero massimo di dischi a 14 (o meno) per le implementazioni che hanno solo un blocco di parità per stripe. Le ragioni per questo limite sono che la probabilità che due dischi del sistema si rompano in successione, cresce con il crescere del numero di dischi. Quando il numero di dischi in un sistema RAID-5 cresce, il MTBF del sistema nel suo complesso può persino diventare minore di quello di un singolo disco. Questo succede quando la probabilità che si rompa un secondo disco degli $N - 1$ rimanenti, tra il tempo di accorgersi, sostituire e ricreare il primo disco guasto, diventi maggiore della probabilità che un singolo disco si guasti.

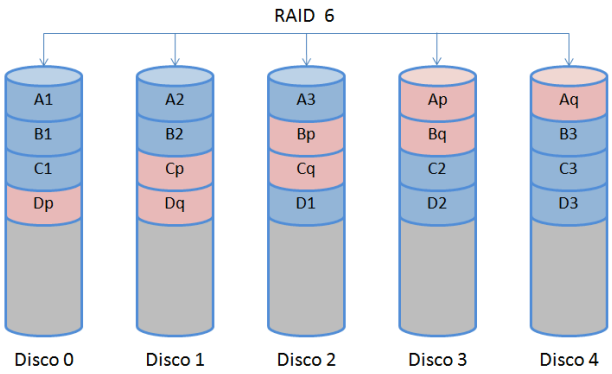
Bisogna ricordare che più dischi insieme aumentano il calore, che abbassa il vero MTBF di ogni disco. Inoltre, i dischi di uno stesso gruppo comprati nello stesso periodo potrebbero raggiungere la fine della loro vita insieme, abbassando in maniera significativa il MTBF del sistema. È buona norma, non sempre seguita dai produttori di server, inserire in RAID dischi identici, ma provenienti da partite differenti, ovvero con numeri di serie e/o date e luogo di produzione distinti e lontani. È semplicemente falsa e - come abbiamo visto - anche controproducente l'affermazione, che spesso si trova in

alcune aste online o su alcuni mercatini, che vorrebbe una coppia di dischi con numeri di serie contigui come perfettamente adatta all'utilizzo in RAID.

Nelle implementazioni con più di 14 dischi, o in situazioni dove è necessaria grande ridondanza dei dati, viene usata spesso una implementazione RAID-5 con doppia parità (detta anche RAID-6), che riesce a gestire il guasto contemporaneo di due dischi.

***RAID 6***

Un sistema RAID-6 usa una divisione a livello di blocchi con i dati di parità distruibuiti due volte tra tutti i dischi. Non era presente tra gli originali livelli RAID.




Nel RAID-6, il blocco di parità viene generato e distribuito tra due stripe di parità, su due dischi separati, usando differenti algoritmi per il calcolo degli stripe di parità nelle due "direzioni".

Il numero minimo di dischi in un volume RAID-6 è 4 ed è evidentemente più ridondante del RAID-5, ma è molto inefficiente quando viene usato in un numero limitato di dischi. Per limitare o annullare l'inefficienza del RAID-6, nei sistemi *enterprise* vengono spesso dedicati processori ad-hoc per il calcolo della parità (ASIC).

I sistemi RAID-2, RAID-3 e RAID-4 non sono trattati in quanto desueti.

***LUN***

Le dimensioni dei dischi sono attualmente dell'ordine di 0,6 Terabyte nel caso di dischi a 10.000 RPM o 15.000 RPM, o di 1,5 o 2 Terabyte, nel caso di dischi a 7200 RPM.

Nei casi più comuni, il numero di dischi che compone un volume RAID-5 va da 5 a 8. Le dimensioni complessive del singolo volume, al netto della parità, si aggirano intorno a qualche Terabyte. Tuttavia spesso la necessità di spazio dedicato ad uno scopo specifico, o da destinare al singolo host, è molto inferiore a tali valori. Per questo motivo i Disk Array Controller sono in grado di suddividere lo spazio complessivo di un volume RAID in partizioni di dimensioni più piccole, scelte dall'amministratore, dette LUN (*Logical Unit Number*) e di esportarle agli host come dei veri e propri hard disk virtuali.

## 3.2 Storage Area Network

Una Storage Area Network (SAN) è una rete ad alta velocità (generalmente vari Gigabit/s) di dispositivi di memorizzazione di massa condivisi; un dispositivo di memorizzazione di massa (storage) è una macchina che può essere composta da uno o più dischi per contenere dati. I protocolli attualmente più diffusi per l'utilizzo degli storage sono FC (Fibre Channel) ed iSCSI (Internet SCSI).

Più precisamente, il dizionario tecnico pubblicato dalla Storage Networking Industry Association (SNIA) definisce una rete SAN nei seguenti termini:

- *Una rete il cui scopo principale è il trasferimento di dati tra sistemi di computer ed elementi di storage e tra elementi di storage. Una rete SAN consiste in un'infrastruttura di comunicazione, che fornisce connessioni fisiche e in un livello di gestione, che organizza connessioni, elementi di storage e sistemi di computer in modo da garantire un trasferimento di dati sicuro e robusto.*

***Architettura SAN***

Un'architettura SAN lavora in modo che tutti i dispositivi di memorizzazione siano disponibili a qualsiasi server della rete LAN o MAN di cui la SAN in questione fa parte; una SAN può essere anche condivisa fra più reti interconnesse, anche di natura diversa: in tal caso uno dei server locali fa da ponte fra i dati memorizzati e gli utenti finali. Uno dei vantaggi di un'architettura di questo tipo è che tutta la potenza di calcolo dei server è utilizzata per le applicazioni, in quanto i dati non risiedono direttamente in alcuno di questi.

In una rete SAN le periferiche di storage sono connesse ai server attraverso una topologia costituita essenzialmente da canali - solitamente in fibra ottica - e da hub, switch router e bridge che in teoria consente la coesistenza di sistemi e dispositivi di storage di natura eterogenea, sebbene nella pratica gli aspetti di interoperabilità costringano spesso a creare reti SAN omogenee. Questo permette di evitare un sovraccarico della rete dato che tutto il traffico è gestito da questi dispositivi.

Normalmente una SAN utilizza dischi collegati con una struttura di tipo RAID per migliorare le prestazioni e aumentare l'affidabilità del sistema (i Disk Array descritti in precedenza).

***Vantaggi dell'architettura SAN***

Le aziende devono poter accedere ai dati in modo rapido e sicuro e quindi la filosofia dell'architettura SAN è quella di poter integrare tutte le caratteristiche dei tradizionali sistemi di memorizzazione:

1. Alte prestazioni
2. Alta disponibilità
3. Scalabilità
4. Facilità di gestione

Tutto questo con le caratteristiche di connettività e accesso distribuito del network computing, attraverso un'architettura di rete dedicata alla gestione e archiviazione dei dati, in grado di non sovraccaricare i server nelle operazioni di scrittura e lettura dei dati, da e verso lo storage.

Le reti SAN forniscono una serie di indubbi vantaggi rispetto ai dispositivi di storage connessi direttamente ai server (*Direct Attached Storage*). Offrono una connettività *any-to-any* tra server e dispositivi di storage, aprendo in tal modo la strada al trasferimento diretto di dati tra periferiche di memorizzazione (dischi o tape), con conseguenti indubbi miglioramenti dell'efficienza dello spostamento dei dati e di processi, quali il backup o la replica dei dati.

***Applicazioni***

L'impiego di Fibre Channel o di qualsiasi altra tecnologia di networking proposta per le reti SAN consente di:

1. raggiungere distanze di connettività superiori e prestazioni migliori rispetto a quanto non sia possibile con le tecnologie SCSI, SAS e SATA
2. facilitare il compito di centralizzare la gestione dello storage che comporta anche l'adozione di strategie di gestione remota e di protezione dei dati
3. consolidamento dello storage e del clustering dei sistemi
4. condividere i dati tra piattaforme diverse

Garantendo alte prestazioni e un accesso diretto ai dischi (Block I/O) le SAN facilitano lo spiegamento di database centralizzati e di applicazioni enterprise quali il Data Warehousing. Inoltre trainano altre iniziative quali il consolidamento dello storage, la protezione dei dati e il disaster recovery. Possono beneficiare di queste prestazioni tutte quelle applicazioni che richiedono un'elevata ampiezza di banda, quali ad esempio:

1. Storage e data consolidation
2. Salvataggio di database
3. Applicazioni distribuite
4. Applicazioni cluster
5. Alta affidabilità
6. Archivi di immagini, foto, grafica, CAD e dati multimediali
7. Controllo e gestione dati
8. Disaster recovery.

## *3.3 Network Attached Storage*

Un *Network Attached Storage* (NAS) è un dispositivo collegato ad una rete di computer la cui funzione è quella di condividere tra gli utenti della rete un'area di storage (o disco).

I NAS sono generalmente dei computer ridotti, con a bordo il minimo necessario per poter comunicare via rete. Oggigiorno i NAS più diffusi sono in pratica dei PC con a bordo il sistema operativo Linux (a volte non visibile dall'utente), e

numerosi hard disk per l'immagazzinamento dati (che possono essere connessi al PC attraverso una Storage Area Network). Questa architettura ha il vantaggio di rendere disponibili i file contemporaneamente su diverse piattaforme come ad esempio Linux, Windows e Unix (o Mac OSX), in quanto il sistema operativo implementa i servizi di rete per tutti gli standard più diffusi quali ad esempio FTP, Network File System (NFS) e Samba per le reti Windows.

Un sistema NAS può essere utilizzato come nodo di una SAN, data la scalabilità di tale architettura.

***Vantaggi e svantaggi dei NAS***

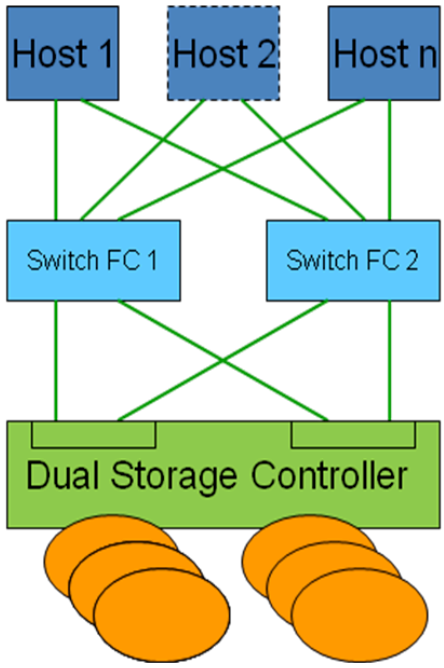
Il vantaggio dei NAS, oltre a centralizzare l'immagazzinamento dei dati in un solo posto invece che spargerli su diversi PC di una rete, è che sono unità altamente specializzate dal punto di vista delle prestazioni e della condivisione dei dati.

Nell'ambito di questa architettura, lo svantaggio maggiore è costituito principalmente dall'enorme quantità di dati che passa sulla rete e dai limiti di stabilità di NFS e degli altri filesystem utilizzabili in rete.

## 3.4 Soluzioni tecniche implementate ai LNF

Il sistema di storage ad uso generale del Servizio di Calcolo è basato su una Storage Area Network i cui elementi di core sono 2 switch Fibre Channel di marca Brocade (modello 4900) ciascuno dei quali con 64 porte da 4 Gb/s.

A tali switch sono connessi gli storage controller e gli host secondo lo schema di principio riportato nella figura a fianco, che prevede una serie di apparati e connessioni ridondate al fine di evitare disservizi per singoli punti di rottura (*no single point of failure*).




Lo schema riportato è molto semplificato in quanto, nella realtà, gli storage controller connessi alla SAN sono attualmente tre e gli host sono varie decine, a volte connessi attraverso altre coppie di switch.

Inoltre agli switch sono connesse molte altre periferiche Fibre Channel, quali unità a nastro e librerie a nastri per l'archiviazione e il backup dei dati.

Come si può notare dallo schema, l'accesso allo storage è garantito da un doppio cammino. Ogni host è connesso attraverso due interfacce Fibre Channel a entrambi gli switch. Gli Storage Controller, a loro volta, utilizzano due o più unità di gestione, con elettroniche ed intelligenze completamente indipendenti, per connettersi

ai due switch Fibre Channel (lato host) e ai dischi. L'affidabilità dell'informazione sui dischi è garantita dai sistemi di organizzazione dei dati di tipo RAID-5 e RAID-6.

Dallo schema si evince che il punto critico della infrastruttura è l'host. Tuttavia l'host va inteso come un fornitore di servizi e di applicazioni. Esistono metodi, descritti in seguito, che garantiscono l'alta disponibilità dei servizi e delle applicazioni in caso di fallimento dell'host.

***Storage Controller e Disk Array ai LNF***

Nel 2002 fu avviata una gara per l'acquisto di uno Storage Controller Ridondato

Il capitolato tecnico prevedeva una serie funzionalità e di requisiti minimi di cui i più importanti erano:

- Alimentatori e sistema di ventilazione ridondati e sostituibili a caldo
- Storage Controller ridondati e sostituibili a caldo
  - Storage Processor ridondati
  - Firmware e Software di sistema ridondati
  - Sottosistemi di Storage ridondati
  - Gestione dei RAID (livelli 0,1,5) in hardware
- Possibilità di gestione di dischi Fibre Channel e ATA

Lo stesso capitolato chiedeva di specificare molte altre caratteristiche tecniche che avrebbero costituito elementi per la valutazione tecnico/economica dell'offerta, quali ad esempio:

- Architettura delle motherboard degli storage controller
  - CPU utilizzata
  - Memoria RAM
  - Cache Size
  - Localizzazione e tipologia del s/w di sistema
  - Modalità di upgrade del s/w di sistema (on-line/off-line e tempi)
  - Possibilità di notifica dei fault via e-mail
- le caratteristiche dei dischi:
  - casa costruttrice
  - numero di giri al minuto (RPM)
  - capacità nominale
  - Sustained Transfer Rate, average seek time, etc.
- Il numero massimo di dischi installabile nel singolo bay
- La capacità di mass storage del singolo disco (FC e ATA)
- Il numero di bay per i dischi, installabili sul sistema
- La capacità di mass storage complessiva del sistema
- Etc.

A seguito di questa gara fu acquistato un sistema di Marca $EMC^2$ modello Clariion CX500 che nel giro di 2 anni fu affiancato da un altro sistema $EMC^2$, modello Clariion CX300. Il primo è in grado di esportare fino a 120 dischi, il secondo fino a 60; segue la tabella della distribuzione:

| Controller | Dischi Fibre Channel | Dischi ATA | Spazio Totale |
|---|---|---|---|
| Clariion CX500 | 83 x 146GB 10Krpm<br>37 x 300GB 10Krpm | | 23 TeraByte |
| Clariion CX300 | 15 x 300GB 10Krpm | 37 x 500GB 7.2Krpm<br>8 x 750GB 7.2Krpm | 29 TeraByte |

**Tabella 3-1: Spazio disco servito dai controller EMC[2]**

Come si evince dalla tabella, nel 2008 si è raggiunto il limite di espandibilità dei due sistemi EMC[2], e quindi si è acquistato un nuovo sistema di Storage, di fascia "*enterprise*": marca Hitachi-Storagetek modello 9985V.

Le caratteristiche più significative includono un'architettura hardware basata su *crossbar switch*, disegnata per fornire un alto livello di prestazioni e affidabilità, *no single point of failure*, accesso ai dati istantaneo, componenti ridondati e *fault-tolerant*, *data cache mirroring*.

Più in particolare il sistema è in grado di fornire:
- 13.3 GB/s banda aggregata e 800,000 IOPS
- Fino a 48 canali Fibre Channel a 4Gb/s
- Fino a 128 GB di RAM cache
- Supporto per modalità RAID 0, 1, 1+0, 5, e 6 in hardware (ASIC)
- Fino a 240 dischi FC (600GB) o SATA (2TB)
- Virtualizzazione dello Storage interno ed esterno
- Partizionamento e Virtualizzazione degli storage controller

Attualmente è configurato con 19 dischi SATA da 1TB 7.2Krpm, e con 18 dischi FC da 450GB 15Krpm, organizzati a blocchi di 8 in volumi RAID-6 (6 dischi utili più 2 di parità) e alcuni dischi utilizzati come *hot spare*.

Quindi lo spazio disco complessivo (lordo) servito a scopo generico dal Servizio di Calcolo utilizzando la Storage Area Network (rete Fibre Channel) è attualmente di circa 80 TeraByte. Esistono in realtà altre aree di Storage gestite dal Servizio, ma ad uso di esperimenti specifici, nell'ambito di particolari farm di calcolo, di cui si parlerà nel seguito. Esistono inoltre, prevalentemente per motivi storici o di economia, alcuni servizi specifici che alloggiano dischi al loro interno, piuttosto che importarli attraverso la SAN.

La Figura 3-3 illustra lo schema di connessione, nell'ambito della Storage Area Network dei LNF, tra gli host e i sistemi di storage tramite i due switch Fibre Channel principali.

I due rettangoli azzurri denominati DS4900B (sopra e sotto) rappresentano gli switch principali Fibre Channel Brocade a 64 porte, di cui solo le prime 32 abilitate.

Gli altri 4 rettangoli azzurri più piccoli, sulla destra della figura, rappresentano 2 coppie di switch Fibre Channel Brocade interni a 2 chassis per blade system, uno di marca HP e uno di marca IBM, che ospitano al loro interno rispettivamente 8 e 14 *blade system*, tutti connessi all'infrastruttura.

I rettangoli rosa rappresentano gli *storage controller* EMC[2] e Hitachi-StorageTek comprensivi dei dischi, e un Network Attached Storage.

Gli altri rettangoli rappresentano *host* e altri *storage device* quali *tape drive* e librerie a nastro.

Connessioni della Storage Area Network dei LNF

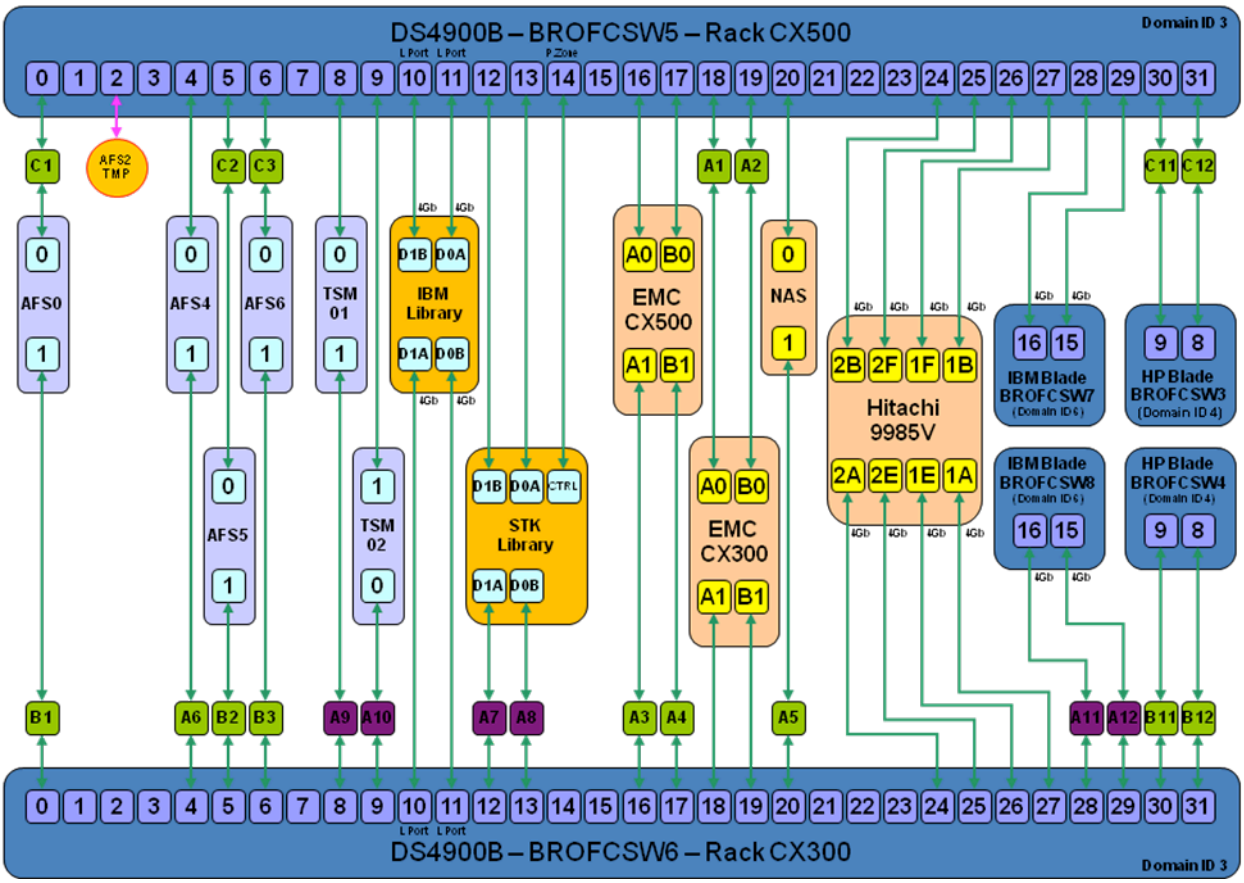



**Figura 3-3: Schema di connessione della SAN dei LNF**

## Legenda:

| | |
|---|---|
| **BROFCSW5/6** | Switch Brocade DS4900B (FOS 5.3.0c) – Domain ID 3) |
| **CX300** | Sistema Storage EMC CX300 (Fw 2.24.300.5.019) |
| **CX500** | Sistema Storage EMC CX500 (Fw 2.19.500.5.040) |
| **NAS** | Sistema NAS EMC Celerra NS600G (5.5.25-2) |
| **Hitachi 9985V** | Sistema Storage Hitachi 9985V |
| **IBM Library** | Libreria a Nastri IBM |
| **STK Library** | Libreria a Nastri SUN/StorageTek |
| **AFS0** | Server Linux |
| **AFS2** | Server IBM (solo test) |
| **AFS4** | Server Linux (al momento non attivo) |
| **AFS5** | Server Linux |
| **AFS6** | Server Linux |
| **TSM01** | Server IBM (Accede solo alle librerie a nastri) |
| **TSM02** | Server IBM (Accede solo alle librerie a nastri) |
| **BROFCSW3/4** | Switch Brocade Blade HP (FOS 5.0.5) – Domain ID 4) |
| **BROFCSW7/8** | Switch Brocade Blade IBM (FOS 5.3.0a) – Domain ID 6) |

***Librerie a nastro***

Il sistema di backup dei LNF si serve di 2 librerie a nastro robotizzate, una IBM ed una Storagetek.

La prima, una IBM 3494 acquistata nel 1999 e aggiornata nel 2006, è composta da tre armadi che ospitano una robotica lineare, e quattro drive di lettura/scrittura della famiglia Magstar modello 3592.

La libreria, espandibile linearmente aumentando il numero degli armadi, è attualmente in grado di alloggiare circa 1000 cassette (cartridge), ognuna delle quali è marcata con un codice a barre univoco. Il robot o braccio meccanico, grazie al codice a barre, identifica le cartridge nelle rispettive posizioni e costruisce un catalogo.

I 4 drive sono in grado di leggere e scrivere sulle cartridge ad una velocità di oltre 40 MB/s e volendo tale velocità può apparentemente aumentare utilizzando metodi di compressione in hardware. Ogni cassetta è in grado di contenere fino a 300 GB di dati (senza tecniche di compressione). La capacità nominale complessiva della libreria è di 300 TB.

La libreria Storagetek è il modello L1400 a robotica circolare (non espandibile), con 2 drive Enterprise modello STK 9940. La libreria è attualmente in grado di ospitare 700 cassette (espandibile fino a 1400). La velocità di lettura/scrittura sulle cassette è di circa 30MB/s (senza compressione), e le cassette sono in grado di contenere fino a 200GB di dati. La capacità nominale complessiva della libreria è di 140 TB (potenzialmente 280 TB con 1400 cassette).

# 4. I servizi informatici centrali

Le infrastrutture informatiche viste nei capitoli precedenti, sono difficilmente fruibili senza l'ausilio fondamentale di una serie di servizi informatici di supporto, alcuni dei quali sono essenziali per il funzionamento, altri meno critici, ma comunque molto utili, spesso irrinunciabili.

Per motivi di affidabilità, quasi tutti i servizi vengono svolti da sistemi dedicati e, dato che oggi la potenza di calcolo, la velocità di Input/Output, la quantità di memoria e tutte le risorse in genere a disposizione di una singola macchina fisica sono esagerate per gestire la maggior parte di tali servizi, conviene ricorrere a metodi di virtualizzazione dei sistemi operativi.

La virtualizzazione permette di eseguire più sistemi operativi (detti virtuali) sulla stessa macchina fisica. I sistemi virtuali condivideranno le risorse h/w della macchina fisica, ma sono indipendenti tra di loro.

Considerando ad esempio una macchina fisica (con architettura x86) di ultima generazione, questa è tipicamente dotata di almeno 2 processori *quad-core* (per un totale di 8 *core* complessivi) e fino a 64GB di memoria RAM. La maggior parte dei servizi che verranno descritti nel seguito non hanno bisogno di tante risorse, ma consumano poca memoria e una piccola percentuale di un singolo core. È evidente quindi che un sistema di virtualizzazione permette di eseguire decine di sistemi virtuali su una singola macchina fisica, salvaguardando l'investimento economico fatto per l'acquisto dell'hardware.

Inoltre, se le risorse di storage sono condivise tra più macchine fisiche, è possibile far partire una macchina virtuale su un'altra macchina fisica con estrema semplicità e velocità, rendendo agevole il ripristino di quei servizi virtuali falliti in seguito al guasto della macchina fisica che li ospita.

Volendo infine realizzare un sistema ad alta affidabilità e disponibilità, è possibile creare un cluster di sistemi fisici, in grado di condividere lo stesso file system, quindi installare le macchine virtuali come servizi failover del cluster, in modo che al fallimento di un nodo fisico del cluster, tutte le macchine virtuali gestite da tale nodo ripartano automaticamente sugli altri nodi dello stesso cluster.

## *4.1 Soluzioni tecniche implementate ai LNF*

Ai LNF la virtualizzazione è realizzata con il s/w opensource Xen incluso nella distribuzione Scientific Linux.

La Scientific Linux è una distribuzione di Linux assemblata dal Fermilab (in USA), con lo scopo principale di realizzare una distribuzione standard del sistema operativo Linux per tutti gli istituti di ricerca esistenti nel mondo, al fine di ottenere una base comune di installazione per i vari esperimenti.

La Scientific Linux è di fatto una distribuzione di tipo Enterprise derivata dalla Red Hat Advanced Server, ricompilata a partire dai sorgenti opensource, e già include il s/w di cluster e quello di virtualizzazione (Xen).

Ai LNF esistono due cluster indipendenti installati su sistemi hardware di tipo *blade*. I sistemi blade sono sistemi ad alta integrazione alloggiati in un apposito chassis che permette di condividere linee di alimentazione e di comunicazione.

Nello stesso chassis trovano spazio una coppia di switch Fibre Channel e una coppia di switch Ethernet. Ogni singola macchina (lama o blade system) ha doppia connettività di rete e doppia connettività allo storage FC. Il singolo chassis può contenere molte lame. Ad esempio, ai LNF, è presente un sistema HP p-class chassis (8 lame totali), un sistema IBM H chassis (14 lame totali), e un sistema Dell Chassis M1000e (16 lame totali).

I sistemi blade o hanno poco disco o sono addirittura diskless. Tuttavia importano tutto lo storage di cui necessitano grazie alla connettività Fibre Channel che permette l'integrazione con l'infrastruttura descritta nel capitolo precedente.

I due cluster sono realizzati con la release 5.4 di Scientific Linux e utilizzano un device logico, realizzato con il s/w Clustered LVM nativo della distribuzione Linux, per condividere lo spazio di storage.

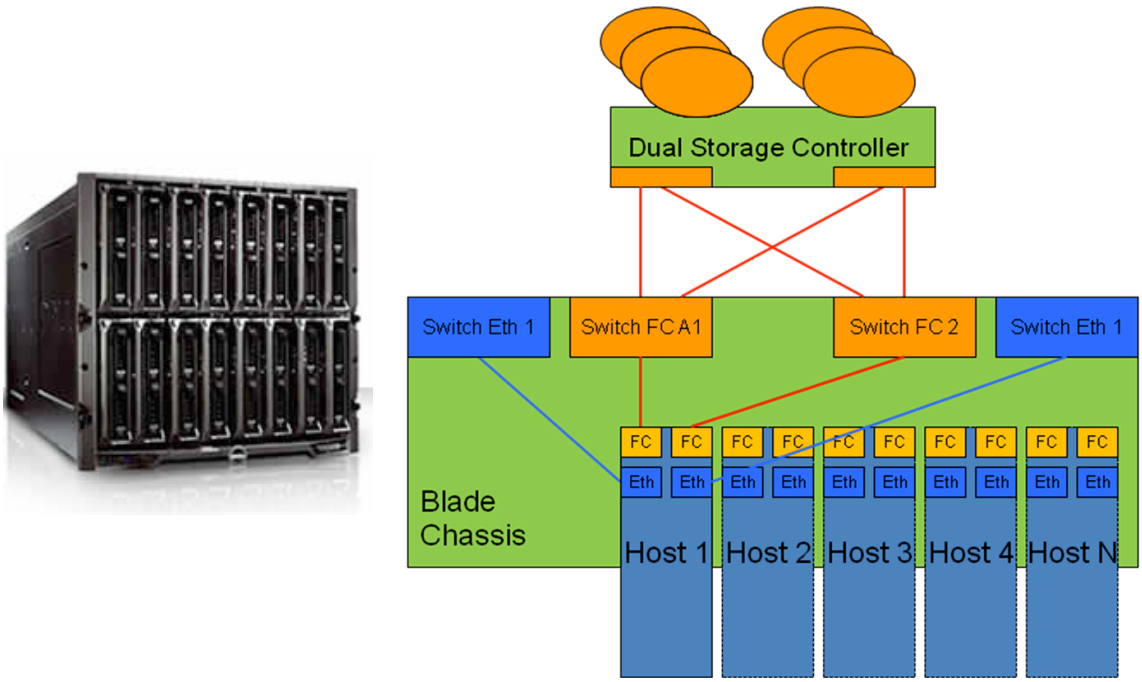



**Figura 4-1: Blade system (foto e schema)**

## 4.2 Servizi di base

Per poter usufruire dei servizi informatici di alto livello, è necessario il supporto di alcuni servizi di base, spesso definiti servizi essenziali.

### 4.2.1 DHCP

Il Dynamic Host Configuration Protocol (DHCP) (protocollo di configurazione dinamica degli indirizzi) è un protocollo che permette ai dispositivi di rete di ricevere la configurazione necessaria per poter operare su una rete basata su Internet Protocol.

In una rete basata sul protocollo IP, ogni calcolatore ha bisogno di un indirizzo IP, scelto in modo tale che appartenga alla sottorete a cui è collegato e che sia unico, ovvero che non ci siano altri calcolatori che stiano già usando quell'indirizzo.

Il compito di assegnare manualmente gli indirizzi IP ai calcolatori comporta un rilevante onere per gli amministratori di rete, soprattutto in reti di grandi dimensioni o in caso di numerosi computer che si connettono a rotazione solo a ore o giorni determinati.

Il protocollo DHCP viene usato anche per assegnare al computer diversi parametri necessari per il suo corretto funzionamento sulla rete a cui è collegato. Tra i più comuni, oltre all'assegnazione dinamica dell'indirizzo IP, si possono citare:

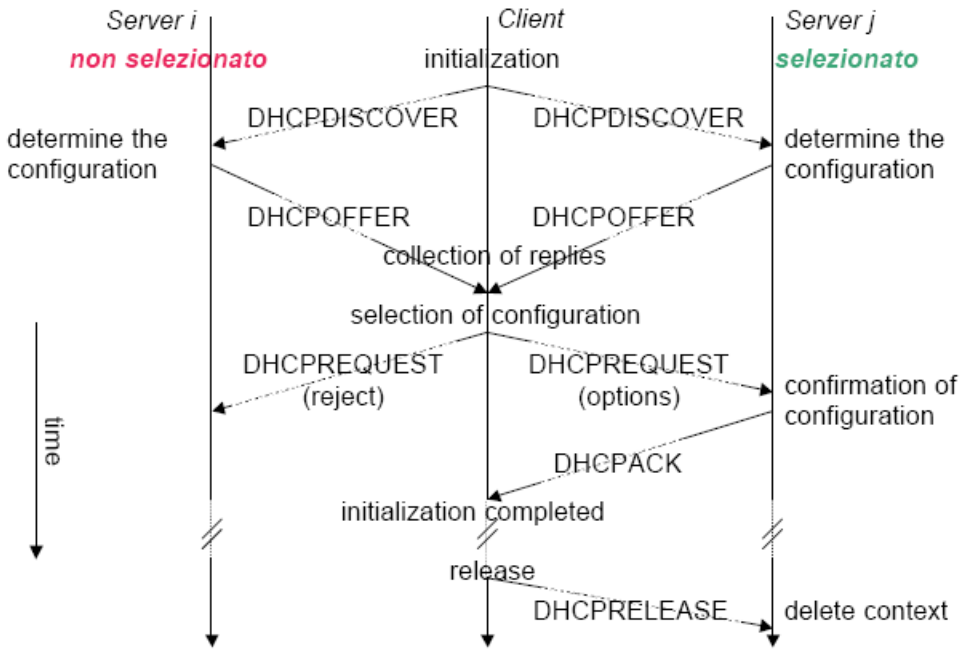
- Maschera di sottorete (*netmask*)
- Default Gateway
- Indirizzi dei server DNS
- Nome di dominio DNS di default
- Indirizzi dei server WINS
- Altri parametri

Il protocollo DHCP prevede che il client invii un pacchetto chiamato DHCPDISCOVER in broadcast, con indirizzo IP sorgente messo convenzionalmente a 0.0.0.0, e destinazione 255.255.255.255. Il pacchetto viene ricevuto da tutto il dominio di broadcast, e in particolare da tutti i server DHCP presenti, che possono rispondere (o meno) con un pacchetto di DHCPOFFER, in cui propongono un indirizzo IP ed altri parametri al client. Questo pacchetto è indirizzato direttamente all'indirizzo di livello datalink del client (che non ha ancora un indirizzo IP), per cui può essere inviato solo da un server che si trovi sullo stesso dominio di broadcast.

Se nel dominio di broadcast ci sono anche uno o più DHCP relay, questi inoltrano il pacchetto al loro server di riferimento, che può rispondere al client sempre attraverso il relay. Il relay agent comunica al server il proprio indirizzo IP sulla sottorete da cui ha ricevuto il pacchetto di DHCPDISCOVER, permettendo al server di capire da quale sottorete è arrivata la richiesta, e quindi offrire un indirizzo per la sottorete giusta. Un server DHCP che debba servire diverse sottoreti IP deve essere configurato per conoscere i parametri di ciascuna (indirizzo della rete, maschera di sottorete, indirizzo di broadcast, indirizzo del gateway).

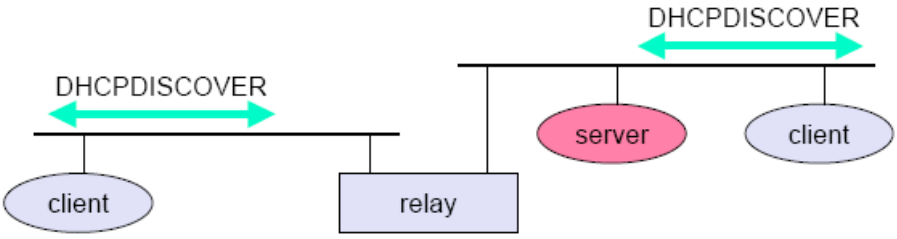
Il client aspetta, per un certo tempo, la ricezione di una o più offerte, dopodiché ne seleziona una, ed invia un pacchetto di DHCPREQUEST in broadcast, indicando all'interno del pacchetto, con il campo "server identifier", quale server ha selezionato. Anche questo pacchetto raggiunge tutti i server dhcp presenti sulla rete (direttamente o tramite un relay).

Il server che è stato selezionato conferma l'assegnazione dell'indirizzo con un pacchetto di DHCPACK (nuovamente indirizzato direttamente all'indirizzo di livello datalink del client, possibilmente attraverso un relay); gli altri server vengono informati che la loro offerta non è stata scelta dal client, e che sulla sottorete è presente un altro server DHCP.




Come si è visto nel paragrafo 2.6.2, ai Laboratori Nazionali di Frascati si è scelto di assegnare gli indirizzi IP pubblici correlandoli all'indirizzo MAC del client. Ovvero ogni nodo connesso in rete si presenterà sempre con lo stesso indirizzo fisico di secondo livello quindi, in funzione di questo, verrà mappato sempre sulla stessa VLAN, gli verrà assegnato sempre lo stesso indirizzo IP e gli verranno attribuiti sempre gli stessi parametri di rete dal DHCP.

Tuttavia occorre tener presente che il pacchetto di richiesta del client è broadcast e rimarrebbe confinato nella VLAN di appartenenza. Affinché tale messaggio arrivi ai DHCP server, gli switch di livello 3, che svolgono la funzionalità di routing tra le varie VLAN interne, sono stati configurati per svolgere anche la funzione di DHCP relay.



Per motivi di affidabilità ai LNF sono stati configurati 2 server DHCP in modalità ridondata, ovvero con lo stesso file di configurazione. Attualmente uno dei due server è una macchina fisica, l'altro è una macchina virtuale. Il file di configurazione mantiene aggiornata la mappatura statica tra gli indirizzi MAC e gli indirizzi IP dei client. Esso risiede su un repository comune (AFS) e, dopo ogni modifica, con un semplice script viene copiato su entrambe le macchine e vengono fatti ripartire entrambi i servizi (restart dei processi *daemon*).

### 4.2.2 DNS

Come è noto, in rete i computer vengono indicati per comodità con dei nomi facili da ricordare, ma ciò che veramente conta e consente ai computer di parlarsi è l'indirizzo IP numerico. Difficilmente gli utenti della rete si rassegnerebbero ad usare i numeri, ma ancora più difficilmente le funzioni di routing della rete potrebbero funzionare sui nomi. Occorre pertanto un sistema che consenta all'utente di indicare per nome il computer desiderato, in grado di convertire tale nome nel corrispondente indirizzo IP, necessario affinché il computer stesso possa essere raggiunto tramite la rete. Il sistema che si occupa di questa importantissima funzione si chiama Domain Name System e viene indicato come DNS.

Possiamo immaginare il DNS come una immensa tabella di conversione. Essendo impensabile che una siffatta tabella risieda in un luogo solo e che decine di migliaia di amministratori di sistema la accedano in continuazione per aggiornarla, la soluzione adottata è quella del database distribuito. Vale a dire che, grazie ad una organizzazione gerarchica, l'intera rete è divisa in zone, ciascuna delle quali è servita da un server DNS che possiede i dati solo per le risorse appartenenti alla zona stessa. L'amministratore di sistema responsabile per le risorse di quella zona avrà quindi da aggiornare solamente il proprio server DNS, senza bisogno di dire a nessun altro che tal nome ha cambiato indirizzo o cose del genere.

I domini assegnati e gestiti dal Servizio di Calcolo dei LNF sono:

*lnf.infn.it*
*ac.infn.it*

In questo caso il dominio di primo livello è .it che identifica la nazione Italia, poi viene .infn che identifica l'Ente, poi .lnf e .ac che identificano rispettivamente i Laboratori Nazionali di Frascati e l'Amministrazione Centrale dell'INFN.

Quindi un host di nome lxcalc nella rete dei LNF avrà un nome su Internet che sarà *lxcalc.lnf.infn.it.*

Ogni volta che una macchina collegata in rete deve accedere a un host di Internet, ad esempio per aprire una pagina web, viene interrogato il DNS server della rete locale che demanderà eventualmente la richiesta ad altri server DNS esterni (seguendo la gerarchia dei domini).

È quindi evidente quanto sia importante il buon funzionamento dei server DNS. Nel caso che il server DNS non funzioni, anche con la rete perfettamente funzionante,

non sarebbe praticamente utilizzabile nessun servizio (funzionerebbero infatti solo quei servizi che accedono ad un host tramite numero IP). Per questo motivo il server DNS dei LNF è realizzato in maniera ridondante. Inoltre dato che il servizio di DNS non impegna eccessivamente la macchina, né come CPU, né come rete, è stato scelto di installarlo sulla stesse macchine ove risiede il server DHCP.

In questo caso la macchina fisica funge da server DNS primario e la macchina virtuale da server DNS secondario. Ogni aggiornamento del file di configurazione avviene sul server primario, in quanto il protocollo del servizio DNS prevede implicitamente l'allineamento dei server secondari.

Esiste anche un terzo server secondario, esterno alla LAN dei LNF, sito presso la sede INFN di Roma1 (Università "La Sapienza").

### 4.2.3 Server di autenticazione Kerberos

In un ambiente distribuito ed eterogeneo come quello dei L.N.F. è necessario individuare un approccio alla sicurezza tale da richiedere all'utente di dimostrare la propria identità per ogni servizio utilizzato, garantendo inoltre l'identità dei server verso i client. E proprio questo il servizio offerto da Kerberos.

Kerberos presuppone la presenza di un'architettura distribuita client/server e impiega uno o più server Kerberos per fornire un servizio di autenticazione.

Il protocollo di autenticazione Kerberos nasce come un sistema di autenticazione sicura su reti non sicure, ed è stato ideato per soddisfare i seguenti principali requisiti:

- ***Sicurezza:*** un'operazione di intercettazione di traffico di rete non deve essere in grado di ottenere le informazioni necessarie per sostituire un altro utente. Più in generale Kerberos deve far in modo che un potenziale cracker non trovi un anello debole.

- ***Affidabilità:*** per tutti i servizi che fanno affidamento su Kerberos per il controllo degli accessi, la chiusura del servizio Kerberos significa mancanza di disponibilità dei servizi supportati. Pertanto Kerberos deve essere altamente affidabile e deve impiegare un'architettura con server distribuiti, in maniera tale da non rendere l'intero sistema legato ad una singola macchina.

- ***Trasparenza:*** l'utente non dovrebbe accorgersi dell'attuazione dell'autenticazione, tranne per la richiesta iniziale di una password.

- ***Scalabilità:*** il sistema deve essere in grado di supportare un gran numero di client e server. Questo dunque suggerisce l'impiego di un'architettura distribuita e modulare.

Per supportare questi requisiti, il meccanismo di Kerberos prevede che venga utilizzato un servizio di autenticazione esterno che impieghi un protocollo fidato, nel senso che i client e i server confidano in Kerberos che effettua la reciproca

autenticazione. Supponendo che il protocollo Kerberos sia ben realizzato e che il server Kerberos stesso sia sicuro, allora il servizio di autenticazione può essere considerato "sicuro". È bene ricordare che la sicurezza del server Kerberos non deve mai essere data per scontata, ma deve essere verificata con la massima cura.

#### 4.2.3.1 Realm Kerberos

Un ambiente Kerberos è costituito da un server Kerberos, una serie di client ed una serie di server di applicazioni con alcuni requisiti, ad esempio:

1. Il server Kerberos deve conservare in un database il codice utente (*principal*) e la password *hash* di tutti gli utenti partecipanti. Tutti gli utenti sono registrati sul server Kerberos.

2. Il server Kerberos deve condividere una chiave segreta con ciascun server. Tutti i server sono registrati sul server Kerberos. Un realm è proprio la descrizione di questo ambiente. Le reti di client e di server operanti in organizzazioni amministrativamente differenti costituiscono realm differenti.

#### 4.2.3.2 Protocollo di autenticazione

Quella che segue è una descrizione del protocollo. In breve, il client si autentica presso l'*Authentication Server* (AS) che gli fornisce un ticket di sessione per accedere al *Ticket Granting Server* (TGS), poi si autentica presso il *Ticket Granting Server* e riceve il *ticket* per aprire una sessione di comunicazione con il *Service Server* (SS). In dettaglio (rif. Figura 4-2):

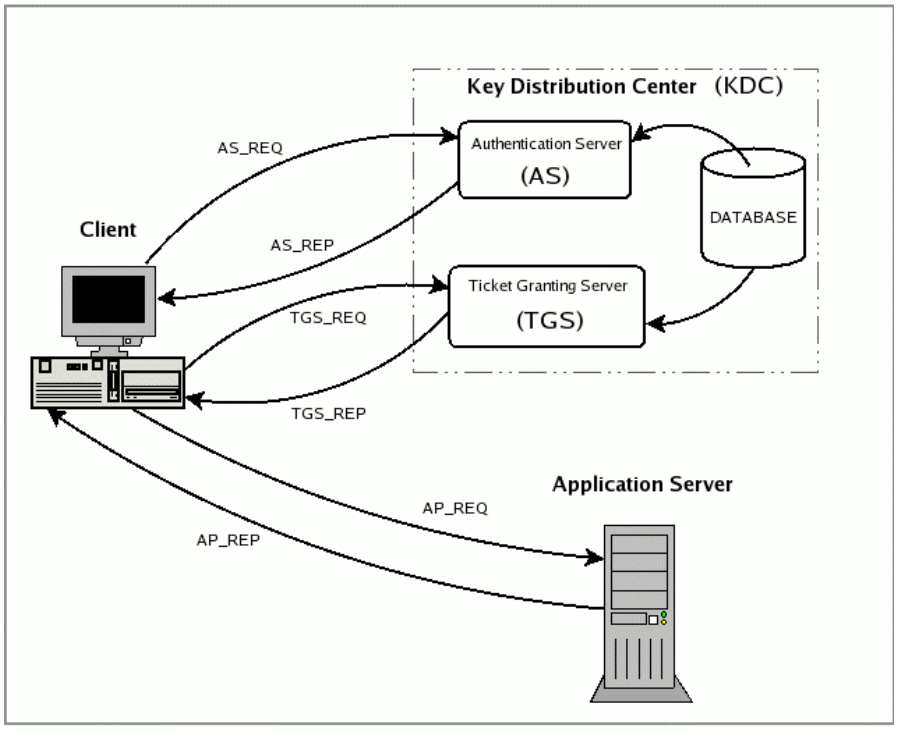



**Figura 4-2: Protocollo kerberos**

***Utente: Autenticazione di base***

1. Un utente inserisce *username* e *password* sul client.

***Client: Autenticazione AS***

1. Il client manda un messaggio non crittografato all'Authentication Server richiedendo i servizi per l'utente. Né la chiave segreta né la password vengono inviate all'AS.
2. L'AS controlla se il client è nel suo database. Se c'è, invia due messaggi al client:
   - Messaggio A: Chiave di sessione client-TGS crittografata usando la chiave segreta dell'utente.
   - Messaggio B: Ticket-Granting Ticket (che include l'identificativo del client, l'indirizzo di rete, il tempo di validità del ticket e la chiave di sessione client-TGS) crittografata utilizzando la chiave segreta di TGS.
3. Quando il client riceve i messaggi A e B, decrittografa il messaggio A ottenendo la chiave di sessione client-TGS. Questa chiave è utilizzata per le successive comunicazioni con TGS (si noti che il client non può decrittografare il Messaggio B, che è stato crittografato con la chiave segreta di TGS). A questo punto il client possiede i mezzi per autenticarsi presso il TGS.

***Client: Autenticazione TGS***

1. Quando richiede dei servizi, il client invia i seguenti due messaggi a TGS:
   - Messaggio C: composto dal Ticket-Granting Ticket (mandatogli dall'AS nel messaggio B) e dall'identificativo del servizio richiesto.
   - Messaggio D: autenticatore (*Authenticator*) (che è formato da identificativo del client e *timestamp*), crittografato usando la chiave di sessione client—TGS.
2. Ricevendo i messaggi C e D, TGS decrittografa il messaggio C con la propria chiave e dal messaggio estrae la chiave di sessione client—TGS che utilizza per decrittografare il messaggio D (autenticatore). A questo punto invia i seguenti due messaggi al client:
   - Messaggio E: Ticket client-server (che include l'identificativo del client, l'indirizzo di rete del client, il periodo di validità e la chiave di sessione client-server) crittografato utilizzando la chiave segreta del server che offre il servizio.
   - Messaggio F: Chiave di sessione client-server crittografato usando la chiave di sessione client-TGS.

***Client: Autenticazione SS***

1. Ricevendo i messaggi E e F dal TGS, il client può autenticarsi presso il SS. Il client si connette al SS e invia i seguenti due messaggi:
   - Messaggio G: Ticket client-server crittografato usando la chiave segreta di SS.
   - Messaggio H: un nuovo autenticatore, che include l'identificativo del client, il timestamp è crittografato usando la chiave di sessione client-server.

2. Il server decrittografa il ticket usando la sua chiave segreta e invia il seguente messaggio al client per confermare la propria identità e la volontà di fornire il servizio al client:
    - Messaggio I: il timestamp trovato nell'autenticatore incrementato di uno, crittografato utilizzando la chiave di sessione client-server.
3. Il client decrittografa la conferma usando la chiave di sessione client-server e controlla che il timestamp sia correttamente aggiornato. Se lo è, il client può considerare affidabile il server e iniziare a effettuare le richieste di servizio.
4. Il server fornisce i servizi al client.

#### 4.2.3.3 Kerberos ai LNF

Il servizio di autenticazione è estremamente critico per il funzionamento dei servizi informatici, per cui è opportuno che sia sicuro ed affidabile. A tale scopo, ai LNF il servizio di autenticazione Kerberos relativamente al realm *LNF.INFN.IT* è implementato su quattro server: un KDC (*Key Distribution Center*) principale e tre slave. Tutti e quattro sono in grado di effettuare l'autenticazione, il ché implica che tutti devono contenere il database degli utenti con le chiavi di autenticazione. L'aggiornamento del Database (ovvero la creazione di nuove credenziali di autenticazione, piuttosto che il cambio password, etc.) avviene sempre sul server principale, mentre la sincronizzazione dei database tra il server principale e i server slave è prevista dal protocollo Kerberos.

Le *username* definite nel realm dei LNF sono circa 1500. Il mapping con gli utenti è uno a uno, ovvero ogni utente ha una sola username. Non esistono username generiche di sigle o esperimenti. Hanno diritto ad avere una username tutti i dipendenti, associati e ospiti dell'INFN che lavorano o collaborano con esperimenti o progetti dei LNF.

La richiesta deve essere fatta tramite un'opportuna procedura web, in cui il nuovo utente deve specificare il progetto di appartenenza e il responsabile di account del progetto stesso. Quest'ultimo deve approvare la richiesta. Il Servizio di Calcolo si occupa di verificare che la posizione del richiedente sia regolarizzata amministrativamente e quindi crea la username (con una password iniziale generata random) e comunica le credenziali al richiedente. L'utente dovrà cambiare la password al primo uso, in accordo con alcune regole di base:

- deve avere lunghezza minima 8 caratteri,
- deve contenere caratteri di almeno 2 tra le seguenti classi di caratteri:
    - lettere alfabetiche maiuscole
    - lettere alfabetiche minuscole
    - caratteri numerici (cifre da 0 a 9)

L'autenticazione è richiesta per tutti i servizi informatici non pubblici dei LNF, quali ad esempio: il login interattivo, l'accesso al servizio di posta elettronica, l'accesso ai file system di rete, l'accesso ad applicazioni web, etc..

Per accedere a tutte le risorse in cui è richiesta l'autenticazione, l'utente deve inserire a richiesta sul proprio client la username e la password per ottenere un ticket

che può avere validità di 10 giorni. Con tale ticket l'utente non ha più necessità di reinserire le credenziali per accedere ai servizi sopra elencati, ad accesso kerberizzato.

### 4.2.4 Server di autenticazione RADIUS

RADIUS (Remote Authentication Dial-In User Service) è un protocollo AAA (*authentication, authorization, accounting*) utilizzato in applicazioni di accesso alle reti o di mobilità IP. RADIUS è attualmente lo standard de-facto per l'autenticazione remota, prevalendo sia nei sistemi nuovi che in quelli già esistenti. Tutti gli apparati di rete integrano il protocollo di autenticazione RADIUS.

Il processo di autenticazione ha inizio quando un client crea un pacchetto RADIUS Access-Request, includendo almeno gli attributi User-Name e User-Password, e generando il contenuto del campo identificatore. Il processo di generazione del campo identificatore non è specificato nel protocollo RADIUS, ma è solitamente implementato come un semplice contatore incrementato ad ogni richiesta.

Il campo authenticator contiene una Request-Authenticator, ovvero una stringa di 16 byte scelta in modo casuale. L'intero pacchetto è trasmesso in chiaro, a parte per l'attributo User-Password, che è protetto nel modo seguente: il client e il server condividono una chiave segreta. Tale chiave viene unita con la Request Authenticator, e l'intera stringa viene sottoposta a una funzione hash MD5 per la creazione di un valore di 16 ottetti, sottoposto a sua volta a un XOR con la password immessa dall'utente (e se tale password è più lunga di 16 ottetti, vi è un calcolo MD5 addizionale, utilizzando il testo cifrato anziché la Request Authenticator).

Il server riceve il pacchetto Access-Request e verifica di possedere la chiave segreta per il client. In caso negativo, il pacchetto viene silenziosamente ignorato. Poiché anche il server è in possesso del segreto condiviso, è possibile utilizzare una versione modificata del processo di protezione del client per ottenere la password in chiaro. Quindi il server consulta il database per convalidare username e password; se la password è valida, il server crea un pacchetto Access-Accept da rimandare al client. In caso contrario, crea un pacchetto Access-Reject e lo invia al client.

Entrambi i pacchetti Access-Accept e Access-Reject utilizzano lo stesso valore identificatore del pacchetto Access-Request del client, e hanno una Response Authenticator nel campo Authenticator. La Response Authenticator è la funzione hash MD5 del pacchetto di risposta con l'associata Request Authenticator, concatenata con il segreto condiviso.

Quando il client riceve un pacchetto di risposta, si accerta che esso combaci con una precedente richiesta utilizzando il campo identificatore. Se non esiste alcuna richiesta con lo stesso identificatore, la risposta è silenziosamente ignorata. Quindi il client verifica la Response Authenticator utilizzando lo stesso calcolo effettuato dal server, ed infine comparando il risultato con il campo Authenticator. Se la Response Authenticator non coincide, il pacchetto è silenziosamente ignorato.

Se il client riceve un pacchetto Access-Accept verificato, username e password sono considerati corretti, e l'utente è autenticato. Se invece riceve un pacchetto Access-

Reject verificato, username e password sono scorretti, e di conseguenza l'utente non è autenticato.

RADIUS è un protocollo che utilizza pacchetti UDP per trasportare informazioni di autenticazione e configurazione tra l'autenticatore e il server RADIUS. L'autenticazione è basata su username, password e, opzionalmente, risposta a una richiesta di riconoscimento (una sorta di "parola d'ordine"). Se l'autenticazione ha successo, il server RADIUS invia le informazioni di configurazione al client, inclusi i valori necessari a soddisfare il servizio richiesto, come un indirizzo IP e una maschera di sottorete per PPP o un numero di porta TCP per telnet.

Uno dei limiti del protocollo RADIUS è l'autenticazione basata esclusivamente su password: la password è trasmessa o in forma hash (utilizzando l'algoritmo di hashing MD5), oppure sottoforma di risposta a una richiesta di identificazione (CHAP-password). L'Extensible Authentication Protocol (EAP) rende RADIUS capace di lavorare con una varietà di schemi di autenticazione, inclusi chiave pubblica, Kerberos e smart card.

L'access point agisce da traduttore EAP-RADIUS tra il client wireless e il RADIUS server. Esso utilizza il protocollo EAP per comunicare con il client e il protocollo RADIUS per comunicare con il server RADIUS. L'access point incapsula le informazioni (come lo username o la chiave pubblica) in un pacchetto RADIUS che inoltra al server RADIUS. Quando il server rimanda una delle possibili risposte (Access-Accept/Reject/Challenge), l'access point spacchetta il pacchetto RADIUS ed inoltra la risposta al client in un pacchetto EAP.

La RFC 2869 (RADIUS Extensions) specifica gli attributi opzionali da settare sui pacchetti RADIUS per indicare al server RADIUS che si sta utilizzando il protocollo EAP. Poiché il pacchetto EAP include un campo per specificare quale metodo di autenticazione è in uso, il server RADIUS implementa l'autenticazione richiamando un'apposita procedura.

#### 4.2.4.1 RADIUS ai LNF

Ai LNF, come già visto nei capitoli precedenti, il protocollo RADIUS è implementato con l'utilizzo del protocollo EAP ed in particolare con il TTLS (*Tunnelled Transport Layer Security*) che permette l'inoltro di informazioni tramite un canale cifrato con un protocollo di cifratura a chiave simmetrica. Dopo che il client ha inviato le informazioni di autenticazioni ed ottenuto risposta, il tunnel collassa.

Ai LNF ove esiste già un sistema di autenticazione centralizzato basato su kerberos, per non duplicare i database di autenticazione, si è impostato RADIUS in modo tale da demandare l'autenticazione a Kerberos. In tal modo RADIUS si comporta come Gateway di autenticazione.

***Applicazione per la gestione dei visitatori***

L'unica eccezione è rappresentata dai visitatori occasionali che vogliono collegarsi alla rete dei LNF attraverso il captive portal. In tal caso le autenticazioni non

esistono su kerberos, quindi RADIUS li tratta in modo differenziato ed esplica autonomamente la funzione di autenticazione, verificando in un database locale.

Il database locale viene popolato attraverso un'apposita applicazione per la gestione dei visitatori (sviluppata in casa in linguaggio Java), che popola un database Oracle nazionale contenente le anagrafiche degli ospiti occasionali, ivi incluse le digitalizzazioni dei documenti di identità personali, in aderenza alla legge antiterrorismo ("Pacchetto Pisanu", N. Decreto Legge Consiglio dei Ministri 22 luglio 2005).

### 4.2.5 File System Distribuito (*Andrew File System*)

Il file system Andrew, o AFS è un file system distribuito sviluppato dalla Carnegie Mellon University, all'interno del progetto Andrew. Il nome è stato dato in onore di Andrew Carnegie e Andrew Mellon. L'utilizzo principale di questo file system è nell'elaborazione distribuita.

AFS ha diversi vantaggi rispetto ai tradizionali file system distribuiti, in particolare riguardo alla sicurezza e alla scalabilità. Non è raro in ambito commerciale che questo file system supporti oltre 50.000 clients. AFS utilizza Kerberos per eseguire le autenticazioni, e implementa liste per il controllo degli accessi alle singole cartelle per utenti e gruppi (Access Control List). Il *caching* a livello del client permette un miglioramento delle prestazioni, limitando gli accessi al file server e i sovraccarichi della rete. Nell'AFS i file vengono mantenuti nella cache su richiesta della singola workstation. Le operazioni di lettura e scrittura vengono eseguite direttamente sulla copia mantenuta nella cache locale. Quando un file che ha subito delle modifiche viene chiuso, la parte modificata viene copiata sul file presente sul server. La consistenza della cache viene mantenuta da un meccanismo denominato callback. Quando un file si trova nella cache di un client, il server ne prende nota, e in caso tale file venga modificato, si impegna a notificarlo ai client che ne hanno una copia. Tale meccanismo viene a cadere, e quindi riavviato, ogni qualvolta un server, un client o un errore nella rete creino un time-out. Il ripristino del callback consiste nel controllo dello stato dei file nelle cache dei vari client e non richiede la ritrasmissione dei file.

Una conseguenza della strategia di sicurezza adottata nel file system è che AFS non supporta un vasto database condiviso, o l'aggiornamento dei record di file condivisi tra client. Questo è una deliberata scelta progettuale basata sulle necessità dell'ambiente di elaborazione utilizzato nell'università.

Una caratteristica peculiare dell'AFS è il volume, una organizzazione gerarchica ad albero dei file e delle cartelle di ordine inferiore. I volumi sono creati dall' amministratore e collegati a specifici nomi di percorso in una cella AFS. Una volta creato, gli utenti del file system possono creare cartelle e file, senza preoccuparsi della locazione fisica dei dati. Un volume può avere una quota assegnata, in modo da limitare lo spazio a disposizione per l'immagazzinamento dei dati. A seconda delle necessità, l'amministratore può trasferire un volume su un diverso server o una diversa locazione del disco, tale procedura risulta trasparente agli utenti.

I volumi AFS possono essere duplicati in copie di sola lettura. Quando viene eseguito un accesso a tali copie, i client possono accedere a copie dei file in sola lettura. Inoltre gli utenti non sono coscienti della locazione spaziale della propria copia di sola lettura; l'amministratore può creare e spostare tali volumi a secondo le necessità. AFS garantisce che i dati presenti in un volume di sola lettura siano consistenti con quelli del volume modificabile al momento della creazione del volume stesso.

Lo spazio dei nomi dei file su una postazione di Andrew si divide in condiviso e locale. Lo spazio dei nomi condiviso è identico per tutte le postazioni. Lo spazio dei nomi locali è unico per ogni postazione. Esso contiene file temporanei a utilizzo della singola postazione e link simbolici a file presenti nello spazio dei nomi condiviso.

Andrew file system ha profondamente influenzato la versione 4 del popolare file system della Sun Microsystems, Network File System (NFS). Inoltre, una variante di AFS, il Distribuited File System (DFS) è stato adottato dalla Open Software Foundation nel 1989 come parte del loro ambiente di elaborazione distribuito.

#### 4.2.5.1 AFS ai LNF

L'implementazione di AFS attualmente utilizzata ai LNF è OpenAFS. La distribuzione opensource, contiene sia la parte server che la parte client praticamente per tutte le piattaforme di OS unix, Windows e Mac.

Quest'ultimo aspetto è fondamentale per l'uso in un ambiente come quello universitario o di ricerca in genere, in quanto tali ambienti sono tipicamente molto eterogenei.

Ai LNF ad esempio sono presenti client di tipo Windows, Mac OS e Linux. L'utente può accedere ai propri spazi di storage in modalità indipendente dal client e dalla piattaforma, sia lavorando dalla propria postazione di lavoro client, sia lavorando su un server centrale messo a disposizione dal Servizio di Calcolo. Inoltre l'accesso è garantito ed efficiente anche stando connessi sulla rete remotamente da qualunque postazione internet.

I server AFS in produzione sono 10. Ognuno di essi è direttamente connesso alla Storage Area Network vista nel paragrafo 3.4. La gran parte delle LUN servite dalla Storage Area Network sono servite agli AFS server che importano tali LUN come dischi fisici.

I server AFS organizzano tale spazio sotto forma di volumi e lo esportano sulla rete come file system distribuito sotto la cella *lnf.infn.it*

Lo spazio è indirizzabile sotto l'albero globale, visibile e raggiungibile da tutto il mondo: */afs/lnf.infn.it*

In particolare sono stati creati i seguenti sottorami

1. */afs/lnf.infn.it/user*
   dedicato alle home directory degli utenti

2. */afs/lnf.infn.it/project*
   dedicato alle aree sperimentali e ai progetti di ricerca
3. */afs/lnf.infn.it/system*
   dedicato alle aree di sistema

Tali aree di lavoro sono protette da ACL (Access Control List). La granularità della protezione arriva alla singola directory (o subdirectory). Una volta che l'utente si è autenticato (via kerberos) potrà avere accesso alle aree a lui dedicate, in accordo con le ACL impostate. Ogni utente può cambiare le ACL delle proprie aree. Il responsabile informatico di gruppo (detto responsabile di account) può cambiare le ACL delle aree comuni di esperimento o di progetto. Grazie alla flessibilità delle ACL, ogni utente può discriminare facilmente chi può accedere alle proprie directory e, al limite, renderle pubbliche, ovvero leggibili da tutto il mondo.

Anzi, per scelta implementativa, quando viene creata una nuova *username*, a questa nuova utenza viene associata un'area di storage AFS limitata da opportuna quota iniziale (ovviamente espandibile), indirizzabile con */afs/lnf.infn.it/user/<username>* con dentro già create 3 directory con le relative ACL iniziali:

1. *private*
   leggibile e scrivibile solo dal nuovo utente
2. *public*
   scrivibile dall'utente e leggibile da tutto il mondo
3. *www*
   scrivibile dall'utente e leggibile dai server WWW dei LNF (per esportare pagine web alla url *http://www.lnf.infn.it/~<username>*)

Tutte le aree AFS sono sottoposte ad una una politica di snapshot e di backup. Una peculiarità di AFS è quella di creare *snapshot* del volume, ovvero una fotografia del contenuto del volume ad un certo momento. Gli snapshot, volumi di sola lettura (*readonly*), richiedono poco spazio disco in quanto non duplicano i dati, bensì contengono i puntatori ai blocchi dei volumi di origine.

Gli snapshot possono essere utilizzati dagli utenti come backup dei dati. Il recupero può avvenire puntando l'area */afs/lnf.infn.it/backup* dove vengono posizionati tutti i volumi si snapshot.

Gli snapshot delle home directory degli utenti sono programmati a cadenza giornaliera (durante la notte), mentre gli snapshot delle aree di progetto e di esperimento sono programmati a cadenza settimanale.

Ciò implica che se un utente ha erroneamente perso un file, può autonomamente recuperarne la versione del giorno precedente con un semplice comando di copia da disco a disco.

Tuttavia questa procedura presenta un grosso limite. Se l'utente infatti si accorge di aver perso il file dopo 2 giorni, lo stesso file non è più reperibile nel volume di snapshot.

Per sopperire a tale limite sono programmate le procedure di backup vero e proprio, che avviene su nastro con gli strumenti e le metodologie viste nel paragrafo 3.4

Anche i backup hanno cadenza giornaliera per le home directory degli utenti e cadenza settimanale per le aree di progetto e di esperimento. L'accesso ai dati di backup deve essere richiesto al Servizio di Calcolo ed è effettuabile solo dall'amministratore dei sistemi e dello storage.

Le politiche di mantenimento dei file sul sistema di backup sono diverse a seconda del volume. Nel caso delle home directory degli utenti il sistema di backup (Tivoli Storage Manager) prevede di mantenere fino a 15 versioni dello stesso file e fino a 365 giorni dopo la rimozione dal disco. In tal modo l'utente è certo di ritrovare un file anche se lo ha perso fino ad un anno prima, purché non lo abbia modificato più di 15 volte. Nel caso lo abbia modificato più di 15 volte, troverà fino alla quindicesima versione precedente.

### 4.2.6 Sistema di backup e archiviazione (Tivoli Storage Manager)

Tivoli Storage Manager (TSM) è un sistema s/w di classe enterprise per la gestione centralizzata del backup e dell'archiviazione dei dati. Il software abilita l'utente ad inserire dati sia attraverso la funzionalità di *backup* sia attraverso la funzionalità di *archive* e analogamente consente il *restore* e il *retrieve*.

Più in dettaglio TSM fornisce una vasta gamma di tipologie di protezione dei dati attraverso l'automazione di procedure basate su politiche predefinite dall'amministratore:

- *Backup and recovery.*
- *Archiving and retrieval.*
- *Disaster recovery.*
- *Space Management.*
- *Online database and application protection.*

Il sistema è basato su un'architettura client-server. Il server connesso alla storage area network, vede le unità a nastro e le librerie come direttamente connesse. Inoltre alloggia al suo interno un'area di disco per la gestione di un database e della cache.

La comunicazione tra client e server avviene attraverso la rete LAN. Tutte le procedure di backup e archiviazione passano quindi per la rete Ethernet. Quando viene lanciata una procedura di backup o archiviazione, il server deposita le informazioni su un'apposita area locale di cache (area di stage su disco) e solo successivamente si occuperà di trasferire le informazioni immagazzinate nella cache, verso le unità a nastri, ottimizzando in tal modo i tempi d'attesa per l'esecuzione delle archiviazioni (dovuti ad esempio all'occupazione delle stesse unità a nastri).

Nel frattempo il server registra tutte le informazioni relative alle transazioni sull'apposito database. Il database contiene tutti i puntatori ai file di cui si richiede il backup o l'archiviazione su nastro con la relativa posizione nella libreria, ivi incluso il numero della cassetta su cui è archiviato.

È evidente che tale database è particolarmente critico. La perdita del DB porterebbe alla perdita delle informazioni sui nastri. Per questo il database è replicato su disco e anche archiviato su nastro.

Il sistema prevede che l'amministratore programmi la cronologia dei backup e delle archiviazioni delle aree di storage da proteggere, da svolgere in modalità automatizzata, non assistita, tipicamente durante le ore di minore attività lavorativa.

#### 4.2.6.1 TSM ai LNF

Ai LNF una macchina IBM P55 (quadriprocessore Power PC) è dedicata al sistema di backup e archiviazione. Tale macchina è partizionabile in hardware e ciò ha consentito di utilizzarla come se fossero due server distinti.

Quindi logicamente ci sono due server TSM, ognuno dei quali con la sua area di stage, il suo database e la sua libreria. Uno è connesso alla libreria IBM e l'altro alla libreria Storagetek (rif. paragrafo 3.4).

Ad ogni insieme omogeneo di server o di servizi viene dedicato un pool di cassette per il backup e l'archiviazione. Il pool è dimensionato per contenere decine di copie dello stesso filesystem di ciascun client. Lo stesso vale per le aree AFS dedicate agli utenti o agli esperimenti.

Tutte le notti sono programmati backup di tutti i filesystem più critici, in modo da garantire e salvaguardare il contenuto delle aree di storage da qualunque evento dannoso, sia esso un errore umano di cancellazione involontaria di un singolo file o di un intero tree di directory, sia esso un problema tecnico comportante la perdita di interi volumi.

Le politiche di backup prevedono tipicamente un salvataggio giornaliero delle home directory degli utenti e un salvataggio settimanale delle aree di esperimento. Il restore di un file è sempre possibile alle seguenti condizioni:

Ad esempio, sulle home directory degli utenti (aree AFS):

- Caso di file modificato su disco (ma ancora presente):
  - rimangono su nastro le ultime 15 versioni per una durata massima di 400 giorni (nota: la 16$^{a}$ versione viene persa prima della scadenza dei 400 giorni, quindi entro i 400 gg si ha garanzia di trovare solo le 14 versioni precedenti).

- Caso di file perso su disco:
  - se il file non è stato mai modificato rimane su nastro per 2000 giorni
  - se il file è stato modificato restano le 4 versioni precedenti per 400 giorni e l'ultima versione per 2000 giorni

## *4.3 Servizi informatici fondamentali*

I servizi di base visti in precedenza sono necessari come intelaiatura sulla quale basare tutti i servizi informatici forniti all'utenza. In questo capitolo si tratteranno invece quei servizi di alto livello, spesso irrinunciabili, direttamente fruibili dagli utenti, quali ad esempio il servizio World Wide Web, il servizio di posta elettronica e il servizio di stampa.

Essendo questi servizi ben noti, nella trattazione ci si soffermerà prevalentemente nel descrivere il modello implementato ai LNF piuttosto che il protocollo di funzionamento dei servizi stessi.

### 4.3.1 Il World Wide Web

Il World Wide Web (ragnatela mondiale) è uno spazio elettronico e digitale di Internet destinato alla pubblicazione di informazioni, particolarmente idoneo per la divulgazione di contenuti multimediali.

Il successo della 'ragnatela mondiale' è stato tale che attualmente, per la maggior parte degli utenti, essa coincide con la rete stessa. Sebbene questa convinzione sia tecnicamente scorretta, è indubbio che gran parte dell'esplosione del 'fenomeno Internet' a cui abbiamo assistito in questi ultimi anni sia legata proprio alla diffusione di questo strumento.

Il successo di World Wide Web ha naturalmente suscitato l'interesse di una enorme quantità di nuovi autori ed editori telematici, interesse che ha determinato dei ritmi di crescita più che esponenziali. Nel 1993 esistevano solo duecento server Web, oggi ce ne sono milioni.

Il suo utilizzo di massa lo rende uno strumento economico e irrinunciabile per tutte le organizzazioni che abbiano interesse a divulgare informazioni e a fornire servizi informatici ad un vasto pubblico.

#### 4.3.1.1 Il World Wide Web ai LNF

Ai LNF il servizio Web è stato implementato con lo scopo di raggiungere i 3 seguenti obiettivi:

1. Fornire un servizio efficiente, affidabile e altamente disponibile.
2. Fornire a ciascun utente, gruppo, o esperimento, la possibilità di divulgare informazioni scientifiche in maniera completamente autonoma, attraverso la gestione di uno spazio di storage privato ed individuale ove inserire i contenuti.
3. Fornire a tutti gli utenti ed in particolare ad uno specifico servizio nazionale (Dataweb) un'infrastruttura locale di Data Warehouse per la gestione di dati scientifici, amministrativi e procedurali, e per consentire la pubblicazione di applicazioni web.

Per raggiungere il primo obiettivo sono stati implementati 4 server web in modalità ridondata, su macchine virtuali XEN con sistema operativo Scientific Linux versione 5. Il servizio Web (il demone httpd) è fornito dal s/w opensource Apache versione 2.2.

I quattro server condividono le stesse aree di storage attraverso il protocollo AFS e hanno file di configurazione sempre perfettamente allineati. Questo permette a ciascun server di svolgere il lavoro di ogni altro. Ovvero i server sono utilizzati in modalità di bilanciamento del carico (*load balancing*).

Il bilanciamento del carico è realizzato tramite una specifica funzionalità degli switch centrali della LAN dei LNF, denominata *Server Load Balancing* (SLB) descritta nel successivo paragrafo 4.3.4. La stessa funzionalità garantisce contemporaneamente anche un servizio di alta disponibilità.

Il secondo obiettivo viene raggiunto attraverso la pubblicazione di specifiche aree afs dedicate ai servizi e agli esperimenti, nonché alla pubblicazione di una particolare sottodirectory delle home directory di ciascun utente (rif. paragrafo 4.2.5.1). Ad esempio le due aree:

*/afs/lnf.infn.it/project/www/<experiment>/*
*/afs/lnf.infn.it/user/<username>/www/*

Sono pubblicate ai rispettivi indirizzi web:

*http://www.lnf.infn.it/<experiment>*
*http://www.lnf.infn.it/~<username>*

Infine il terzo obiettivo viene raggiunto attraverso la realizzazione di vari database server. In particolare esistono due database server Opensource MySQL e un database server Oracle.

Per l'alta affidabilità del database server MySQL si è scelto di adottare due server fisici in modalità Master – Replica. La modalità Master – Replica non prevede il bilanciamento del carico. Il database Replica è sempre sincronizzato con il database Master (quello dove avvengono le letture e le scritture). Tuttavia nel caso di fallimento del Master, la Replica può prenderne il posto previo intervento del sistemista. Per questo non si può parlare di alta disponibilità.

Il database server Oracle è di installazione molto più recente ed è implementato in modalità RAC (Real Application Cluster) su due macchine fisiche dedicate allo scopo in modalità esclusiva. In questo caso la soluzione fornisce implicitamente sia il bilanciamento del carico, sia l'alta affidabilità, sia l'alta disponibilità.

L'idea è quella di migrare, con il tempo, tutti i dati dal database MySQL a quello Oracle, e modificare tutte le applicazioni perché utilizzino il database server cluster della Oracle.

In aggiunta ai servizi già descritti, ai LNF è anche installato un server di audio e video streaming Adobe Flash Media Server per la fornitura di contenuti multimediali.

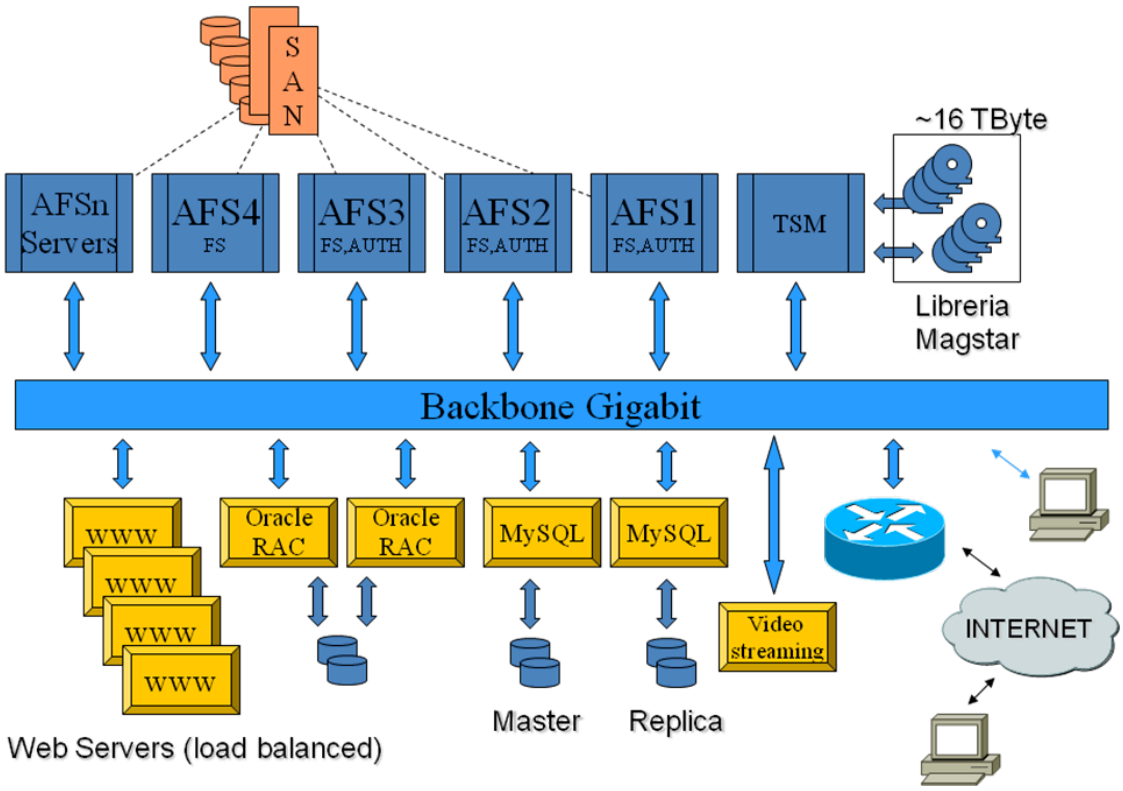



**Figura 4-3: Web Servers e Data warehouse ai LNF**

Basato sulla struttura rappresentata in figura, il servizio Web dei Laboratori Nazionali di Frascati svolge una serie di servizi che hanno rilevanza locale, ma anche nazionale. I domini web (o virtual host) che serve sono i seguenti:

*http://www.lnf.infn.it/*
*http://www.ac.infn.it/*
*http://www.infn.it/*
*http://www.asimmetrie.it/*
*http://scienzapertutti.lnf.infn.it/*
*http://www.nottedellaricerca.eu/*
*http://hadronphysics.infn.it/*
*http://www.lp01.infn.it/*
*http://frontierscience.lnf.infn.it/*

In particolare, sul virtual host *www.infn.it*, il Servizio Dataweb (dell'Amministrazione Centrale) ha implementato il portale web dell'INFN. Tramite il portale, previa autenticazione basata sul sistema di autenticazione nazionale dell'INFN e previa identificazione del client a mezzo di *cookie* (ideata, progettata e sviluppata all'interno del Servizio di Calcolo dei LNF), ogni utente INFN può accedere ad una serie di dati e di servizi amministrativi e procedurali, sviluppati dallo stesso servizio Dataweb. Ad esempio è possibile eseguire una procedura di associazione, oppure presentare un bilancio preventivo, ricevere assegnazioni economiche, etc..

Oltre alla struttura rappresentata nella precedente figura, il Servizio di Calcolo dei LNF gestisce anche alcune altre macchine (virtuali) installate per fornire in modalità esclusiva altri servizi Web specifici di rilevanza nazionale (INFN):

- ***Indico***: sviluppato al CERN è un'applicazione web per la gestione delle conferenze e dei seminari, permette di gestire l'agenda e il materiale documentale delle presentazioni:
  *http://agenda.infn.it/*
- ***Dokuwiki***: software opensource principalmente utilizzato per la creazione e la gestione di documentazione di ogni tipo. Particolarmente adatto per gli sviluppatori o per i gruppi di lavoro per lasciare documentazione e traccia del lavoro svolto.
  *http://wiki.infn.it/*
- ***Oracle Collaboration Suite***: software della Oracle che fornisce un set integrato di strumenti collaborativi costruiti su un'unica piattaforma di classe enterprise. Permette all'utente di costruire un proprio calendario, un proprio framework per la gestione di progetti e lavori, di gestire documenti, di condividere il desktop, etc..
  *http://ocs.infn.it/*
  Nota: dato il costo della soluzione, si stanno valutando strumenti alternativi come ***Alfresco***.

### 4.3.2 Il servizio di posta elettronica

La posta elettronica o email (*electronic mail*) è un servizio Internet grazie al quale ogni utente può inviare o ricevere dei messaggi. È l'applicazione Internet più conosciuta e più utilizzata attualmente.

È la controparte digitale ed elettronica della posta ordinaria e cartacea. A differenza di quest'ultima, il ritardo con cui arriva dal mittente al destinatario è normalmente di pochi secondi o minuti.

Per la realizzazione degli esperimenti di ultima generazione è necessaria la collaborazione di centinaia di scienziati e tecnici. È evidente che lo strumento della posta elettronica acquista una tale importanza per una veloce e continua comunicazione scientifica, che lo rende uno degli strumenti informatici irrinunciabili per la ricerca.

I componenti fondamentali del sistema di e-mail sono:

- i client (tecnicamente MUA, *Mail User Agent*), utilizzati per accedere ad una casella di posta elettronica e per inviare messaggi;
- i server, che svolgono due funzioni fondamentali:
  - ricevere i messaggi in arrivo ed in partenza e smistarli (MTA, *Mail Transfer Agent*);
  - immagazzinare i messaggi per uno o più utenti (MS, *Message Store*).

I protocolli tipicamente impiegati per lo scambio di email sono l'SMTP, usato per l'invio, la ricezione e l'inoltro dei messaggi tra server, e POP e IMAP, usati per la ricezione e consultazione dei messaggi da parte degli utenti.

I client richiedono la configurazione dei server da contattare, e sono quindi adatti principalmente a computer usati regolarmente. È anche molto diffusa la possibilità di consultare una casella e-mail attraverso il web.

#### 4.3.2.1 Il servizio di posta elettronica ai LNF

Ai LNF il servizio di posta è svolto da otto server, di cui due implementati su macchine fisiche e sei su macchine virtuali. In particolare cinque server (due fisici e tre virtuali) sono dedicati allo smistamento dei messaggi e altri tre (tutti virtuali) dedicati alla consegna locale e all'immagazzinamento (*store*).

**Server per lo smistamento (Mail Transfer Agent)**

***Mail eXchanger***

Il protocollo di smistamento è il Simple Mail Transfer Protocol (*SMTP*).

SMTP è un protocollo relativamente semplice, testuale, nel quale vengono specificati uno o più destinatari di un messaggio; verificata la loro esistenza, il messaggio viene trasferito. Il protocollo SMTP utilizza il protocollo di livello transport TCP. Il client apre una sessione TCP verso il server sulla porta 25.

SMTP supporta soltanto i caratteri a 7 bit (codifica ASCII). Questo rappresenterebbe un limite per la trasmissione di caratteri con encoding diverso da quello inglese (ad esempio per la trasmissione delle lettere accentate), oppure per la trasmissione di file binari o di documenti multimediali (immagini, suoni video) a 8 bit.

MIME (*Multipurpose Internet Mail Extensions*) risolve questo limite del protocollo SMTP, definendo i meccanismi per mandare altri tipi di informazioni contenute nei messaggi di posta elettronica. Tipicamente viene utilizzato un sistema di conversione di ogni byte da 8-bit a 7-bit, in modo da ricondurlo ad un carattere con codifica ASCII, tale da poter essere trasmesso dal protocollo SMTP.

Per associare il server SMTP a un dato nome di dominio (DNS) si usa un Resource Record di tipo MX (Mail eXchange) che è un particolare tipo di record DNS.

Quando un utente di internet invia un mail all'indirizzo

*<e-mail.adress>@lnf.infn.it*

il server SMTP utilizzato dall'utente internet accetterà il messaggio ed effettuerà una query DNS per individuare i server SMTP che svolgono il ruolo di mail exchanger per il dominio *lnf.infn.it.* Otterrà come risposta che i server possibili sono due (con pari priorità), quindi sceglierà uno dei due per inoltrare il messaggio. In caso di fallimento, sceglierà l'altro server. La ridondanza dei due server di Frascati garantisce un'alta disponibilità del servizio.

I server MX del dominio *lnf.infn.it* sono configurati con il software opensource *Berkeley Sendmail* per svolgere le seguenti funzioni:

- inoltro di messaggi dall'interno della LAN verso l'interno
- inoltro di messaggi dall'interno della LAN verso l'esterno
- inoltro di messaggi dall'esterno verso l'interno
- inoltro di messaggi dall'esterno verso l'esterno (esclusivamente se il mittente è autenticato)

Il protocollo SMTP nasceva quando la rete Internet era privilegio di pochi e non c'erano ancora i tentativi di abuso informatico che si registrano ormai da anni. Per questo il protocollo nasceva senza particolari protezioni e non prevede ad esempio che il mittente del mail debba autenticarsi per inoltrare il messaggio. Questa scelta tecnica, che ha funzionato per anni, oggi si rivela un limite del protocollo, che consente a chiunque, nascondendo la propria identità, di inviare qualunque tipo di messaggio.

Per questo motivo il protocollo SMTP veicola grandi quantità di messaggi indesiderati, spesso pubblicitari (detti *spam*), a volte contenenti addirittura codice maligno (*malware*), ovvero software creato con il solo scopo di causare danni più o meno gravi al computer su cui viene eseguito. Il volume di spam oggi è talmente elevato da rappresentare oltre il 95% del traffico dei mail ricevuti dai mail server (Mail exchanger) dei LNF.

***Milter***

Per difendersi da questa tipologia di attacchi, ai LNF è stata installata una macchina (virtuale) con funzioni di filtro. Tutti i messaggi pervenuti ai mail exchanger, prima di essere smistati, passano per tale macchina utilizzando un protocollo specifico detto *milter*.

Un primo filtro sui mail pervenuti viene effettuato dal milter con un sistema di *greylisting*, che opera rifiutando in prima istanza tutti i messaggi ricevuti, confidando sul fatto che gli spammer non eseguono un secondo tentativo. Con questo sistema già l'80% circa dei mail spam viene rigettato.

Il milter dei LNF è configurato anche con un software specifico della Sophos (*PureMessage*), in grado di effettuare verifiche sul contenuto dei messaggi e classificarli eventualmente quali spam o malware.

Dato che la tipologia di spam e di malware è in continua evoluzione, questa macchina è connessa con un centro servizi della Sophos per effettuare gli aggiornamenti utili per la classificazione. Tali aggiornamenti avvengono con cadenza oraria, in modo da garantire una maggiore protezione.

Il software della Sophos, oltre a riconoscere i malware e gli spam noti, è in grado di fare delle valutazioni su ciascun messaggio in transito, anche con metodi euristici, analizzando sia l'intestazione del messaggio, sia le parole contenute nel corpo, attribuendo un indice di probabilità che quel messaggio sia o meno spam o malware, a seguito del confronto con un parametro di soglia preimpostato.

I mail a cui viene attribuito un indice di probabilità molto maggiore del parametro di soglia, vengono rigettati dal sistema. I messaggi a cui viene attribuito un indice di probabilità vicino al parametro di soglia preimpostata, vengono etichettati

come spam (inserendo la stringa [==SPAM==] nell'oggetto del messaggio) e inviati comunque al destinatario.

Grazie a questa soluzione combinata viene filtrato il 98% del traffico mail e praticamente il 100% dei malware in transito sui mail exchanger dei LNF.

***SMTP***

Il milter svolge un grande lavoro sulla rete, dovendo smistare e filtrare tutti i messaggi in transito per il dominio dei LNF. Si sono sperimentate spesso delle inefficienze del milter per via del volume di traffico di posta che deve analizzare. Queste inefficienze si evidenziano come ritardi nella risposta o, in taluni casi, addirittura come mancata risposta (server time-out).

Questo fatto non rappresenta un problema ai fini del funzionamento del servizio di mail (da server SMTP a server SMTP), in quanto i mail exchanger fanno innumerevoli tentativi, con cadenza di pochi minuti, prima di desistere dall'inoltro del mail (tipicamente dopo 5 giorni, con avviso al mittente).

Tuttavia ha creato disagi ai client interni alla LAN che, in alcuni casi, ricevevano direttamente una segnalazione d'errore al mittente da parte dei server MX.

Per ovviare a questo inconveniente sono state installate altre due macchine virtuali che svolgono la funzione di SMTP server di inoltro per gli utenti dei LNF. Tali macchine sono Scientific Linux 5, in cluster fra loro, e accettano i mail inviati dai client (MUA) per inoltrarli ai mail exchanger locali. Anch'esse garantiscono un servizio di alta affidabilità e disponibilità grazie alla funzionalità di Server Load Balancing descritta nel successivo paragrafo 4.3.4.

I server SMTP sono configurati, con il software opensource *Berkeley sendmail*, in modo da accettare su connessione cifrata (usando il *Transport Layer Secure*), tutti i mail provenienti da utenti autenticati tramite certificato digitale X.509, tramite credenziali d'accesso Kerberos (Username e Password), oppure tramite Kerberos/GSSAPI ovunque connessi su internet.

**Server per l'immagazzinamento (Message Store)**

***IMAP***

L'Internet Message Access Protocol (IMAP) è un protocollo di comunicazione per la ricezione di e-mail, ovvero per fornire l'accesso alle caselle di posta elettronica ad un client (dotato di MUA). L'attuale versione è la "4 revision 1". La porta predefinita del demone IMAP sull'host è la 143 (protocollo TCP). Se si utilizza una connessione sicura tramite SSL, allora la porta è la 993.

Il protocollo IMAP è stato progettato per consentire ad un client di accedere alle proprie caselle di posta elettronica gestite e conservate dai server di immagazzinamento. IMAP è nato come alternativa al protocollo POP (vers. 3). Quest'ultimo scarica la posta direttamente sul client, cancellandola dal server; mentre con il protocollo

IMAP è possibile conservare copia delle proprie e-mail sul server per accedervi in un secondo momento anche da altri computer.

Il server IMAP permette la catalogazione dei mail in opportune cartelle separate (*folder*) organizzate ad albero, nell'ambito dello spazio di storage gestito dal server e dedicato al singolo utente (*quota di storage per utente)*. È evidente che in questo modo è necessario che i server IMAP gestiscano un'ampia area di storage, adeguata al volume di mail che mediamente gli utenti mantengono nell'area di storage a loro dedicata.

Ai LNF l'area dedicata al Message Store è attualmente di 2 TB. È un'area servita dalla Storage Area Network ai tre server IMAP che la condividono attraverso OCFS2 (Clustered File System opensource della Oracle versione 2).

I server IMAP sono realizzati con il software opensource Dovecot installato su un cluster di tre nodi (macchine virtuali) con Sistema Operativo Scientific Linux 5. Anche in questo caso l'alta disponibilità è garantita dalla funzionalità di Server Load Balancing descritta nel successivo paragrafo 4.3.4. Inoltre, sugli stessi server, è installato e configurato SMTP *Berkeley sendmail* che è necessario per la consegna dei messaggi nelle caselle di posta elettronica degli utenti destinatari.

Le caratteristiche principali con cui tali server IMAP sono stati installati sono le seguenti:

- Organizzazione dei mail con formato *maildir*: è un sistema di immagazzinamento dei mail che prevede che ogni messaggio venga tenuto in un file separato con un nome unico, e ogni folder sia una directory. Dovecot, oltre a gestire tale formato, ha inoltre la particolarità di generare dei file di indici per accelerare l'accesso ai mail. Questa organizzazione rispetto alla precedente *mbox* (che prevedeva di avere tutti i messaggi di un singolo folder scritti in un grande file sequenziale), ha permesso di migliorare i tempi di accesso ai mail di un ordine di grandezza.
- Gestione della quota disco *per utente*: ad ogni utente è riservato uno spazio minimo di 1GB, ovviamente espandibile.
- Accesso alla propria casella di posta elettronica da qualunque postazione (interna o esterna alla LAN dei LNF) tramite un protocollo cifrato *Secure Socket Layer*.
- Autenticazione integrata con i server kerberos del realm LNF.INFN.IT: gli utenti possono accedere alla loro casella di posta elettronica inserendo le credenziali uniche rilasciate dal Servizio di Calcolo per l'accesso a tutti i servizi informatici di dominio.
- Autenticazione Kerberos/GSSAPI: avendo un MUA che supporta il protocollo GSSAPI, non è necessario che l'utente inserisca le proprie credenziali d'accesso se è già in possesso di un ticket kerberos sul proprio client (*Single Sign On*).

Il sistema di mail così realizzato è un sistema efficiente e scalabile. È infatti possibile aggiungere facilmente altri nodi al cluster SMTP o al cluster IMAP per aumentare le capacità di elaborazione dei due servizi.

**Client (Mail User Agent)**

Il Servizio di Calcolo dei LNF suggerisce di utilizzare un client d'accesso opensource: Mozilla Thunderbird. Questo client ha la peculiarità di esistere per tutte le piattaforme (Windows, MacOS, Linux, etc.) ed ha la caratteristica di essere molto funzionale. Inoltre implementa tutte le caratteristiche sopra elencate, necessarie per integrarsi con i server IMAP dei LNF.

In realtà funzionano anche altri client d'accesso, quali ad esempio Mac OS Mail (che arriva già con il Sistema Operativo) oppure Microsoft Outlook. Tuttavia potrebbero esserci problemi di integrazione con tutte le funzionalità implementate sui server IMAP, prima fra tutte l'autenticazione GSSAPI.

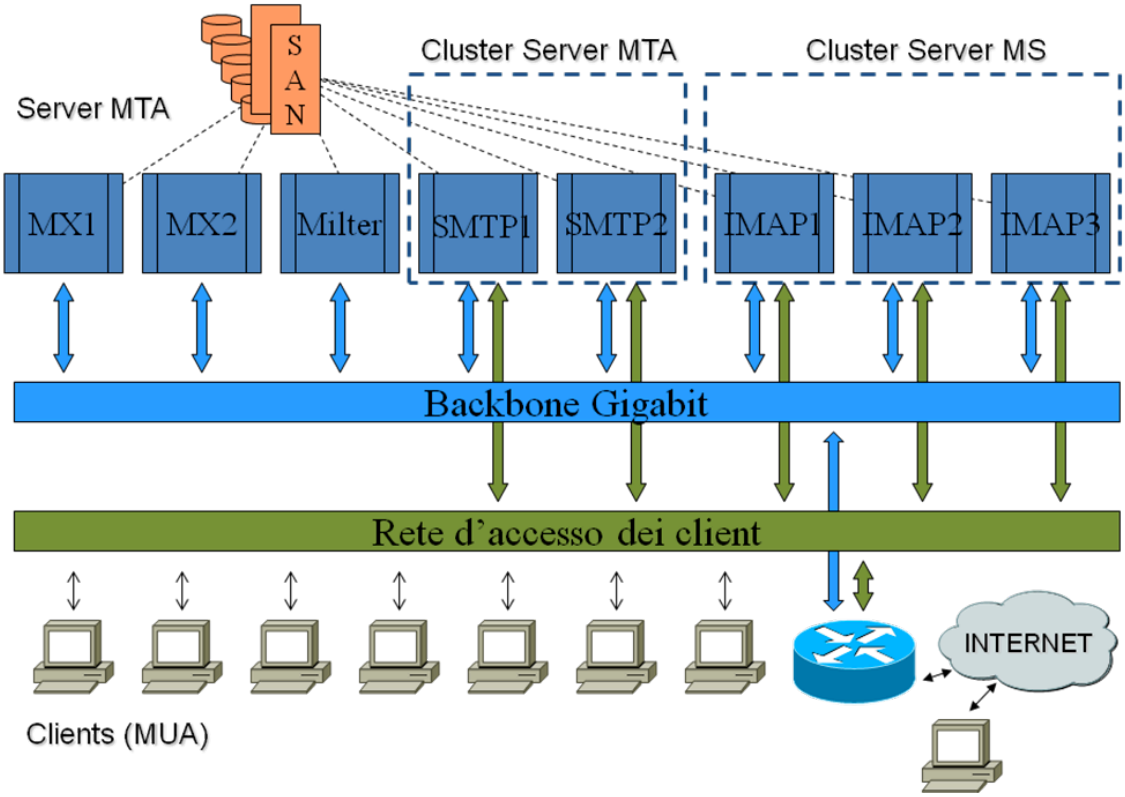



**Figura 4-4: Mail Servers ai LNF**

## 4.3.3 Il servizio di stampa

Per la stampa su carta di documenti in formato elettronico è spesso richiesto un servizio centralizzato per il controllo delle code di stampa e per il controllo delle stampanti.

In effetti non è consigliabile che le stampanti di volume vengano viste sulla rete da tutti i client in modalità diretta. In questo caso infatti, in caso di stallo, sarebbe difficile individuare la provenienza del problema.

Quindi ci si affida spesso ad un Printing Server centralizzato che mappa le stampanti dipartimentali in modalità diretta ed esclusiva, e tutti i client passano per questo server centrale per stampare documenti su tali stampanti di volume.

#### 4.3.3.1 Il servizio di stampa ai LNF

Ai LNF il servizio di stampa è basato su soluzione Microsoft Windows Server 2008 Enterprise Edition. Due macchine fisiche (biprocessori dual-core) sono dedicate al dominio Windows che è attualmente in via di implementazione ed integrazione con le altre infrastrutture informatiche. Grazie alle peculiarità intrinseche della soluzione Microsoft, le due macchine sono in *Failover Cluster*. Pur non permettendo il bilanciamento di carico, il failover cluster permette di gestire molti servizi informatici, rendendoli altamente affidabili e disponibili. Infatti i servizi gestiti dal failover cluster sono serviti da uno solo dei due nodi di cluster, ma sono in grado di passare in pochi istanti e in maniera del tutto trasparente agli utenti, all'altro nodo non appena il primo fallisce.

Molti servizi centrali del Windows domain sono gestiti attraverso tale cluster e, tra questi, il servizio centrale di stampa.

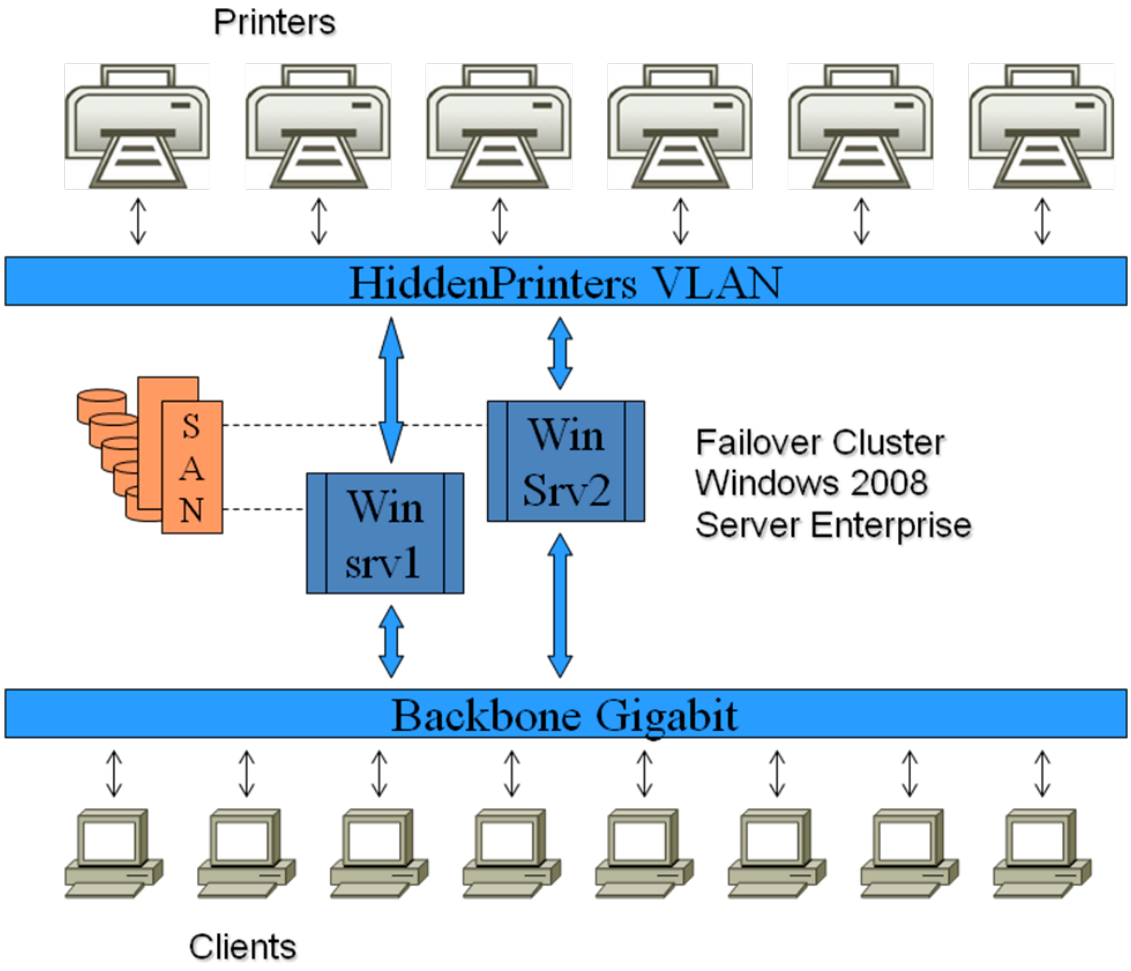



**Figura 4-5: Schema del servizio di stampa ai LNF**

Il servizio di stampa di Windows Server 2008 permette di mappare tutte le stampanti dipartimentali ed esportarle sulla rete tramite tutti i protocolli conosciuti per la gestione delle code di stampa. Ai LNF le stampanti servite sono una dozzina; sono tutte mappate dal server attraverso lo *Standard TCP/IP Port Protocol* e servite ai client attraverso due protocolli: *Netbios, LPR/LPD*.

Le stampanti dipartimentali sono mappate in rete su una VLAN a cui i printing server hanno accesso esclusivo in modo da evitare che i client possano vedere le stampanti direttamente.

I client quindi potranno mappare le code di stampa accedendo al server centrale utilizzando uno dei tre protocolli a disposizione. Per i client Windows si consiglia di usare il protocollo Netbios; infatti in tal caso il client non ha bisogno di cercare i driver specifici per la stampante mappata, in quanto quelli corretti vengono serviti direttamente dal server al client al momento della prima installazione della coda sul client. Questa funzionalità è molto apprezzata dall'utenza, in quanto, oltre a semplificare le operazioni di installazione delle code di stampa, risulta estremamente utile per ridurre i fail delle stampanti dovuti ad erronea installazione dei driver sui client.

### 4.3.4 Alta disponibilità e Service Load Balancing

L'alta disponibilità dei servizi è garantita da una specifica funzionalità dei due switch centrali del Servizio di Calcolo (descritti nel paragrafo 2.5.10), denominata *Server Load Balancing* (SLB).

Tramite tale funzionalità, è possibile definire l'indirizzo di un servizio virtuale in modo da rappresentare un gruppo omogeneo di server reali in ambiente cluster o server farm. I client richiedono il servizio all'indirizzo IP del server virtuale secondo quanto indicato dal DNS locale; in realtà la richiesta arriva ad uno dei due switch centrali, che provvede ad eseguire un Network Address Translation, per ridirigerla immediatamente ad uno dei server reali appartenente al gruppo.

Lo smistamento delle connessioni verso i server reali avviene con latenze estremamente basse secondo un algoritmo di round-robin, ovvero le nuove connessioni vengono ridirette ai server reali secondo una selezione circolare. Tuttavia le nuove connessioni tra un generico client e il server virtuale vengono ridirette dallo switch sempre verso lo stesso server reale, qualora lo stesso client abbia già altre connessioni stabilite, oppure ne abbia già stabilite in precedenza (ovvero in un periodo di tempo configurabile).

Gli switch sono inoltre in grado di verificare lo stato dei server reali a cui vengono ridirette le richieste dei client, monitorandone le risposte e, in caso di fail di uno di essi, lo escludono dalla selezione circolare prevista dall'algoritmo di round-robin, ovvero ridirigono le richieste ai rimanenti server reali attivi appartenenti al gruppo.

Un'altra caratteristica della funzionalità di SLB è quella di poter essere gestita dal servizio HSRP (rif. paragrafo 2.5.14), ovvero di poter continuare a funzionare anche in caso di fail di uno dei due switch centrali che condividono la funzionalità stessa.

Questa importante funzionalità degli switch cisco consente di raggiungere due principali obiettivi: alta disponibilità e bilanciamento del carico dei server.

Il Servizio di Calcolo ha dedicato una specifica VLAN al servizio di *Server Load Balancing* a cui è stata attribuita una network IP tra quelle private (indirizzamento 172.17.0.0/16). I server che svolgono servizi critici sono stati configurati con indirizzi IP su tale network privata, mentre gli indirizzi dei servizi virtuali sono stati assegnati sulle network pubbliche, in modo da poter essere indirizzati anche dall'esterno della LAN dei LNF.

In particolare sono gestiti tramite la funzionalità SLB i server web, i server smtp e i server imap descritti nei paragrafi precedenti. Nella Figura 4-6, a titolo di esempio, è illustrato lo schema di funzionamento della funzionalità SLB relativamente al servizio web.

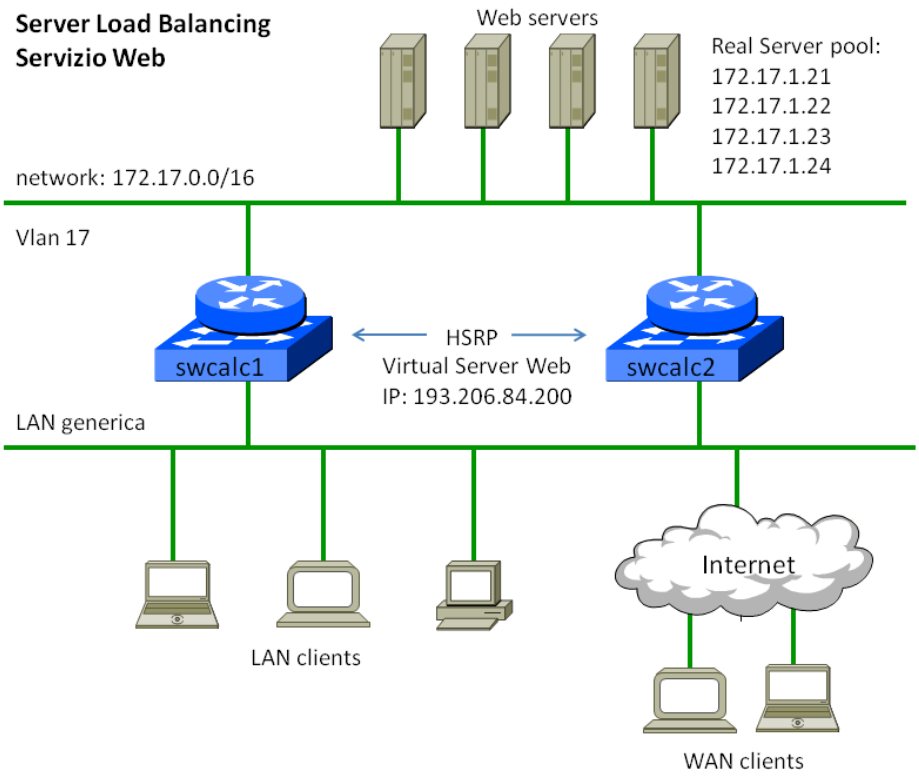



**Figura 4-6: Funzionalità di Server Load Balancing**

# 5. Le griglie computazionali

Il termine GRID (griglia computazionale) è stato coniato ispirandosi alla rete elettrica che in inglese si chiama appunto *electric power grid*. L'energia elettrica necessaria a un Paese, infatti, viene di solito prodotta in diverse centrali e quindi distribuita a tutti gli utenti secondo la richiesta. Di conseguenza, quando giriamo un interruttore, non sappiamo da quale centrale proviene l'energia che fa accendere la nostra lampadina: può essere la centrale più vicina o, se l'energia di questa è già totalmente assorbita da altri utenti, quella di una centrale più lontana. Analogamente, la Grid metterà a disposizione grandi risorse condivise di calcolo e di memoria di archivio alle quali l'utente potrà attingere senza conoscere la provenienza di ciò che sta sfruttando in quel momento. In un futuro, per il mondo della ricerca e per strutture come le aziende, le amministrazioni pubbliche e gli ospedali potrebbe divenire possibile fare a meno di propri sistemi di calcolo per svolgere operazioni complesse: sarebbe sufficiente possedere un semplice PC dal quale accedere a risorse disponibili in centri esterni attraverso i servizi GRID.

È stato l'INFN a proporre il primo sviluppo pionieristico della GRID in Europa nel 2000 con INFN-GRID: il primo progetto nazionale di questo tipo approvato in Europa. Lo sviluppo della GRID ha poi avuto un notevole impulso nel 2001, grazie al progetto del CERN chiamato Datagrid e finalizzato a mettere a disposizione delle comunità scientifiche la prima GRID Europea sperimentale. Nel 2002, grazie anche al fondamentale contributo dell'INFN per il progetto europeo DataTag, è stata fatta la prima dimostrazione di comune funzionamento fra la GRID europea e quella statunitense. È ora il momento del progetto Egee (Enabling Grid for E-sciencE) il quale si propone di realizzare un'infrastruttura comune a livello europeo per permettere a gruppi di ricercatori distribuiti in tutta l'Europa e nel mondo di condividere in modo trasparente risorse distribuite di calcolo, di memoria di archivio e di dati. In sostanza è un progetto di seconda generazione per realizzare non più delle sperimentazioni bensì una vera GRID di produzione, con lo scopo cioè di mettere a disposizione di tutti gli scienziati europei una "e - infrastruttura" che consenta loro di fare scienza superando facilmente le barriere geografiche e nazionali.

Una "e - infrastruttura" è dunque l'insieme di internet e di GRID; in altre parole è l'unione di una rete a banda larga, che collega tutte le reti nazionali, e dei servizi per la condivisione di risorse di calcolo, memoria di archivio e basi di dati disponibili in ciascun paese.

In Italia la rete informatica a larga banda, a cui sono attualmente collegate le Università e gli enti di ricerca italiani, è la rete GARR realizzata e sviluppata dall'INFN a partire dal 1998 per incarico del Ministero dell'Istruzione, dell'Università e della Ricerca Scientifica.

Oggi su questa rete la maggior parte delle informazioni viaggia a una velocità di vari Gb/s, contro i circa 8 Mb/s di bit delle linee commerciali Adsl più veloci. L'opera di costruzione di GRID come quella INFN o di EGEE, sebbene sia concepita per i ricercatori, ha importanti ricadute nel mondo produttivo e in definitiva su tutti i cittadini allargando la base delle conoscenze disponibili per esempio attraverso la formazione di giovani che sono poi in grado di trasferire alla società la conoscenza acquisita.

## 5.1 Sistemi di calcolo distribuito tradizionali

In generale un sistema distribuito è un insieme di calcolatori interconnessi da una rete in cui le comunicazioni avvengono tramite messaggi.

Un esempio di sistema distribuito è Internet, che si estende a livello mondiale comprendendo risorse fisicamente molto distanti tra loro, in cui computer con funzioni diverse e connessi da reti di vario tipo si scambiano messaggi basati su disparati protocolli.

Un sistema di calcolo distribuito è un sistema di molti processori distribuiti su rete locale o geografica, interconnessi tra loro, accessibili agli utenti nel modo più trasparente possibile e capaci di cooperare tra loro alla soluzione di un problema (ad es. un'applicazione dell'utente).

Un sistema di calcolo distribuito, rispetto ai più tradizionali sistemi di calcolo basati su singoli server o su mainframe, presenta alcuni vantaggi:

- Basso costo in rapporto alle prestazioni.
- Potenza integrata scalabile.
- Distribuzione delle risorse di calcolo su più sedi (es. banche, aziende, industrie, istituzioni scientifiche, università).
- Condivisione di dati su più sedi (es. database comuni).
- Affidabilità complessiva dell'intero sistema (riduzione dei single points of failure).

Tuttavia gestire un sistema distribuito introduce indubbiamente anche qualche problema, tra i quali ad esempio:

- Complessità di Gestione ("globale" e "controllata").
- Riservatezza e Sicurezza delle informazioni.
- Disaster Recovery su scala geografica.

Un sistema distribuito deve avere i seguenti requisiti minimi:

- *Trasparenza:* il sistema deve apparire come un sistema singolo.
- *Flessibilità:* ovvero adattabilità a nuove esigenze.
- *Affidabilità (reliability):* caratteristica intrinseca del sistema distribuito: se una macchina ha problemi il job deve poter migrare su altre macchine.
- *Efficienza (performance)*: il sistema nel suo complesso deve essere efficiente. La valorizzazione dell'efficienza complessiva può avvenire secondo i seguenti parametri (*performance metrics*):
  - *Response time*.
  - *Throughput* (numero di job in un'ora, strettamente dipendente dal tipo di job: *cpu bound* o *I/O bound*).
  - *Uso di banda trasmissiva sulla rete*.

  Un altro importante parametro che misura l'efficienza di un sistema distribuito è la *granularità di parallelismo* (*fine-grain* o *coarse-grain*).
- *Scalabilità:* l'espandibilità è un requisito fondamentale di un sistema distribuito.

Con l'evoluzione della tecnologia delle reti di trasmissione dati e la conseguente crescita delle prestazioni dei canali di comunicazioni, oggi si parla spesso di *High Throughput Computing* (HTC), particolarmente idoneo per problemi di calcolo in cui occorre processare una grande quantità di dati indipendenti (nel qual caso si parla di calcolo "in parallelo" e non di calcolo parallelo). Un sistema HTC è caratterizzato da:

- Prestazioni di insieme piuttosto che alte prestazioni del singolo programma.
- Ridondanza piuttosto che totale affidabilità dei singoli componenti.

I sistemi HTC trovano applicazione nella fisica delle alte energie, nella biologia, e, più in generale, nell'utilizzo condiviso di risorse da parte di molti utenti indipendenti.

### 5.1.1 Condor: esempio di calcolo distribuito

*Condor* è un sistema di calcolo distribuito sviluppato presso l'Università del Wisconsin. L'INFN collabora allo sviluppo e alla configurazione di tool su wide area network e all'adattabilità alle proprie esigenze di calcolo.

L'idea della soluzione Condor è quella di poter sfruttare macchine private temporaneamente inattive facendoci girare programmi che hanno bisogno di un elevato tempo di cpu. L'obiettivo è quello di restituire immediatamente la macchina al proprietario nell'istante in cui ne avesse bisogno, e continuare l'elaborazione in corso (senza perdere i calcoli già effettuati) su altre macchine a disposizione del pool, attraverso un sistema di switching automatico.

Come funziona Condor: s'individuano le macchine che aderiscono all'iniziativa formando quello che viene detto un Pool di Condor. Tra le macchine del Pool s'individuano quelle che svolgeranno la funzione di *Central Manager* e di *Checkpoint Server*:

- *Central Manager*: individua nel pool una macchina *idle* che risponda ai requisiti richiesti dal job che deve essere lanciato (ram, clock, sistema operativo, distanza, etc.).
- *Checkpoint Server*: Importa periodicamente dalle macchine su cui girano i job le immagini dei processi. Qualora la macchina su cui gira un job non sia più disponibile, vuoi perché serve al suo proprietario vuoi perché sia andata in crash, il Checkpoint preleverà l'ultima immagine del processo e la sposterà su un'altra macchina indicata dal Central Manager; il job ripartirà dal punto dove si era fermato.

Condor presenta indubbiamente dei grandi vantaggi, il più importante dei quali è l'ottimizzazione dell'uso delle risorse informatiche a disposizione dell'ente. Tuttavia, Condor è indicato solo per elaborazioni che coinvolgono CPU e RAM, dato che le operazioni di I/O non sono ammesse sulle macchine del Pool; per questo Condor è indicato solo per un tipo di calcolo specifico (molto usato dai fisici teorici).

### 5.1.2 Calcolo parallelo

Nei problemi di larga scala, ove sia richiesta grande potenza computazionale da utilizzare in un singolo programma, oppure grande mole di dati complessi e correlati da elaborare, può avvenire che il singolo computer non abbia risorse sufficienti per l'elaborazione in tempi accettabili.

Da questa esigenza si è evoluto il calcolo parallelo in ambito *High-Performance Computing* (HPC), inizialmente utilizzato su supercomputer dotati di più processori identici e memoria condivisa (*Simmetric Multi-Processing*), poi evolutosi verso soluzioni che prevedevano l'adozione di più elaboratori in un'architettura a memoria distribuita, attraverso particolari canali di comunicazione ad altissime prestazioni (ad esempio *infiniband*).

Con l'evoluzione delle reti LAN, vengono sempre più spesso usate farm di workstation. Tale tendenza deriva direttamente dall'economia di utilizzo di tale tipologia di macchine (ad esempio basate su piattaforma x86) e dalla contemporanea crescita delle prestazioni dei canali di comunicazioni messi a disposizioni dalle attuali infrastrutture di reti locali.

Per la comunicazione delle applicazioni tra i vari nodi che costituiscono la farm di calcolo (tipicamente in cluster), il modello dominante usato oggi nell' HPC, rimane il protocollo di comunicazione *Message Passing Interface* (MPI).

Il *Message Passing Interface* è lo standard de facto per la comunicazione tra nodi appartenenti a un cluster di computer che eseguono un programma parallelo sviluppato per sistemi a memoria distribuita. MPI, rispetto alle precedenti librerie utilizzate per il passaggio di parametri tra nodi, ha il vantaggio di essere molto portabile e veloce (MPI è stato implementato per moltissime architetture parallele e viene ottimizzato per ogni architettura). Gli obiettivi di MPI sono: alte prestazioni, scalabilità e portabilità.

L'applicazione deve essere opportunamente costruita e sviluppata per poter essere parallelizzata su più nodi della farm, mediante l'uso della libreria MPI. L'efficienza nella parallelizzazione dipende molto dall'abilità del programmatore.

## 5.2 GRID

I Grid computing o sistemi Grid sono un'infrastruttura di calcolo distribuito, utilizzati per l'elaborazione di grandi quantità di dati, mediante l'uso di una vasta quantità di risorse. In particolare, tali sistemi permettono la condivisione coordinata di risorse all'interno di un'organizzazione virtuale.

Il termine "Griglia" è stato coniato intorno alla metà degli anni Novanta. Il vero e specifico problema alla base del concetto di Griglia è la condivisione coordinata di risorse all'interno di un'organizzazione virtuale dinamica e multi-istituzionale (Virtual Organization, brevemente indicata con VO).

La condivisione non è limitata solo allo scambio dei file, ma si estende all'accesso diretto a computer, storage, software e in generale a tutto l'hardware necessario alla risoluzione di un problema scientifico, ingegneristico o industriale. Gli

individui e le istituzioni, che mettono a disposizione della griglia le loro risorse per la medesima finalità, fanno parte della stessa VO.

Caratteristica comune dei progetti Grid è la necessità di disporre un ambiente di calcolo data-intensive, all'interno del quale le applicazioni possano accedere a grandi quantità di dati geograficamente distribuiti in maniera veloce ed affidabile. Onere della Grid è far operare tali applicazioni nel miglior modo possibile. È facile osservare che nessun computer attualmente in commercio sarebbe in grado, da solo, di elaborare simili moli di dati in tempi ragionevoli; tuttavia la condivisione di risorse quali CPU e dischi, opportunamente coordinati, può dare l'impressione all'utente di accedere ad un supercomputer virtuale, con una incredibile potenza computazionale e capacità di memorizzazione in grado di sopportare grandi carichi di lavoro.

Dall'idea di far apparire tutta l'architettura di un Grid come un unico supercomputer virtuale, celando all'utilizzatore tutta la complessità interna e mostrandogli solo i benefici, nasce l'esigenza di progettare e realizzare uno schedulatore di risorse: il Resource Broker. Esso è uno dei componenti critici del sistema di gestione delle risorse, ha il compito di assegnare le risorse ai job, in modo da soddisfare le esigenze delle applicazioni e del sistema. Le risorse di cui esso deve tenere traccia e gestire, includono sistemi di calcolo e sistemi di immagazzinamento dati.

Lo scheduling è un campo tradizionale dell'informatica, ma nonostante siano state studiate molte tecniche per numerose tipologie di sistemi (da uniprocessore a multiprocessore ai sistemi distribuiti), le caratteristiche tipiche delle griglie di dati rendono molti di questi approcci inadeguati. Infatti, mentre nei sistemi tradizionali le risorse e i job sono sotto il diretto controllo dello schedulatore, le risorse delle griglie sono geograficamente distribuite. Queste ultime sono di natura eterogenea e appartengono a diversi individui o organizzazioni, ciascuna con le proprie politiche di scheduling, modelli di costo di accesso differenti, carichi di lavoro e disponibilità di risorse che varia dinamicamente nel tempo. La mancanza di un controllo centralizzato, insieme alla presenza di utenti che generano job, molto diversi l'uno dall'altro, rendono la schedulazione più complicata rispetto a quella dei sistemi di calcolo tradizionali.

È quindi evidente che i punti chiave su cui si basa il Grid computing sono:

- Capacità di negoziare, secondo regole stabilite, la condivisione di risorse (computers, software, dati, etc.) da parte di organizzazioni o istituzioni (scientifiche, industriali, governative, etc.) che agiscono da Virtual Organizations.
- Importanza di definire protocolli standard per consentire l'interoperabilità e realizzare una infrastruttura comune.

Si può quindi definire il sistema Grid individuando i tre punti fondamentali sottoelencati:

1. **Coordinare risorse che non devono essere soggette ad alcun controllo centralizzato**
   (es. nodi di calcolo e database di istituzioni sparse su territorio nazionale e nel mondo, senza la necessità del controllo tipico di un sistema a gestione locale, pur garantendo la sicurezza e la realizzazione delle politiche di utilizzo all'interno di un'organizzazione virtuale).

2. **Usare protocolli e interfacce standard, open, general purpose** (essenziali per assicurare in modo trasparente funzionalità di base quali autenticazione, autorizzazione, ricerca e accesso alle risorse).
3. **Assicurare un'elevata qualità di servizio (QoS - *Quality of Service*)** (es. tempi di risposta, throughput, disponibilità, sicurezza, coallocazione di risorse).

Infine, per completezza di informazione, nella Tabella 5-1 sono riportate le principali differenze tra i sistemi di calcolo distribuito convenzionale e i sistemi di calcolo Grid.

| Sistemi di calcolo distribuito convenzionali | Sistemi di calcolo Grid |
|---|---|
| pool virtuale di nodi di calcolo | Pool virtuale di risorse |
| L'utente ha accesso ai nodi del pool | L'utente ha accesso al pool ma non ai nodi |
| l'accesso al nodo implica l'utilizzo di tutte le risorse del nodo | l'accesso ad una risorsa può essere ristretto |
| l'utente è a conoscenza delle caratteristiche dei nodi | l'utente non ha idea delle caratteristiche delle risorse |
| i nodi appartengono solitamente ad un singolo dominio di gestione | le risorse appartengono a più domini di gestione |
| elementi nel pool: 100-1000, statici | elementi nel pool: >> 1000, dinamici |

**Tabella 5-1: Sistemi di calcolo distribuito convenzionale e GRID**

### 5.2.1 L'architettura GRID

Il problema da affrontare nella realizzazione di un'architettura Grid, come prima definita, è la condivisione coordinata di risorse su larga scala in un contesto di organizzazione virtuale, multi-instituzionale e dinamica.

Tale architettura deve poter identificare le componenti principali del sistema, specificare scopo e funzioni di queste componenti, indicare come queste componenti interagiscono fra di loro e definire servizi e protocolli comuni per garantire l'interoperabilità attraverso la rete (flessibilità di aggiungere nuovi utenti, servizi e piattaforme hw/sw in modo dinamico) per costituire così un sistema aperto.

Per il raggiungimento di questi obiettivi il modello di riferimento è l'architettura a clessidra, rappresentata con un diagramma a blocchi interconnessi nella Figura 5-1. Il centro (neck) della clessidra definisce un piccolo insieme di astrazioni di base (core) e di protocolli (servizi di base); la parte superiore contiene high level services (o behaviors) che si basano sui servizi e protocolli sottostanti; la parte inferiore contiene le risorse della grid.

Si osservi la figura sottostante per una rappresentazione dell'architettura con un diagramma a bocchi interconnessi.

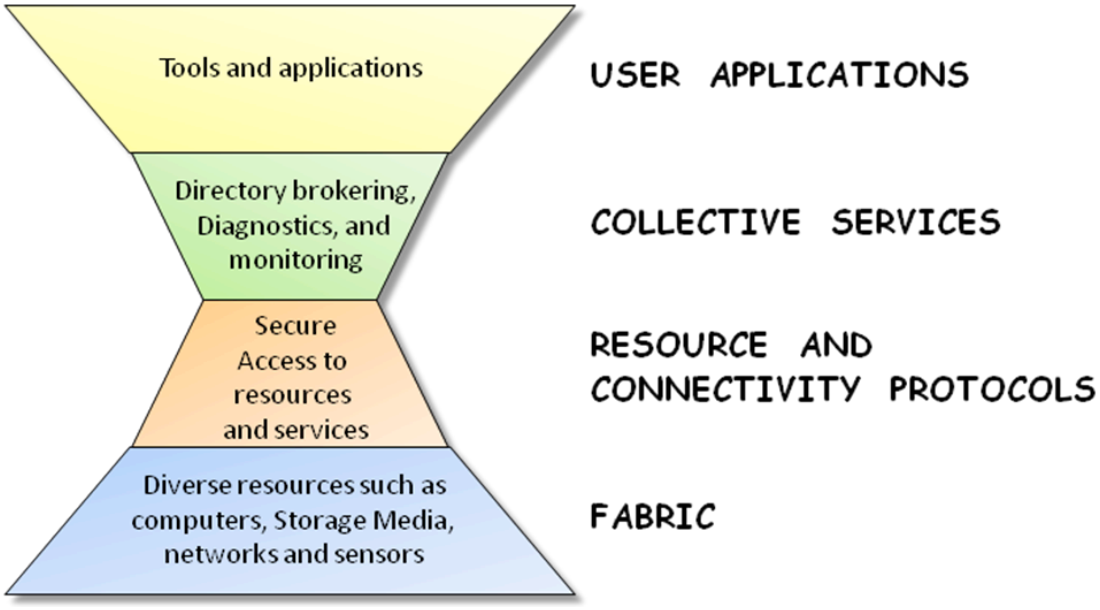



**Figura 5-1: Architettura a clessidra**

Segue la descrizione dettagliata dei vari blocchi che costituiscono l'architettura di Grid illustrata in Figura 5-1.

***Fabric Layer***

Il Fabric Layer fornisce le risorse per l'accesso condiviso da parte della Grid, quali, ad esempio:

- Risorse computazionali
- Sistemi di storage
- Cataloghi
- Risorse di rete
- Sensori

Le risorse devono assicurare un meccanismo di enquiry che consenta di scoprire la loro struttura, lo stato e la loro capacità, oltre ad un meccanismo di management per il controllo della qualità del servizio offerto.

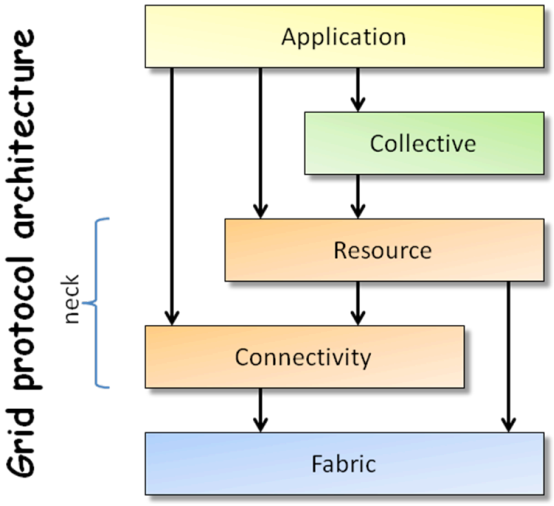



**Figura 5-2: Architettura dei protocolli Grid**

***Connectivity Layer***

Il Connectivity Layer definisce i protocolli base per la comunicazione e l'autenticazione. I protocolli di comunicazione abilitano lo scambio dei dati fra le risorse del fabric layer e si basano sui protocolli dell'architettura Internet (IP, TCP, UDP, DNS, RSVP). I protocolli di autenticazione, costruiti sui servizi di comunicazione, forniscono meccanismi crittografici sicuri per verificare l'identità di utenti e risorse e devono avere le seguenti caratteristiche:

- *Single sign on*: l'utente si deve autenticare una volta sola.
- *Delegation*: propagazione delle credenziali ai programmi.

- *Integration with local security*: non sostituiscono ma si mappano nell'environment locale.
- *User-based trust relationships*: il sistema di security non deve richiedere che i sistemi di security locali interagiscano tra di loro per configurare l'ambiente di sicurezza.

Il *Globus Toolkit*, il primo software open source utilizzato per la costruzione delle griglie computazionali, usa i protocolli GSI (*Grid Security Infrastructure*) per l'autenticazione (certificati con formato X.509), la protezione della comunicazione (estende i protocolli di TLS - Transport Layer Security) e l'autorizzazione.

***Resource Layer***

Il Resource Layer definisce i protocolli per negoziare, iniziare, monitorare, controllare e addebitare l'utilizzo di risorse singole, cioè non distribuite. Questi protocolli utilizzano le funzioni del Fabric Layer per accedere e controllare risorse locali. Due classi principali di protocolli dello strato Resource sono gli *Information protocol*s per ottenere informazioni sulla configurazione e lo stato di una risorsa e i *Management protocols* per negoziare l'accesso alla risorsa condivisa.

A ciò si aggiungono i servizi che costituiscono l'Information Service: GIIS (Grid Index Information Service); GIS (Grid Information Service); GRIS (Grid Resource Information Service); GSS (Generic Security Service). L'*Information Service* ha un ruolo fondamentale nella grid perché sta alla base del *resource discovery* e del *decision making*.

***Collective Layer***

Il Collective Layer definisce protocolli e servizi (oltre ad API e SDK) che non sono associati a una singola risorsa ma a una collezione di risorse. Essi si basano sui protocolli definiti nel Resource e nel Connectivity layer, quindi possono implementare una vasta gamma di servizi senza porre nuovi requisiti sulle risorse condivise.

Esempi di servizi:

- *Directory services*: consentono ai membri di una VO di identificare le risorse a disposizione della VO.
- *Co-allocation, scheduling and brokering services*: consentono ai membri di una VO di richiedere e schedulare l'allocazione di una o più risorse.
- *Monitoring and diagnostics services*: consentono il monitoraggio delle risorse di una VO, inclusi gli attacchi (intrusion detection).
- *Data replication services*: consentono la gestione ottimizzata delle risorse di storage di una VO per massimizzarne le prestazioni (tempi di risposta, affidabilità, etc.).
- *Grid-enabled programming systems*: consentono l'utilizzo di modelli di programmazione utili in ambienti Grid per l'implementazione dei vari servizi Grid (es. Message Passing Interface).
- *Workload management systems and collaboration frameworks*: chiamati anche PSE (Problem Solving Environments) per l'utilizzo e la gestione dei carichi di lavoro in ambienti collaborativi.

- *Software discovery services*: consentono di scegliere software e piattaforme adatte per il problema specifico da risolvere.
- *Community authorization servers*: consentono di gestire le politiche di accesso alle risorse di una comunità in modo da renderle fruibili all'utente.
- *Community accounting and payment services*: consentono di raccogliere informazioni sull'utilizzo delle risorse ai fini di resoconti, pagamenti, limitazioni di uso.
- *Collaboratory services*: consentono lo scambio di informazioni all'interno di vaste comunità di utenti.

Il *Globus Toolkit* utilizza i servizi di cui sopra e altri, fra cui MDS (*Meta Directory Service*) per la gestione informativa delle risorse.

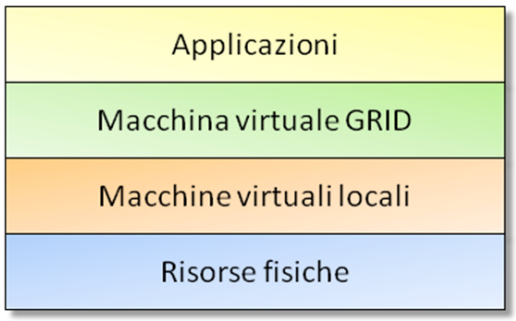
***Applications Layer***

L'Application Layer è l'insieme delle applicazioni degli utenti appartenenti ad una Virtual Organization. Le applicazioni sono costruite sulla base dei servizi definiti in ciascuno strato.

### 5.2.2 Il middleware di GRID

Il Grid middleware di base è la componente software che realizza una macchina virtuale di griglia; esso va inteso come strato di mezzo tra i sistemi operativi delle singole risorse e le applicazioni distribuite.

Nella figura il GRID middleware è rappresentato dal livello "Macchina virtuale GRID" posizionato tra le "Applicazioni" e le "Macchine virtuali locali".




Storicamente il GRID middleware viene inizialmente realizzato con il Globus Project a partire dal 1998, che diviene lo standard "de facto" per le tecnologie GRID.

Il Globus project nasce dallo sviluppo open source del Globus Toolkit, e realizza un insieme di librerie e servizi "nucleo" dell'infrastruttura Grid" tramite un'architettura a livelli in un modello modulare (*bag of services*).

Alla fine degli anni 90 il GRID middleware evolve verso il progetto LCG (*LHC Computing GRID*), sviluppato (open source) dal CERN e dalle comunità di fisica delle alte energie a partire dal Globus Toolkit 2.

Dal 2004 evolve ulteriormente tramite il progetto gLite (*Grid Lite*). Sviluppato nell'ambito del progetto europeo EGEE, gLite rappresenta un middleware di nuova generazione orientato ai servizi.

#### 5.2.2.1 Il middleware gLite

Grid Lite è il middleware Europeo completamente riscritto a partire dal modello del Globus toolkit 2.4. Realizzato con l'obiettivo di fornire una piattaforma multidisciplinare, sulla base dei feedback di alcune comunità pilota, gLite nasce per fornire servizi di basso livello sui quali costruire applicazioni scientifiche e commerciali.

gLite prevede un modello fortemente strutturato per l'implementazione dei livelli *Collective*, *Resource* e *Connectivity*, descritti nel paragrafo 5.2.1, interagendo con prodotti di terze parti per i livelli *Application* e *Fabric*; quindi offre protocolli di base per l'accesso alle risorse computazionali, di storage e per la creazione di virtual instrument.

In particolare, il modello gLite si basa sui concetti di *Elemento* e *Sito* ed implementa due classi di servizi: i servizi *core* e *collective*. I servizi di tipo *core* sono servizi locali che permettono di condividere risorse di calcolo, di storage e virtual instruments, sono implementati da GRID Element e sono presenti in tutti i siti. I servizi *collective* sono servizi distribuiti o centralizzati che lavorano al di sopra delle risorse locali e che di fatto costituiscono l'infrastruttura distribuita, sono implementati da Grid Element e hanno valenza collettiva.

***Concetto di Elemento***

Un Elemento (*GRID Element*) in gLite è un host che fornisce uno o più servizi (di tipo collective o core) ed i metodi per accedervi. Un GRID Element è in grado di pubblicare informazioni sulle caratteristiche e sullo stato dei servizi ed è in grado di interagire con altri elementi GRID o con gli utenti direttamente.

Esempi di GRID Element sono:

- *Computing Element* (CE): fornisce tutti i servizi necessari per l'accesso alle risorse computazionali, interagisce con i local resource manager, pubblica le informazioni sui suoi servizi.
- *Storage Element* (SE): fornisce i servizi e le interfacce per la gestione dello storage locale in ambiente GRID in maniera trasparente all'utente.
- *Worker Node* (WN): fornisce i servizi per il calcolo, ovvero le macchine utilizzate per la computazione pura.
- *Site BDII* (SBDII): Pubblica le informazioni sulle risorse locali.

***Concetto di Sito***

Un *Sito* in gLite è un insieme di risorse che possiede almeno un site BDII, un CE, un SE ed una serie di WN. Nell'architettura gLite i *Siti* rappresentano i nodi della GRID, centri di risorse e di servizi locali (rif. Figura 5-3 e Figura 5-4).

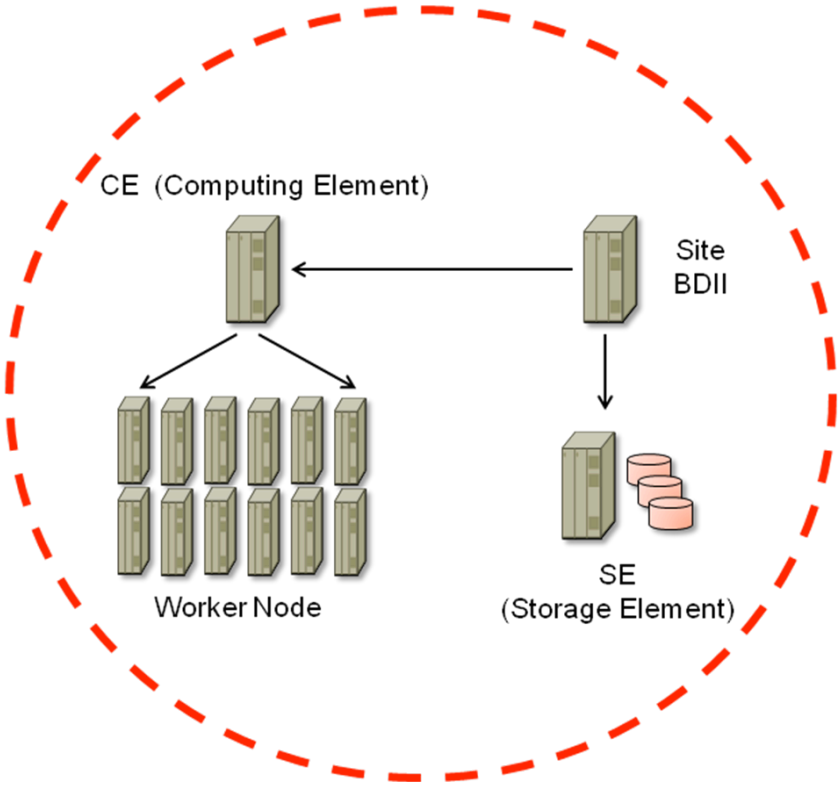



**Figura 5-3: Sito gLite**

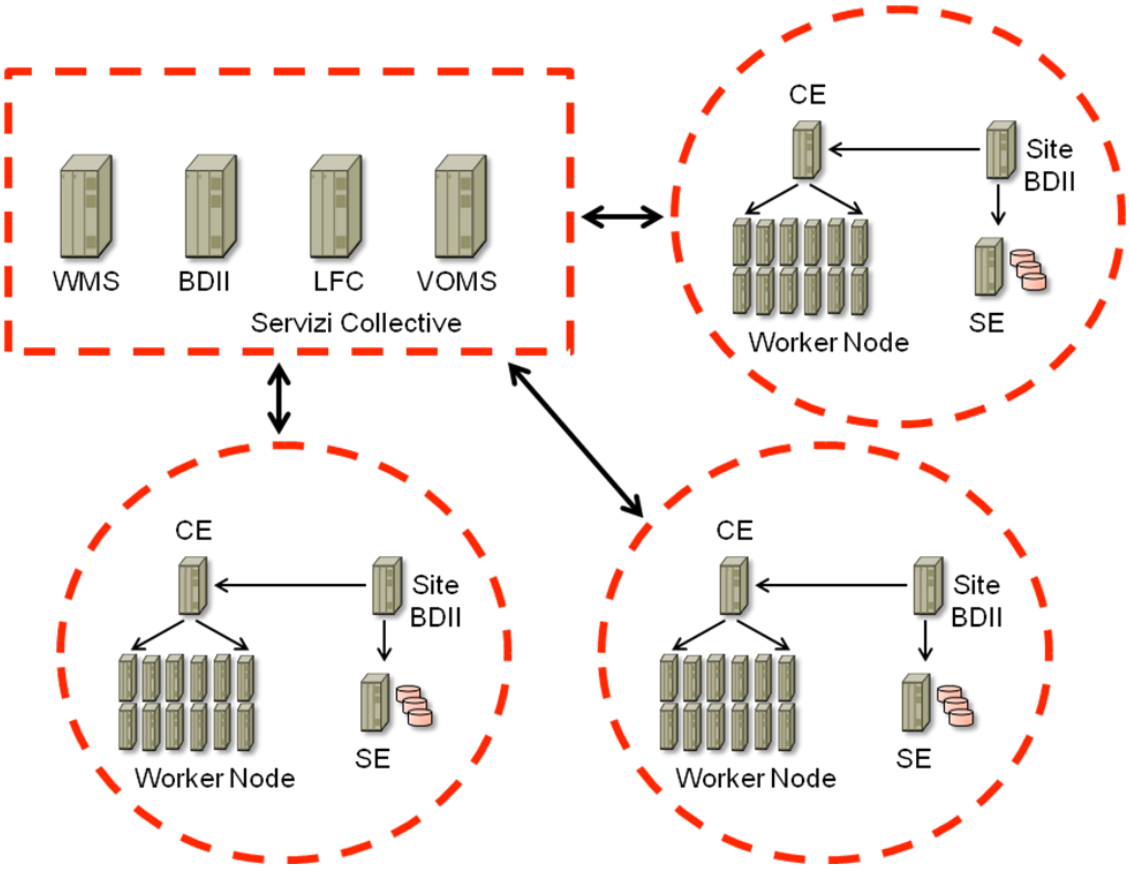



**Figura 5-4: Il middleware gLite (schema di una VO)**

***I Servizi Collective***

I servizi collective di gLite sono:

- *Virtual Organization Membership Server* (VOMS): è un servizio per la gestione delle membership ad una VO e per la gestione delle policy di utilizzo delle risorse.
- *Workload Management System* (WMS): ha una visione globale della GRID e schedula l'esecuzione dei job sui singoli SITI.
- *Logging and Bookkeeping Server* (LB): tiene traccia dei job e del loro stato.
- *Top BDII* (TBDII): Pubblica le informazioni su tutte le risorse della GRID e viene usato dal WMS.
- *Logical File Catalog* (LFC): fornisce i servizi di naming ed un filesystem virtuale di alto livello per il mapping di nomi logici a file fisici distribuiti sugli Storage Element.

Essendo così strutturato, il middleware gLite prevede un'architettura di tipo SOA (Service Oriented Architecture) che garantisce una maggiore modularità, un alto livello di astrazione e consente il riutilizzo del codice e delle applicazioni di terze parti.

I servizi comunicano tra di loro attraverso interfacce e protocolli ben definiti che possono comprendere coordinamento di servizi e scambio di dati. Molti servizi oggi sono basati su tecnologie web-services.

Il *Middleware* è quindi il vasto insieme di librerie e servizi, per la massima parte sviluppate in ambiente open source, che permettono all'utente finale della Grid di utilizzare la "macchina virtuale" per sottomettere le proprie applicazioni ed accedere ai propri dati sulla Grid.

Nei paragrafi che seguono verranno trattati in dettaglio i servizi di base del Middleware: servizio di sicurezza (GSI: Autorizzazione, Autenticazione), servizi informativi (*Grid Information Service*), servizio di gestione delle risorse (*Workload Management*), servizio di gestione dei dati (*Grid Data Management*) e servizio di monitoring.

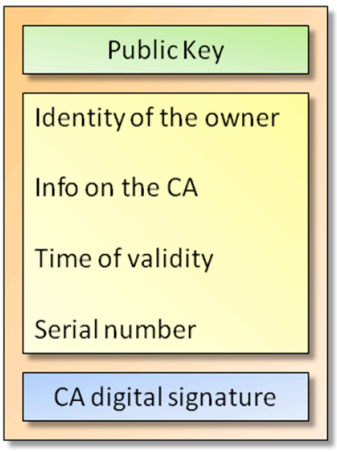
#### 5.2.2.2 La sicurezza: Grid Security Infrastructure (GSI)

Le politiche di accesso alle risorse di Grid sono gestite attraverso un controllo locale e flessibile. Le credenziali di una entità (utente o server) sono fornite attraverso certificati digitali X.509 rilasciati da parte di *Certification Authority* (CA).

I certificati digitali X.509 (rif. figura seguente) sono una soluzione basata su crittografia asimmetrica a chiave pubblica e privata.

La crittografia asimmetrica è caratterizzata dal fatto che se il mittente cifra un messaggio con la chiave pubblica del ricevente, quest'ultimo potrà decifrare il messaggio soltanto utilizzando la sua chiave privata. Più in generale quello che è criptato con la chiave pubblica può essere decriptato solo con la chiave privata e vice-

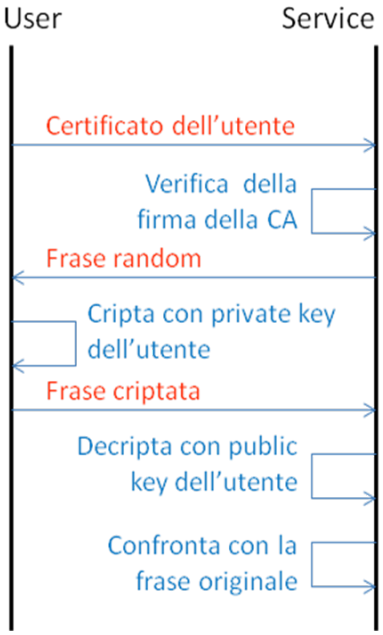
versa. Non si può mai ricavare la chiave pubblica da quella privata, né quella privata da quella pubblica.




Un certificato X.509 può essere utilizzato sia per firmare un documento o un messaggio, sia per l'autenticazione. Nel primo caso, da un messaggio di lunghezza variabile, una particolare funzione produce una stringa di lunghezza fissa (*hash*). L'*hash* viene criptato con la chiave privata del mittente e decriptato con quella pubblica, verificando l'integrità. In questo modo si ottiene sia il controllo sull'identità del mittente, sia la verifica dell'integrità del messaggio.

Nel secondo caso, le fasi dell'autenticazione sono le seguenti:

- ogni richiesta di autenticazione è accompagnata dal certificato dell'utente;
- la CA deve essere accreditata (*trusted*) presso il servizio ricevente, affiché questo possa verificare e validare il certificato dell'utente firmato dalla CA.
- Il servizio ricevente invia all'utente una frase random.
- L'utente cripta tale frase random con la chiave privata e la reinvia al ricevente.
- La stringa viene decriptata dal ricevente con la frase pubblica dell'utente.
- La stringa viene confrontata con quella inviata. Se coincide, il servizio ricevente non può rifiutare la richiesta di autenticazione.




***Autenticazione***

Per evitare che un certificato possa essere intercettato, viene creato un certificato proxy delle credenziali (locale e temporaneo) utilizzato per tutte le operazioni. Il certificato proxy è un'estensione di X.509, ha una durata molto limitata e svolge la funzione di delegato per l'autenticazione, permettendo ad un processo remoto di autenticarsi per conto dell'utente. L'utente firma il proxy con la propria chiave privata, immettendone la password di protezione.

I certificati proxy sono gli unici a viaggiare nella grid ed hanno coppie di chiavi pubblica e privata con chiave privata non criptata, per cui non è richiesto l'intervento umano. Un'applicazione può durare più a lungo del certificato proxy, poiché un *server myproxy* può rinnovare automaticamente i certificati proxy di applicazioni ancora in corso, che stanno per scadere.

***Autorizzazione***

Essere autenticato non dà diritti di esecuzione su una Grid, poiché l'utente deve appartenere ad un gruppo autorizzato all'uso delle risorse (ovvero ad una Virtual Organization).

Il certificato proxy X.509 non contiene informazioni sulla VO, che viene invece individuata attraverso il *grid-mapfile* presente sulle risorse. Il grid-mapfile stabilisce i diritti di un utente su una specifica risorsa, in base alla sua VO di appartenenza e le entry del grid-mapfile mappano gli utenti della Grid autorizzati su utenti del sistema locale. Quindi, per far sì che un utente possa utilizzare effettivamente le risorse distribuite di calcolo e di storage, occorre che appartenga ad una Virtual Organization (VO). Ogni utente può appartenere a più VO, presso le quali si identifica con il suo certificato personale; inoltre, in ogni VO un utente può avere differenti ruoli e differenti privilegi, poiché il servizio VOMS (*Virtual Organization Membership Service*) estende le informazioni del certificato proxy X.509 con VO membership, gruppo, ruolo e privilegi.

#### 5.2.2.3 Il modello informativo

Il modello Informativo (*GIS: Grid Information System*) è il sistema che si occupa di raccogliere le informazioni sullo stato delle risorse della griglia. Il GIS individua le risorse disponibili e il loro stato di salute, per ottenere importanti informazioni utili su cui basare scelte per la sottomissione dei job.

Il GIS si occupa inoltre di monitorare lo stato delle risorse localmente e di pubblicare le opportune informazioni adottando un modello riconosciuto da tutti i componenti della griglia. I requisiti del sistema informativo di grid sono:

- *espressività*: il modello dei dati capace di rappresentare le strutture rilevanti di un sistema distribuito;
- *estendibilità*: capacità di incorporare informazioni aggiuntive, non incluse nel modello dei dati di partenza;
- *dati dinamici*: i cambiamenti sui dati di ogni risorsa devono essere resi prontamente disponibili;
- *flessibilità di accesso*: capacità di lettura, aggiornamento e ricerca sui dati;
- *sicurezza*: restrizioni di accesso e/o aggiornamento delle informazioni;
- *performance*: rapidità di accesso alle informazioni usate frequentemente;
- *scalabilità*: qualità di servizio costante all'aumentare dei componenti;
- *costo*: i costi di gestione delle informazioni in termini di risorse usate devono essere contenuti;
- *uniformità*: standard nella rappresentazione delle risorse e uniformità delle API.

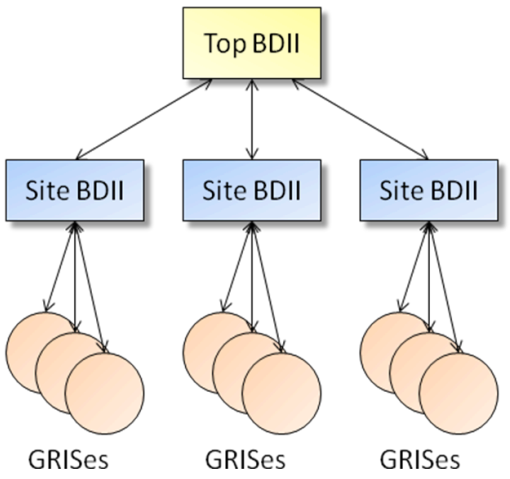
L' architettura del modello informativo di grid è a livelli. Al livello più basso c'è il *Grid Resource Information Server* (GRIS) che gestisce l'informazione sullo stato di una data risorsa (CE, SE). Quindi si ha un GRIS per ogni risorsa, ovvero un insieme di script e sensori che estraggono dati utili sulla risorsa.

Al livello medio c'è il *Site BDII* (Berkeley Database Information Index) che gestisce l'informazione a livello di 'sito'. Ogni sito deve avere un Site BDII, che colleziona e mette in cache le informazioni ottenute interrogando periodicamente i sistemi GRIS sottostanti.

Al livello Alto c'è il *Top BDII* che ottiene informazioni da tutte le risorse registrate. Esiste un BDII centralizzato per ogni VO, o multi-VO su base regionale.

Il modello informativo di Grid (illustrato nella figura a destra) è gerarchico con un sistema di raccolta delle informazioni basato su LDAP. L'informazione è mantenuta nelle foglie di un albero e viaggia fino alla sua radice. Le foglie dell'albero sono GRIS e i nodi intermedi sono Site BDII.

I client possono interrogare l'albero ad ogni livello; più alto è il livello cui si fa l'interrogazione e meno recente sarà l'informazione ottenuta.




Le risorse disponibili in Grid devono essere descritte in maniera precisa e sistematica al fine di fornire le seguenti funzionalità: discovery, allocazione (brokering), controllo delle prestazioni, troubleshooting, e monitoring.

Il modello informativo di Grid, usato come base del Grid Information Service, è il GLUE (*Grid Laboratory for a Uniform Environment*); è stato sviluppato per gran parte dall'INFN e adottato anche da Globus e da gLite, allo scopo di realizzare una collaborazione tra progetti di fisica delle alte energie europei ed americani, per la definizione di un modello informativo comune tra domini Grid diversi, indipendente dall'implementazione.

Lo schema astratto GLUE viene mappato nel data model LDAP per l'implementazione. Le informazioni relative alla Grid sono fornite da numerosi server LDAP (GRIS, BDII) ed organizzate gerarchicamente in un Directory Information Tree (DIT), in cui ogni nodo del DIT è una entry LDAP. Il servizio LDAP è utilizzato anche come linguaggio di query e come protocollo.

Nello schema di GLUE le entità da descrivere si classificano come siti e servizi. I *siti (system)* sono insiemi di risorse connesse che operano come un'entità funzionale; i *Servizi*, forniti dai siti, svolgono una funzione coerente ad un richiedente.

GLUE è progettato per fornire informazioni su *computing resource*, *storage resource* e *network resource*.

Il *computing resource* è caratterizzato dalla potenza di calcolo offerta e dalla gestione delle richieste attraverso un sistema di code (*queue*); le policy presenti su ciascuna coda permettono di differenziare il servizio offerto (es. durata massima di un job, numero massimo di job running, tipologia di CPU allocate).

Per cui i parametri che è necessario pubblicare nell'Information System per l'utilizzo di risorse di computing sono: l'*execution environment* (sistema operativo, librerie disponibili, etc.), il *quality of service* (tempo di risposta stimato, etc.), lo *status* (numero di job in esecuzione, etc.), la *policy* (tempo massimo di esecuzione, CPU assegnate, etc.), l'*access right* e il *location* (Uniform Resource Locator, etc.).

Lo *storage resource* offre spazio di memorizzazione tramite lo Storage Element (SE): un sistema che fornisce *Storage Service* e *Storage Space*; l'SE può essere costituito da un semplice disk server oppure da un sistema di storage più complesso.

Lo *Storage Service* gestisce le risorse disco e nastro in termini di Storage Space, mascherando i dettagli hardware. I file, gestiti secondo le policy del sistema, vengono trasferiti da/verso i vari Storage Space utilizzando un insieme di data access protocol (es. GridFTP, rfio, etc.).

Lo *Storage Space* è associato ad una directory e assegnato ad una Virtual Organization, è soggetto a *policy* (dimensione massima di un file, numero massimo di file, etc.) e a regole di accesso (*access control base rules*) e fornisce informazioni sul proprio stato (spazio disponibile, spazio utilizzato). Quindi i parametri necessari per l'utilizzo delle risorse di storage sono: il *protocollo di accesso* (gridftp, rfio, etc.), il *quality of service* (affidabilità, disponibilità, etc.), lo *status* (spazio disponibile, etc.), la *policy* (dimensione massima di un file, file lifetime, etc.), l'*access right* e il *location* (Uniform Resource Locator, etc.).

#### 5.2.2.4 Gestione delle risorse

La gestione delle risorse di Grid è effettuata attraverso il *Workload Management System* (WMS) che è un insieme di componenti del middleware il cui scopo è distribuire i job nella griglia.

In particolare il ruolo del *Workload Manager* (*WM*) è soddisfare le richieste di gestione dei job che vengono dai suoi client. Il Workload Manager, in effetti, si occupa di inoltrare un job ad un Computing Element che soddisfi requisiti stabiliti nella descrizione del job. La selezione del CE è il risultato di un processo detto di *Matchmaking*. Le componenti del WMS sono:

- La *User Interface* (UI) è il punto di accesso a Grid per gli utenti.
- Il *Workload Manager* (WM) è la componente centrale, che parla con tutti gli altri componenti per gestire i job.
- Il *Resource Broker* (RB) è responsabile del processo di *matchmaking*, cioè di trovare le risorse adatte ai requisiti del job. In precedenti versioni del middleware era un server a parte, in gLite è solo un componente del software.
- L' *Information Super Market* (ISM) è una cache delle informazioni raccolte dai BDII, necessarie per il matchmaking. Esso si aggiorna dinamicamente effettuando un polling sulle risorse e ricevendo notifiche da servizi.

- Il *Logging & Bookkeeping* (LB) tiene traccia di tutti gli eventi relativi ai job gestiti dal WM e fornisce agli utenti le informazioni sullo stato dei loro job.
- La *Task Queue* permette di accodare i job nel WMS nel caso non ci siano risorse disponibili al momento della sottomissione.
- Il *WMProxy* è il componente, sviluppato come Web Service, attraverso il quale la UI passa le richieste al WM. Esso riceve una delega dall'utente, in modo che il WMS possa interagire con gli altri servizi (gridftp, cataloghi dati) con le credenziali dell'utente che ha sottomesso il job.
- Il *MyProxy* è un servizio che consente di rinnovare il certificato proxy dell'utente se questo è in scadenza prima del completamento dei suoi job. Anche il MyProxy server riceve la delega dall'utente e fornisce al WMS il proxy rinnovato quando questo lo richiede.

***Job Description Language***

Un file di testo, con estensione *.jdl*, contiene tutte le informazioni da specificare per l'esecuzione di un job della Grid. Queste informazioni sono utilizzate dal WMS per allocare la risorsa migliore e per mandare il job in esecuzione. Alcuni esempi di informazioni specificate nel file di descrizione del job sono: i requisiti "fisici" del job (spazio disco, memoria, processori, durata presunta, etc.), i requisiti "logici" (sistema operativo, librerie, s/w, etc.), i dati di input e di output.

Più in dettaglio, le categorie di attributi sono:

- *job attribute*: definiscono le caratteristiche del job;
- *computing resource attribute*: specificano i requirement in termini di risorse di computing;
- *storage resource attribute*: specificano i requisiti in termini di risorse di storage (protocollo di accesso, logical file name, etc.).

***Resource Broker***

Il *Resource Broker* è il componente che effettua il *matchmaking*: il suo compito è quello di individuare la risorsa di computing "migliore" su cui sottomettere il job dell'utente. Il *Resource Broker* interagisce con l'*Information Super Market*, che ha in cache le informazioni raccolte dall'Information System; quindi, per l'esecuzione del job, seleziona il Computing Element che soddisfi i requirement specificati nel file di descrizione. Qualora più Computing Element soddisfassero i requirement, allora viene scelto il Computing Element con *rank* più alto che viene determinato secondo un algoritmo di default, oppure in base a parametri indicati dall'utente nel file jdl.

Alcuni esempi di scenari di sottomissione dei job sono i seguenti:

1. sottomissione diretta dei job: l'utente specifica il Computing Element su cui deve essere sottomesso il job e il WMS non effettua il matchmaking;
2. sottomissione di job senza richieste sullo storage: il Resource Broker effettua il matchmaking interroga l'ISM per ottenere l'elenco delle risorse che soddisfano i requirement e le credenziali dell'utente e, se più Computing Element soddisfano i requirement, allora viene scelto quello con rank più alto;

3. job submission con indicazione di risorse di storage: il Resource Broker effettua il matchmaking, interroga l'ISM per ottenere l'elenco delle risorse che soddisfano i requirement e le credenziali dell'utente e seleziona i Computing Element "vicini" agli Storage Element. se il risultato della selezione consiste in più Computing Element, allora viene scelto il Computing Element con rank più alto.

Nella Figura 5-5 è rappresentato il processo di sottomissione di un job batch nella Grid e la funzione svolta dal Resource Broker.

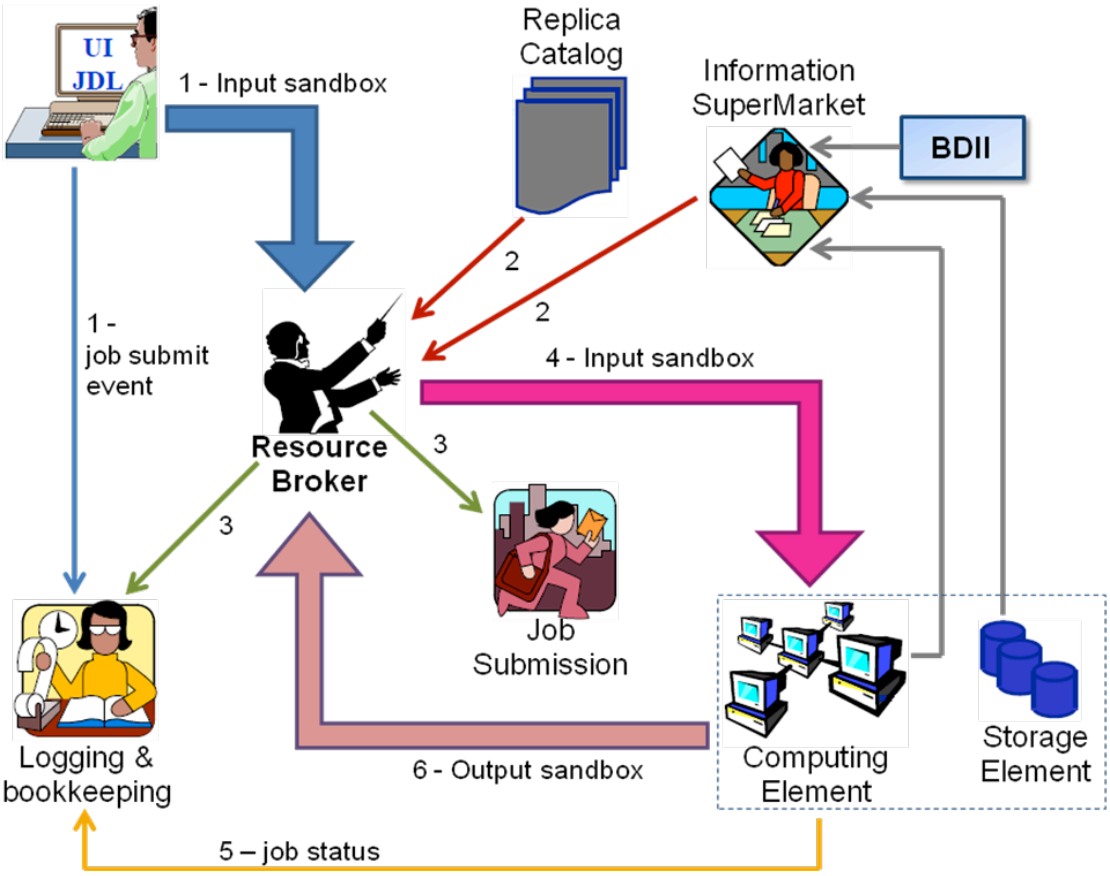



**Figura 5-5: Job submission**

### 5.2.2.5 Gestione dei dati

Il sistema di gestione delle risorse di storage nella Grid deve avere le caratteristiche di *eterogeneità*, *distribuzione geografica* e *descrizione*. Infatti, i dati sono immagazzinati in sistemi che usano differenti tecnologie d'accesso, sono registrati il locazioni distribuite geograficamente (in molti casi non esiste un file system condiviso né un namespace comune) e devono poter essere trasferiti in modo efficiente tra le diverse locazioni. Infine, i dati sono memorizzati in forma di file, quindi occorre un modo per identificare tali file e localizzarli in base al loro contenuto.

Dunque, il sistema di *Data Management* di Grid deve avere le seguenti fondamentali funzionalità:

- localizzazione dei dati
- accesso ai dati locali e remoti
- copia e trasferimento dei dati
- catalogazione e replica dei dati
- gestione dei meta-dati (particolari "dati che descrivono i dati").

Tutti i dati sono memorizzati in una *risorsa di storage*, una combinazione di un hardware e di un software che lo controlla. Poiché una risorsa di storage può variare da un singolo disco ad una libreria di nastri, risorse differenti offrono differenti livelli di *Quality of Service* (QoS).

Lo *Storage Element* è il servizio che consente ad utenti o applicazioni GRID di utilizzare le risorse di storage attraverso tre tipi di interfacce: l'accesso ai dati (I/O), il trasferimento dati (esempio GridFTP) e lo Storage Management (SRM).

Lo *Storage Resource Manager* (SRM) è l'interfaccia che si occupa dell'interazione tra la Grid e le risorse di storage. La definizione di uno standard SRM permette l'interoperabilità tra le molteplici implementazioni poiché il servizio SRM non effettua direttamente il trasferimento dei file tra SE, ma invoca i componenti del middleware che se ne occupano (come GridFTP).

I dati gestiti da un servizio SRM possono essere registrati in disk pool server e/o Mass Storage System (nastri+dischi) e la caratteristica di un SRM è di fornire un file system virtuale che nasconde la complessità del pool di risorse gestite dall'interfaccia. I file possono essere spostati in modo trasparente tra diverse risorse di storage gestite dallo stesso SRM e le risorse sono espandibili dinamicamente.

Inoltre, i file non sono più entità permanenti sullo storage, poiché, mediante la definizione di una *lifetime* specificata, possono avere durata limitata. L'utente può effettuare il *pinning* di un file, impedendone la cancellazione finché ne ha bisogno. L'autorizzazione di accesso ai pool e ai singoli file è gestita mediante le credenziali VOMS ed è implementata una space reservation che consente di preallocare lo spazio di storage alle Virtual Organization.

I tipi di SRM presenti in gLite sono: *DPM* (*Disk Pool Manager*), *Castor (CERN Advanced STORage manager), dCache* e *StoRM (Storage Resource Manager)*. *DPM* consiste di un server che fornisce un singolo punto di accesso verso un pool di server di disco. *Castor* e *dCache* sono costituiti da un server che, mediante un disk buffer, fa da frontend ad un sistema complesso di mass storage (librerie di nastri e/o server di dischi). *StoRM* sfrutta le caratteristiche dei file system paralleli per l'accesso alle aree disco.

Un *data file* rappresenta la granularità più fine dei dati poiché i file su uno Storage Element non possono essere modificati, ma solo cancellati o sovrascritti; possono esistere numerose repliche dello stesso file, in locazioni diverse e devono essere accessibili da ogni punto della Grid. È possibile trasferire file tra file system locali (ad es. la propria User Interface) e la Grid (cioè uno Storage Element).

Un file nella GRID viene identificato, in maniera univoca, dal suo GUID (GRID Unique Identifier, la cui unicità è garantita da un algoritmo), che assolutamente non è user friendly. D'altra parte è possibile associare al file un *Logical File Name* (LFN) , che definisce un alias del GUID, ed è creato dall'utente per meglio identificare il file; I nomi logici LFN sono organizzati in una struttura gerarchica di directory; ad esempio: *lfn:/grid/cms/20090203/run2/track1*. Le copie fisiche dei file sono individuate dalla *Storage URL* o *Site URL* (SURL) che include l'indirizzo dello Storage Element ed il *path* del file (per un SRM il path è quello del file system virtuale), ad esempio: *srm://srm.cern.ch/castor/cern.ch/cms/output10_1*. Infine un file si può indicare anche con il *Transport URL* (TURL), che contiene il protocollo di accesso al file: *gsiftp://pcrd24.cern.ch/flatfiles/cms/output10_1*

Ogni file nella Grid può avere più nomi logici (*alias*) e più copie fisiche (*repliche*), ma un solo GUID (rif. Figura 5-6). I file possono essere replicati per vari motivi. I motivi per cui è consigliabile replicare i file sono: metterli vicino ai CE per questioni di efficienza e resistere a eventuali guasti di un SE.

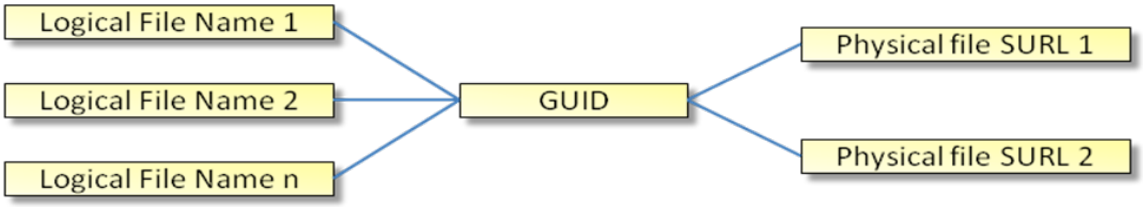



**Figura 5-6: Schema di naming del file**

Il mapping tra LFN, GUID e SURL è fornito da un unico catalogo, chiamato LCG File Catalog (LFC). Un file si può considerare un Grid file solo se è fisicamente presente su uno Storage Element ed è registrato in un LFC. L'LFC gestisce anche i *metadata* di un Grid File: *system metadata* e *user metadata*. I *system metadata* contengono la dimensione del file, che resta sempre invariata e una checksum, che viene calcolata ad ogni replica del file per verificarne l'integrità. Gli *user metadata* descrivono i dati contenuti nel file e possono essere utilizzati per correlare files in insiemi, contenenti dati dello stesso tipo.

La chiave principale di accesso all'LFC è il nome logico, dove lo spazio dei nomi è organizzato ad albero di directory (esempio: */grid/<VO>/<path>*) e i permessi di accesso sono molto simili alle ACL di un filesystem Unix.

***Protocolli di accesso ai dati***

Il protocollo di base per il trasferimento dei file della Grid è GSIFTP, un'estensione di FTP (*standard File Transfer Protocol*) che include autenticazione e encryption dei servizi di sicurezza GSI ed è in grado di trasferire multiple stream di dati in parallelo.

I requisiti richiesti per il trasferimento dei dati in Grid sono: *Velocità*, *Sicurezza* e *Robustezza*. Infatti si vuole utilizzare al massimo la velocità consentita dalle connessioni fisiche, minimizzando l'overhead dovuto a servizi e protocolli; i file

devono essere trasferiti solo tra entità authenticate e i servizi devono essere stabili e va implementata una fault tolerance.

In risposta a tali requisiti è stato realizzato *GridFTP*, basato sul protocollo standard FTP, ma realizzato per effettuare operazioni ad alte prestazioni.

Realizzato nel rispetto degli standard di sicurezza di Grid (GSI), che prevedono l'autenticazione degli utilizzatori (utenti o servizi), *GridFTP* ha la peculiarità di supportare stream multiple di trasferimento in parallelo, sia tra singolo end-point sorgente e singolo end-point di destinazione che tra diversi end-point sorgenti (contenenti repliche dello stesso Grid file) e singolo end-point di destinazione. Inoltre, GridFTP è in grado di effettuare trasferimenti di file parziali, che consentono, tra l'altro, di supportare trasferimenti di dati affidabili e riavviabili dal punto dell'eventuale interruzione.

L'accesso diretto ai dati remoti su un SE avviene tramite il protocollo RFIO (*Remote File Input/Output*), che consente di "leggere" i dati su un SE senza trasferirli, implementando una versione remota delle funzioni di I/O standard Posix, quali *open*, *read*, *write*, *lseek* and *close*.

Esistono una versione sicura ed una insicura di RFIO. La versione insicura può essere usato solo per accedere ai dati in una LAN, tipicamente dai Working Node allo Storage Element, poiché il controllo dell'accesso ai dai è fatto sulla user ID e non sulle credenziali VOMS. L'RFIO in versione sicura (gsirfio) può essere usato per l'accesso remoto ai file anche da una User Interface o da un Working Node esterno al sito, dato che l'autorizzazione e l'autenticazione sono basate su GSI.

#### 5.2.2.6 Monitoring

I Service Level Agreement (SLA) definiscono formalmente metriche di servizio che devono essere rispettate da un fornitore di servizi. Un esempio potrebbe essere a richiesta di servizio attivo 24 ore al giorno, 365 giorni l'anno, con un downtime dell'1% nell'arco temporale di un anno. Un altro esempio di SLA potrebbe essere una connessione di rete punto-punto tra due Grid-site con banda garantita di 1Gbit/s, etc..

Nell'ambito di una Grid di produzione si danno le seguenti definizioni:

- *Scheduled Downtime*: arco di tempo, programmato, in cui un sito Grid non fornisce servizi, ad esempio per manutenzione della sala macchine o per aggiornamenti del sistema operativo.
- *Unscheduled Downtime*: tempo in cui un sito Grid si trova in uno stato di non raggiungibilità a causa di problemi inattesi. È un downtime non programmato.
- *Availability* (*Disponibilità*): misurata in ore, e rappresenta il tempo in cui un sito Grid risulta attivo e funzionante.
- *Reliability*: È un indice di affidabilità si un sito ed è definito nel seguente modo: *Reliability* = *Availability* / (*Availability* + *Unscheduled Downtime*).

Le definizioni precedenti vengono utilizzate per definire dei livelli di SLA a cui un sito in Grid deve aderire in funzione del suo ruolo. Ad esempio, ogni sito inserito nella griglia di produzione deve risultare *Available* (disponibile) almeno per il 70% del tempo nell'arco di un mese e la sua *Site Reliability* deve essere almeno del 75%.

Il *Monitoring* è un servizio di controllo implementabile a tutti i livelli di un infrastruttura IT ed è erogato su tutte le componenti (*layer* tecnologici) dichiarate all'interno del Service Level Agreement. Questo servizio rappresenta la base di partenza per l'attivazione di un servizio di gestione e controllo completo. Alcuni dei principali usi dei servizi di Monitoring sono: supporto per il rispetto del necessario livello di SLA, verifica dello stato delle risorse e dei servizi, individuazione dei fault, esecuzione del troubleshooting, individuazione dei colli di bottiglia e produzione di report statistici.

In ambito Grid i sistemi di Monitoring si basano su sensori e servizi che eseguono misure sulle risorse e sui servizi distribuiti. Il Grid Monitoring è un'attività strettamente collegata ai sistemi informativi, che permettono di propagare e rendere fruibili i dati ottenuti.

Il servizio di monitoring viene effettuato a tre livelli: livello fabric, livello collective e livello application.

***Monitoraggio di livello Fabric***

Il monitoraggio di livello Fabric, monitoring a più basso livello generalmente rivolto agli addetti ai lavori, rappresenta la base per garantire qualsiasi altro servizio al di sopra della pila GRID ed è una problematica comune a tutti i sistemi informatici, ivi compresi Datacenter per la pubblica amministrazione, servizi di rete, telefonia.

Con il monitoraggio di livello Fabric si possono controllare servizi quali: UPS (*uninterruptable power supply*), tensione e consumo sulle prese di alimentazione dei server, impianti e sistemi di condizionamento, temperatura delle CPU, stato delle ventole dei server, temperatura ambientale e dei rack, stato delle testine sul disco, stato della rete, congestioni, traffico e Fault.

Lo strumento più utilizzato per questo scopo è il *Simple Network Management Protocol* (SNMP), un protocollo di comunicazione basato su IP con architettura di tipo client/server (descritto come protocollo di management). Esso nasce per consentire la gestione di intere reti geografiche e per tenere sotto controllo qualsiasi tipo di apparato (il *draft* che descrive il protocollo SNMP è la RFC 1157).

Nell'architettura SNMP, per ogni sottosistema è definita una base dati detta *Management Information Base* (MIB), gestita dal corrispondente *subagent*, che rappresenta lo stato del sottosistema gestito, o meglio, una proiezione di tale stato limitata agli aspetti di cui si vuole consentire la gestione.

Un altro sistema di monitoring molto diffuso di livello fabric è il software Ganglia, nato per controllare lo stato di occupazione e di carico delle CPU, della memoria e di altri parametri per infrastrutture di tipo High Performance Computing. È

un sistema basato su architettura client/server con sensori locali ai nodi da monitorare che estraggono le informazioni dal sistema operativo e le propagano verso un server.

***Monitoring di livello Collective***

È il livello di monitoraggio che controlla lo stato dei servizi Grid, ed ha una doppia utenza: sia gli addetti alla gestione sia gli utenti. Esso viene utilizzato per verificare lo stato di funzionamento dei siti e dell'infrastruttura Grid. Presuppone una conoscenza dell'architettura Grid.

L'aumento del livello di astrazione aumenta la comunità di utenza a cui si rivolge il servizio di monitoraggio.

Il Grid Monitoring può essere definito come la misurazione di parametri significativi delle risorse presenti in Grid. Esso rappresenta un'attività strategica per il calcolo distribuito, necessaria per: *performance analysis*, *resources/services fault detection*, *problems spotting*, *statistics and capacity planning*, *auditing system*.

Il sistema di Grid Monitoring deve avere le seguenti caratteristiche:

- *scalabilità*: il servizio di monitoring deve garantire la stessa qualità del servizio indipendentemente dal numero di risorse monitorate
- *bassa intrusività*: le attività di monitoring non devono compromettere le prestazioni offerte tramite i servizi
- *formato degli eventi di monitoring*: compromesso tra semplicità di utilizzo e compattezza (esempio documento XML)
- *security*: non deve abbassare il livello di sicurezza del sistema
- *architettura distribuita*: evitare il *single point of failure* e favorire delle architetture a componenti modulari
- *aggiornamenti frequenti*: le informazioni collezionate hanno un tempo di utilità molto breve

Uno strumento utilizzato per il monitoring a livello collective è Nagios: un software open-source per il monitoraggio ed il controllo costante di server e di servizi molto usato nel mondo Linux. Nagios è in grado di eseguire controlli su un'ampia serie di servizi quali HTTP, FTP, SSH, numero di processi attivi, carico dei server, numero di utenti collegati, risposta del server ai *ping*, controllo dei DNS, controllo del demone di MySql, fare query con protocollo SNMP e molto altro ancora, e fornisce l'output su web.

In ambito Collective è possibile creare dei sistemi di monitoraggio basati sulle informazioni dei BDII senza l'ausilio di sensori locali alle risorse. Tali sistemi presentano in maniera grafica ed aggregata le informazioni utilizzate dagli altri Grid Services per tutte le operazioni e per il funzionamento globale della Grid. Tali sistemi sono i meno invasivi ma offrono inevitabilmente un set di informazioni limitate.

***Monitoring di livello Application***

Esegue il monitoring dei servizi Grid dal punto di vista dell'utente e dell'applicazione. È rivolto sia agli utenti finali, che per la preparazione di statistiche ed

il calcolo dei parametri di Reliability ed Availability. È altresì utilizzato come strumento di allarme.

Il *Service Availability Monitoring* (SAM) è un insieme di strumenti per il monitoring dei siti Grid in ambiente di produzione e di preproduzione. Prevede un set di test che vengono sottomessi con intervalli regolari e la gestione di un database che conserva il risultato dei test.

I dati raccolti vengono pubblicati con interfaccia web e presentati in maniera fruibile. SAM effettua un monitoring dei servizi Grid dal punto di vista dell'utente. La versione per gLite è basata su Nagios.

SAM prevede due classi principali di test:

- *Test di base*: Sono i plugin generali che verificano lo stato dei servizi basilari di ogni singolo ruolo Grid; ad esempio, nel caso del logical file catalog, esegue dei comandi di scrittura, creazione di directory e cancellazione di file e cartelle; nel caso dei CE sottomette dei job e recupera l'output e così via.
- *Test Personalizzati*: Sono i test aggiuntivi che riguardano il funzionamento di particolari applicativi o servizi specifici di una VO, ad esempio test di funzionamento di librerie o di altri software di specifico interesse.

## 5.3 Il Tier 2 di ATLAS ai LNF

Il Servizio di Calcolo dei LNF collabora alla gestione di uno dei quattro Tier-2 italiani dell'esperimento ATLAS. Il Tier-2 è situato nella sala macchine dell'edificio Calcolo e sfrutta gli stessi impianti e sistemi infrastrutturali dedicati ai servizi centrali.

Il Tier-2 di ATLAS si è costituito attraverso vari acquisti diluiti negli ultimi anni. Ora la farm che costituisce il Tier-2 conta complessivamente 3 armadi (*rack*) con 176 slot di calcolo (per una potenza di calcolo complessiva di circa 1400 HEP SPEC) e 110 TeraByte di spazio disco raw.

***Lo storage***

Un rack è dedicato allo storage, che è composto da tre storage controller duali (ridondati), cui sono connessi dischi in tecnologia SATA:

1. Dual Storage Controller SUN/StorageTek Flexline 210 (4 canali host FC 2Gb/s), con 30 dischi SATA da 450 GB 7200 RPM.
2. Dual Storage Controller SUN/StorageTek 6140 (4 canali host FC 4Gb/s), con 48 dischi SATA da 1 TB 7200 RPM.
3. Dual Storage Controller E4 E6560 (4 canali host FC 4Gb/s), con 24 dischi SATA da 2 TB 7200 RPM.

I volumi sono configurati RAID-5 o RAID-6 e sono serviti in Fibre Channel direttamente ai disk server. I disk server sono quattro, connessi agli storage controller come rappresentato nella Figura 5-7.

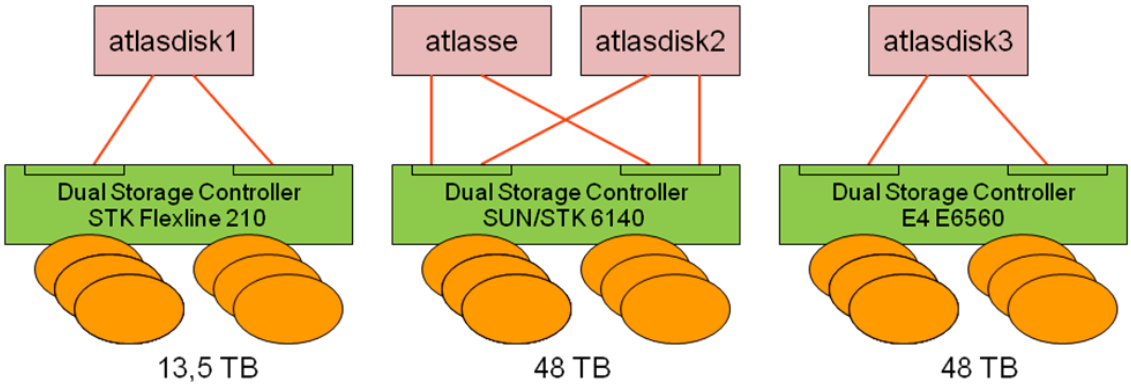



**Figura 5-7: Schema del sistema di storage**

***I server***

Altri due rack ospitano i worker node ed i server della Grid (compresi i disk server appena descritti). In particolare alcune macchine sono di tipo slot 1U (montaggio da rack, altezza 1 unità), altre sono in tecnologia blade, per la maggior parte dual processor quad-core. In dettaglio:

1. 6 Server di tipo slot 1U, dual processor (Intel Xeon), 2GB RAM.
2. Un Blade System p-Class HP con 2 switch Gigabit Ethernet integrati e 2 moduli Fibre Channel Pass-Through; 6 lame BL25p dual processor dual-core, 8GB RAM, daughter card FC.
3. Un Blade System H Chassis IBM con 2 switch Gigabit Ethernet integrati; 14 lame HS21 dual processor quad-core (Intel Xeon), 16GB RAM.
4. 3 Server di tipo slot 1U della E4, dual processor quad-core (Intel Xeon), 16GB RAM.
5. 3 Server di tipo slot 1U twin della E4 (2 macchine indipendenti montate nello spazio di una unità); ogni macchina con dual processor quad-core (Intel Xeon), 24GB RAM.

Ciascuno dei due rack è dotato di uno switch di rete 3com 4500G configurato con 48 porte Gigabit Ethernet e 2 uplink a 10 Gigabit Ethernet. I server all'interno del rack sono connessi alle porte Gigabit Ethernet, mentre l'interconnessione tra i rack avviene grazie ad uno dei due uplink a 10Gb/s. Uno dei due switch 3com è connesso allo switch centrale del Servizio di Calcolo tramite una porta Gigabit Ethernet. Lo schema di interconnessione è rappresentato nella Figura 5-8.

Ai nodi di rete del Tier-2 è dedicata una network IP pubblica di classe C per la raggiungibilità sulla rete Internet (indirizzo: 192.84.128.0/24) e due private per l'accesso alle console e per la gestione degli apparati (indirizzi: 192.168.218.0/24 e 192.168.220.0/24).

Il sistema operativo installato sui *Worker Node* è Scientific Linux versione 5.4, mentre quello installato sui server di disco e sugli altri server è Scientific Linux versione 4.8.

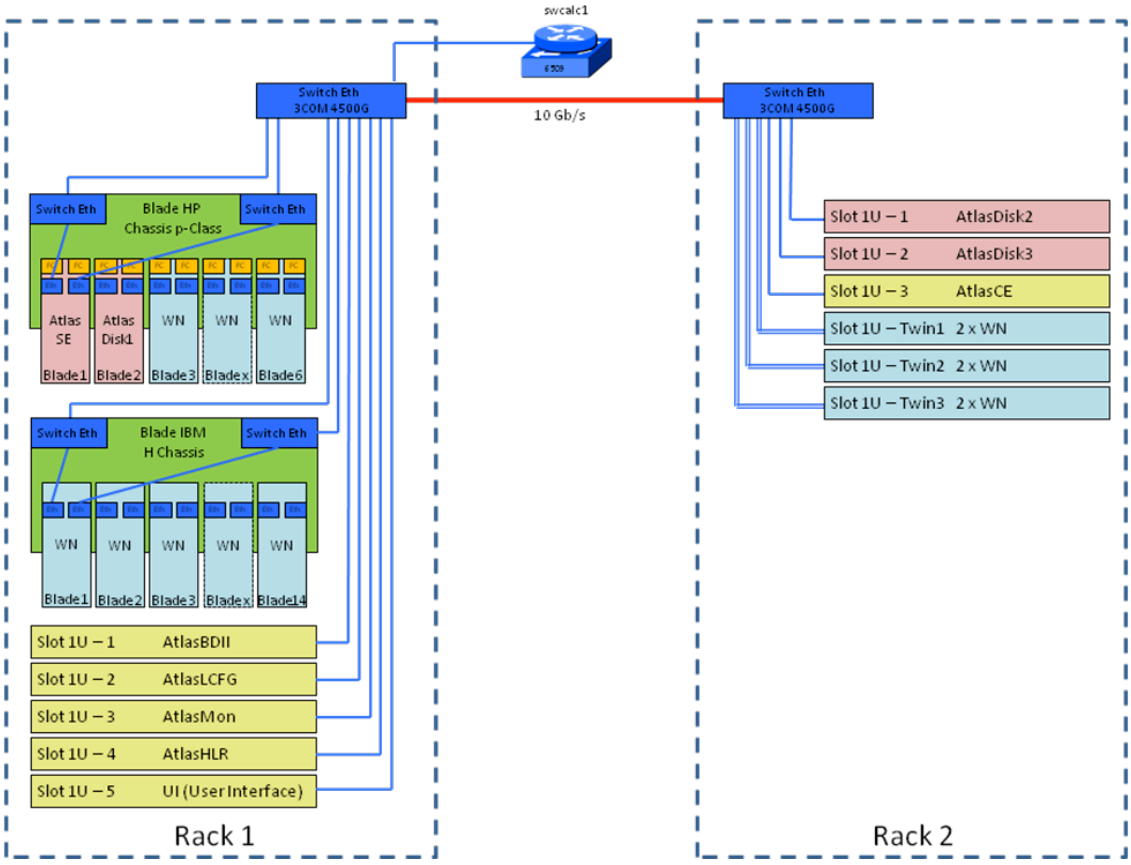



**Figura 5-8: Schema di interconnessione tra i server del Tier-2 di ATLAS**

L'installazione del sistema operativo sui server e sui Worker Node è automatizzata attraverso un servizio svolto dal server AtlasLCFG. Questo servizio è in grado di fornire un'immagine iniziale di boot via rete, sfruttando il protocollo PXE (Preboot eXecution Environment). La macchina da installare fa una richiesta broadcast di boot sulla rete LAN, alla quale rispondono il DHCP server offrendo le impostazioni IP e il PXE server offrendo l'immagine di boot. Il client prosegue l'installazione automaticamente utilizzando un file di configurazione precostruito (kickstart) e un repository di rete contenente le distribuzioni Linux.

### *Il middleware*

Il Tier-2 di ATLAS dei LNF si basa sulla release INFNGRID del middleware gLite (gLite versione 3.1 sui server e gLite versione 3.2 sui Worker Node). Come già visto nel paragrafo 5.2.2.1, gli elementi principali che caratterizzano il middleware gLite sono gli Storage Element il Computing Element, i Worker Node e il site BDII.

Nella precedente figura tali elementi sono ben visibili. In particolare gli Storage Element, rappresentati in rosa, sono implementati da 2 server del blade HP e da due server slot 1U di recente acquisizione; come Storage Resource Manager implementano la soluzione DPM (Disk Pool Manager). In particolare il server ATLASSE svolge il ruolo di DPM head node ed ospita il database MySQL del DPM. Tutti i server di disco svolgono il ruolo di DPM pool node.

Il Computing Element (profilo lcg-CE) è implementato su una slot 1U di recente acquisizione, rappresentata in giallo sul rack di destra. Il sistema di gestione delle code batch è basato sul s/w open-source Torque/PBS (Portable Batch System). I Worker Node sono implementati sia sui blade system (HP e IBM), sia sui server Slot1U twin (rappresentati in celeste in entrambi i rack).

Alcuni servizi che richiedono meno prestazioni, ma non per questo meno critici, sono implementati sulle più vecchie slot 1U della HP (rappresentati in giallo sul rack di sinistra): il site BDII, il server per il monitoring, il server per l'accounting, e la User Interface.

# 6. Cenno alla sicurezza informatica

Nel mondo informatico non bisogna mai trascurare le problematiche relative alla sicurezza, per la salvaguardia dei sistemi da potenziali rischi e/o violazioni dei dati. Sarebbe estremamente superficiale pensare di trovarsi in una condizione di assoluta sicurezza, illudendosi che sia tutto sotto controllo, mentre si gestisce un'infrastruttura informatica non esente da vulnerabilità, costantemente sotto la minaccia di attacchi di vario tipo.

I principali aspetti di protezione dei dati sono:

- *Confidenzialità*: si intende la protezione dei dati e delle informazioni scambiate tra un mittente e uno o più destinatari nei confronti di terze parti. Tale protezione deve essere realizzata a prescindere dalla sicurezza del sistema di comunicazione utilizzato; assume anzi particolare interesse il problema di assicurare la confidenzialità quando il sistema di comunicazione utilizzato è intrinsecamente insicuro (come ad esempio la rete internet). In un sistema che garantisce la confidenzialità, una terza parte che entri in possesso delle informazioni scambiate tra mittente e destinatario, non è in grado di ricavarne alcun contenuto informativo intelligibile. Per assicurarla si ricorre a meccanismi di cifratura e all'occultamento della comunicazione. I meccanismi di cifratura garantiscono la confidenzialità per il tempo necessario a decriptare il messaggio. Per questo motivo occorre stabilire per quanto tempo il messaggio deve restare confidenziale. Non esistono meccanismi di protezione sicuri in assoluto.
- *Integrità dei dati*: si intende la protezione dei dati e delle informazioni nei confronti delle modifiche del contenuto, accidentali oppure effettuate da una terza parte, essendo compreso nell'alterazione anche il caso limite della generazione ex novo di dati ed informazioni. I dati e le informazioni possono essere sia scambiati tra un mittente ed uno o più destinatari, sia memorizzati e/o archiviati su un generico supporto. L'integrità dei dati garantisce anche l'integrità del supporto che li contiene (ad es. CD, DVD...) o di un software, ad esempio un database. Insito nel concetto di integrità vi è la possibilità di verificare con assoluta certezza se un dato o una informazione siano rimasti integri, ossia inalterati nel loro contenuto, durante la loro trasmissione e/o la loro memorizzazione. In un sistema che garantisce l'integrità, l'azione di una terza parte di modifica del contenuto delle informazioni scambiate tra mittente e destinatario, viene quindi rilevata.
- *Disponibilità*: misura l'attitudine di un'entità ad essere in grado di svolgere una funzione richiesta in determinate condizioni ad un dato istante, o durante un dato intervallo di tempo, supponendo che siano assicurati i mezzi esterni eventualmente necessari.

La protezione dagli attacchi informatici viene ottenuta agendo su più livelli: innanzitutto a livello fisico e materiale, *sicurezza passiva*, ponendo i server in luoghi il più possibile sicuri, dotati di sorveglianza e/o di controllo degli accessi; anche se questo accorgimento fa parte della sicurezza normale e non della "sicurezza informatica" è

sempre il caso di far notare come spesso il fatto di adottare le tecniche più sofisticate generi un falso senso di sicurezza che può portare a trascurare quelle semplici.

Il secondo livello è normalmente quello logico che prevede l'autenticazione e l'autorizzazione di un'entità che rappresenta l'utente nel sistema. Successivamente al processo di autenticazione, le operazioni effettuate dall'utente sono tracciate in file di *log*. Questo processo di monitoraggio delle attività è detto *audit* o *accountability*. Il livello logico rientra nella *sicurezza attiva*.

Per sicurezza attiva si intendono le tecniche e gli strumenti mediante i quali le informazioni ed i dati di natura riservata sono resi intrinsecamente sicuri, proteggendo gli stessi sia dalla possibilità che un utente non autorizzato possa accedervi (*confidenzialità*), sia dalla possibilità che un utente non autorizzato possa modificarli (*integrità*).

È evidente che la sicurezza passiva e quella attiva sono tra loro complementari ed entrambe indispensabili per raggiungere il desiderato livello di sicurezza di un sistema.

Le possibili tecniche di attacco sono molteplici, perciò è necessario usare contemporaneamente diverse tecniche difensive per proteggere un sistema informatico, realizzando più barriere fra l'attaccante e l'obiettivo.

Spesso l'obiettivo dell'attaccante non è rappresentato dai sistemi informatici in sé, quanto piuttosto dai dati in essi contenuti, quindi la sicurezza informatica deve preoccuparsi di impedire l'accesso ad utenti non autorizzati, ma anche a soggetti con autorizzazione limitata a certe operazioni, per evitare che i dati appartenenti al sistema informatico vengano copiati, modificati o cancellati.

Le violazioni possono essere molteplici: vi possono essere tentativi non autorizzati di accesso a zone riservate, furto di identità digitale o di file riservati, utilizzo di risorse che l'utente non dovrebbe potere utilizzare etc. La sicurezza informatica si occupa anche di prevenire eventuali Denial of service (DoS). I DoS sono attacchi sferrati al sistema con l'obiettivo di rendere non utilizzabili alcune risorse in modo da danneggiare gli utenti del sistema. Per prevenire le violazioni si utilizzano strumenti hardware e software.

## 6.1 Principali tipologie di attacchi

L'esecuzione di un codice che, sfruttando un *bug* o una vulnerabilità, porta all'acquisizione di privilegi o al denial of service di un computer è chiamato *exploit*.

Ci sono diversi modi per classificare gli exploit. Il più comune è una classificazione a seconda del modo in cui l'exploit contatta l'applicazione vulnerabile. Un *exploit remoto* è compiuto attraverso la rete e sfrutta la vulnerabilità senza precedenti accessi al sistema. Un *exploit locale* richiede un preventivo accesso al sistema e solitamente fa aumentare i privilegi dell'utente oltre quelli impostati dall'amministratore. Segue una tabella con i principali tipi di exploit:

| CATEGORIA | DESCRIZIONE |
|---|---|
| Overflow | Gli exploit di overflow sono tipici dei sistemi Linux/Unix ma con il tempo si sono manifestati anche in ambiente Windows. Sono una categoria di exploit complessa da creare che richiede la conoscenza approfondita del sistema operativo e del linguaggio Assembly. Questi exploit sfruttano la vulnerabilitá della programmazione dinamica (l'uso dei puntatori) in linguaggio C, per eseguire codice a scelta e istruzioni iniettate direttamente dall'aggressore. A loro volta si dividono in altre sottocategorie: *Buffer Overflow, Heap Overflow, Frame Pointer, Format String.* |
| Denial Of Service | *Denial Of Service*(abbreviato con Dos) è un attacco che mira alla negazione di un particolare servizio. Un exploit che nega tutte le connessioni di rete è un DoS; allo stesso modo un exploit in grado di saturare la CPU di un calcolatore impedendogli di effettuare ogni altra cosa è anch'esso un DoS. Si tratta di exploit che non consentono di accedere in maniera illecita ad una risorsa, ma che possono però bloccare l'accesso alla risorsa a chiunque altro. |
| Weakness | Molti protocolli, algoritmi crittografici, sistemi di manutenzione spesso presentano alcune debolezze intrinseche, dovute ad una cattiva progettazione, che li rendono vulnerabili ad attacchi mirati. Un esempio tipico può essere un web server che consente l'accesso libero a pagine di amministrazione, oppure un'applicazione che memorizza le password in chiaro su un certo file a portata di tutti o ancora il router che nasconde una backdoor attivabile tramite un apposito comando segreto. |
| Information Disclosure | Spesso capita che alcune informazioni sensibili (una password, un indirizzo IP privato, un percorso di una directory) vengano rivelate da un server in maniera banale tramite opportune richieste. Gli exploit che fanno uso di queste vulnerabilità non sono molto pericolosi, ma a volte combinati con altri exploit possono diventare un serio problema. |
| SQL-Injection | L'interrogazione dei database attraverso il linguaggio SQL può essere un veicolo di attacco quando le query vengono composte utilizzando stringhe di input passate dall'esterno. L'esempio classico è quello della stringa "**' OR 1=1**" che consente spesso di bypassare l'accesso a quei siti web che interagiscono con un database senza filtrare alcuni caratteri speciali come l'apice. |
| Privilege Escalation | A volte può accadere che una certa istruzione possa eseguire comandi privilegiati a cui l'utente non avrebbe normalmente diritto. Alcuni exploit sfruttano questa tecnica per ottenere l'accesso "root" di un sistema partendo da account di livello piú basso. |

**Tabella 6-1: Principali tipologie di exploit**

Oltre alle tipologie di attacco già viste, esiste una tipologia di attacco chiamata *Man in the middle* (uomo in mezzo); consiste nel dirottare il traffico generato durante la

comunicazione tra due nodi verso un terzo nodo (attaccante). Durante l'attacco è necessario far credere ad entrambi i nodi terminali della comunicazione che l'host attaccante sia il loro interlocutore legittimo. Tuttavia, la maggior parte degli attacchi che rientrano in questa categoria richiedono un accesso fisico alla rete locale.

Per come è strutturato l'ambiente di lavoro, che prevede l'accesso solo ai dipendenti e agli associati attraverso un controllo di identità da parte di un apposito servizio di vigilanza, questo tipo di attacchi risulterebbe estremamente meno probabile e quindi esula da questa trattazione.

Nei paragrafi successivi verranno invece trattati brevemente tutti i tipi di attacchi più frequenti di provenienza da siti remoti, attraverso la rete internet.

## *6.2 Esplorazione dell'obiettivo*

L'attaccante informatico viene spesso definito *hacker* anche se la definizione più corretta dovrebbe essere *cracker*.

Ogni attacco ad una rete viene sempre preceduto da una fase di esplorazione, nella quale viene studiata a fondo la subnet oggetto dell'attacco, cercando di raccogliere quante più informazioni possibili su di essa e sugli host presenti al fine di scegliere le tecniche di attacco più adeguate.

### 6.2.1 Banner grabbing

Se l'obiettivo dell'attacco è un host individuato da uno specifico indirizzo IP, per raccogliere informazioni, l'hacker dapprima esegue un *ping-scan* per trovare tutti gli host attivi sulla subnet del target. Questa prima fase esplorativa fornisce un elenco di possibili punti di attacco.

Una volta ottenuti gli host attivi, l'hacker tenterà di scoprire quali sistemi operativi essi montano e quali servizi e porte TCP/IP sono attive, perché proprio dalla vulnerabilità di una di queste potrebbe riuscire a violare il sistema.

Prima dell'avvento delle tecniche di OS Fingerprinting, basate sull'analisi dello stack e sulla produzione di pacchetti TCP/IP anomali, i metodi più utilizzati erano l'analisi dei "banner" e delle porte aperte, che in modo abbastanza banale spesso rivelano quale sistema operativo è in uso.

### 6.2.2 Port scanning

La tecnica di analizzare i banner di risposta dei servizi attivi fornisce quasi sempre risultati interessanti, ma in molti casi potrebbe non essere sufficiente. L'hacker può ricorrere ad un'altra tecnica che riguarda l'analisi delle porte aperte. Lo studio delle porte aperte e dei servizi in esecuzione spesso rileva in maniera esatta che tipo di sistema operativo è in esecuzione.

Ad esempio, la presenza di porte come la 23 (telnet) o la 22 (ssh) su una macchina Windows è alquanto improbabile, così come la presenza delle porte 135/139/445 (NetBios) su macchine Unix. Esistono molte altre porte significative, come la 1433/1434 che si riferiscono generalmente al database MS-SQL Server e che identificano presumibilmente un sistema Windows; allo stesso modo la porta 3306, relativa a MySQL è più probabilmente presente su macchine Linux.

Un elenco completo di porte e servizi corrispondenti (well know ports) è disponibile su Internet al link:

*http://www.iana.org/assignments/port-numbers*

La corrispondenza fra porte e sistemi operativi viene sempre concepita in termini probabilistici perché nulla vieta ad un host Windows di montare un server SSH sulla porta 22.

### 6.2.3 OS fingerprinting

Determinare un sistema operativo di una macchina in remoto non è del tutto banale; L'*OS Fingerprinting* è il nome della tecnica in grado di risolvere questo problema, determinando in maniera esatta (o con probabilità molto elevata) il tipo di sistema operativo di un certo host attraverso l'analisi di particolari informazioni e caratteristiche fornite dallo stesso host, che in un certo senso costituiscono l'impronta digitale di quel sistema (da qui in termine *fingerprint*).

Oggi esistono diverse tecniche di fingerprinting, implementate da diversi programmi: uno dei tool più noti e facili da usare è *Nmap* (open source).

Le tecniche di fingerprinting implementate da Nmap sono diverse ma tutte basate sull'analisi dello stack del protocollo TCP/IP; una qualsiasi connessione TCP/IP tra due computer viene instaurata seguendo il meccanismo noto come *"Three Way Hand-Shaking"* (accordo a tre mani). Questo protocollo di creazione non richiede che entrambi i lati inizino a spedire pacchetti con lo stesso numero di sequenza, e quindi può essere utilizzato con diversi metodi di sincronizzazione.

Le specifiche di questa procedura vengono stabilite dalla RFC793, anche se nella realtà ciascun sistema operativo la implementa con lievi modifiche. Queste sottili diversità consentono ad Nmap di creare uno schema logico e catalogare i diversi sistemi operativi da remoto.

## *6.3 Malware e Virus informatici*

Si definisce *malware* un qualsiasi software creato con il solo scopo di causare danni più o meno gravi al computer su cui viene eseguito. Il termine deriva dalla contrazione delle parole inglesi *malicious* e *software* e ha dunque il significato letterale di "programma malvagio"; in italiano è detto anche codice maligno.

Si distinguono molte categorie di malware, anche se spesso questi programmi sono composti di più parti interdipendenti e rientrano pertanto in più di una classe: *Virus, Worm, Trojan Horse, Backdoor, Spyware, Dialer, Adware, etc*..

- I *Virus* sono parti di codice che si diffondono copiandosi all'interno di altri programmi, o in una particolare sezione del disco fisso, in modo da essere eseguiti ogni volta che il file *infetto* viene aperto. Si trasmettono da un computer a un altro tramite lo spostamento di file infetti ad opera degli utenti (o via e-mail).
- I *Worm* non hanno bisogno di infettare altri file per diffondersi, perché modificano il sistema operativo della macchina ospite in modo da essere eseguiti automaticamente e tentare di replicarsi sfruttando per lo più Internet.
- I *Trojan horse* sono programmi che, oltre ad avere delle funzionalità "lecite", utili per indurre l'utente ad utilizzarli, contengono istruzioni dannose che vengono eseguite all'insaputa dell'utilizzatore. Non possiedono funzioni di auto-replicazione, quindi per diffondersi devono essere consapevolmente inviati alla vittima. Il nome deriva dal famoso cavallo di Troia.
- Le *Backdoor*, letteralmente "porta sul retro", sono dei programmi che consentono un accesso non autorizzato al sistema su cui sono in esecuzione. Tipicamente si diffondono in abbinamento ad un trojan o ad un worm, oppure costituiscono una forma di accesso di emergenza ad un sistema, inserita per permettere ad esempio il recupero di una password dimenticata.
- Gli *Spyware* vengono usati per raccogliere informazioni dal sistema su cui sono installati e per trasmetterle ad un destinatario interessato. Le informazioni carpite possono andare dalle abitudini di navigazione fino alle password e alle chiavi crittografiche di un utente.
- Gli *Adware* sono programmi che presentano all'utente messaggi pubblicitari durante l'uso, a fronte di un prezzo ridotto o nullo. Possono causare danni quali rallentamenti del pc e rischi per la privacy in quanto comunicano le abitudini di navigazione ad un server remoto.

Tutte queste tipologie di malware impediscono il corretto funzionamento di una macchina, sia di una postazione di lavoro client, sia di un server. Mentre è relativamente facile proteggere un server, è molto più difficile proteggere i nodi client della rete, soprattutto perché il controllo e la gestione dei client è spesso effettuata ad opera degli stessi utenti.

La pericolosità dei malware non è da sottovalutare, in quanto, un nodo *infettato* sulla LAN può aprire connessioni verso l'esterno mettendo a rischio la sicurezza di altri nodi sulla rete e rendendo tutta l'infrastruttura vulnerabile. Inoltre essendo a volte particolarmente *contagiosi*, ovvero propagandosi velocemente tra più client sulla rete, potrebbero essere in grado di congestionare il traffico della rete locale o saturare la banda del link d'accesso alla rete geografica.

## 6.4 Mail spam

Lo *spamming* è l'invio di grandi quantità di messaggi indesiderati, generalmente commerciali, attraverso l'e-mail. Il principale scopo dello spamming è la pubblicità, il cui oggetto può andare dalle più comuni offerte commerciali a proposte di vendita di materiale pornografico o illegale, come software pirata e farmaci senza prescrizione medica, da discutibili progetti finanziari a veri e propri tentativi di truffa. Uno *spammer*, cioè l'individuo autore dei messaggi spam, invia messaggi identici a migliaia di indirizzi e-mail. Questi indirizzi sono spesso raccolti in maniera automatica dalla rete mediante spambot ed appositi programmi, ottenuti da database o semplicemente indovinati usando liste di nomi comuni.

Per definizione lo spam viene inviato senza il permesso del destinatario ed è un comportamento considerato inaccettabile dalla maggior parte degli utenti di Internet. Gli utilizzatori del servizio di e-mail trovano lo spam fastidioso e con contenuti spesso offensivi, mentre i fornitori del servizio vi si oppongono anche per i costi del traffico generato dall'invio indiscriminato.

Dal punto di vista della sicurezza informatica, pur non essendo grave e pericoloso come una intrusione diretta in un server o in un client, o come una violazione di dati, sondaggi hanno indicato che lo spam è considerato uno dei maggiori fastidi di Internet; l'invio di questi messaggi costituisce una violazione del contratto *"Acceptable Use Policy"* (condotta d'uso accettabile) dei servizi di Internet.

Essendo l'attività di spamming generalmente perseguita, gli spammer cercano sempre di nascondere la propria identità e spesso riescono a inviare le e-mail spam attraverso smtp server non adeguatamente protetti. All'occorrenza, in alcuni casi, dopo essere riusciti a violare un nodo qualunque su internet e ad acquisirne i privilegi amministrativi, procedono con l'installarci un server smtp con l'unico scopo di inviare grandi quantitativi di mail spam.

## 6.5 Soluzioni di protezione adottate ai LNF

Come già accennato, l'accesso fisico nell'area geografica dei LNF è sottoposto a controllo di identità da parte di un apposito servizio di vigilanza e ciò semplifica in parte le metodologie di protezione da adottare. In particolare fa presumere che un eventuale primo attacco ai server possa partire con maggiore probabilità dall'esterno della Local Area Network.

Per le tipologie di attacco che provengono dall'esterno le metodologie di protezione prevedono l'uso di una difesa perimetrale, ovvero una tecnica di difesa posta sulle vie di accesso alla rete geografica. In particolare ai LNF la via di accesso è una sola (link Gigabit Ethernet di connessione alla rete GARR dal router di frontiera dei LNF).

Tuttavia non è da escludere del tutto che un attacco possa partire anche dall'interno della LAN. In effetti qualunque utente, inconsapevolmente, può introdurre

una sorgente di attacchi, connettendo alla LAN un PC portatile che è stato violato (o infettato) all'esterno della rete dei LNF.

Questa evenienza, purtroppo, si verifica soprattutto a causa di Virus o Malware installati su PC portatili configurati in ambiente Microsoft Windows.

### 6.5.1 Difesa perimetrale

Come già descritto nel paragrafo 2.5.12, la via di accesso alla rete geografica per la LAN dei LNF è una sola, per cui è sufficiente un *firewall* posto su tale via per ottenere un buon controllo perimetrale di tutto il traffico che passa dall'esterno verso la LAN e viceversa.

Il *firewall* è un apparato di rete hardware o software che filtra tutti i pacchetti entranti ed uscenti, da e verso una rete o un computer, applicando regole che contribuiscono alla sicurezza informatica della LAN.

La funzionalità di firewall ai LNF è svolta dal router di frontiera che implementa una soluzione di *packet filtering* basata su ACL (*Access Control List*). Le ACL del router sono in grado di filtrare pacchetti sulla base delle informazioni contenute negli *header* dei livelli 2, 3 e 4 della pila del protocollo TCP/IP. Rispetto ad un firewall vero e proprio hanno il limite di non ispezionare il *payload*, ovvero il contenuto dei dati che vengono scambiati. Tuttavia questa limitazione non rappresenta un ostacolo alla funzionalità, in quanto ai LNF non vi è necessità di effettuare il packet filtering basato sul contenuto dei dati scambiati.

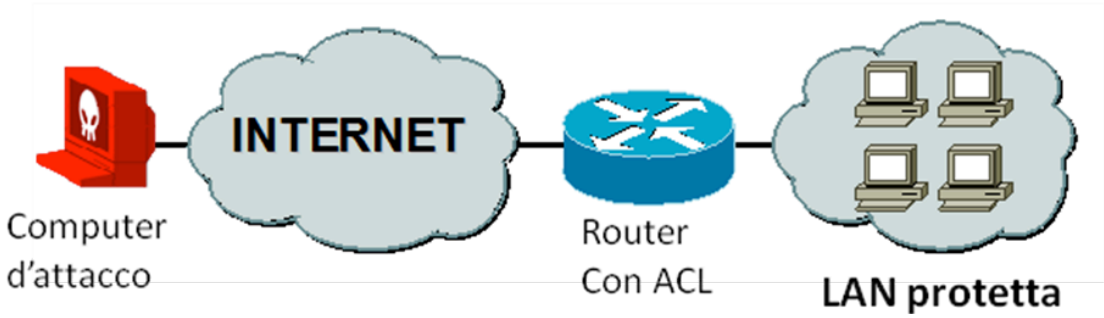



**Figura 6-1: Funzionalità di firewall ai LNF**

In particolare i *router* Cisco mettono a disposizione più tipi di ACL. Ad esempio relativamente al protocollo IP si hanno le *IP standard access list* e le *IP extended access list*. La differenza consiste nel fatto che, utilizzando le *standard ACL*, si possono controllare solamente l'indirizzo IP sorgente e l'indirizzo IP destinatario; mentre, con le *extended ACL*, si possono controllare molti più campi del pacchetto TCP/IP, più precisamente:

- Indirizzo IP sorgente
- Indirizzo IP di destinazione
- Campo protocollo, nell'*header* del livello rete (*Network layer*)
- Numero di porta sorgente (TCP o UDP) nell'*header* del livello trasporto (*Transport layer)*

- Numero di porta di destinazione (TCP o UDP) nell'*header* del livello trasporto (*Transport layer)*

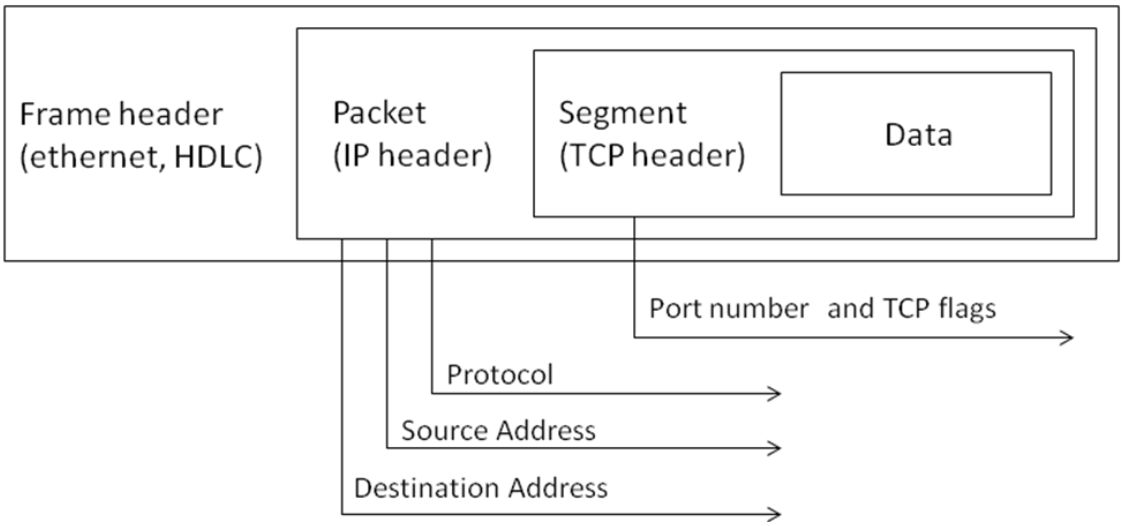



**Figura 6-2: Struttura di un pacchetto TCP/IP**

Le *access list* sono una sequenza di regole di controllo d'accesso che vengono applicate al traffico in transito sull'interfaccia dove l'access list è impostata. Le regole vengono verificate nella sequenza con cui sono scritte e, per ogni pacchetto, al primo *match* l'access list viene abbandonata, ovvero il controllo delle regole successive non viene effettuato. Anche se i controlli vengono effettuati tramite processori dedicati, particolarmente efficienti ed idonei allo scopo (ASIC), il router impiega risorse (processore e tempo) per i vari controlli sulle regole. Maggiore è il numero delle regole in un'access list e maggiore è il tempo di latenza del pacchetto nel router. Per questo motivo conviene costruire le access list mettendo le regole con più *match* all'inizio della sequenza.

Una volta creata una *access list* essa può essere applicata su un'interfaccia del router sia in ingresso che in uscita. Ci sono delle linee guida che dovrebbero essere seguite quando si creano e implementano le *access list* su un router:

- Si può assegnare solo un'*access list* per ogni interfaccia, o direzione. Questo significa se ne può avere solo una in ingresso e una in uscita per ogni interfaccia.
- Si dovrebbe organizzare in modo tale che i test più specifici siano all'inizio della lista.
- A meno che la lista non finisca con il comando *permit all* tutti i pacchetti saranno scartati se non sono compatibili con almeno una regola. Ogni lista dovrebbe avere almeno un comando *permit* altrimenti è come spegnere l'interfaccia (*deny all* è il default).
- Le *access list* sono progettate per filtrare il traffico attraverso il *router*. Esse non filtrano il traffico generato dal *router*.

#### 6.5.1.1 Politiche di sicurezza: packet filtering in uscita

È stata impostata una *ACL standard* sulla porta interna del router di frontiera nel verso di uscita (rif. Figura 6-3).

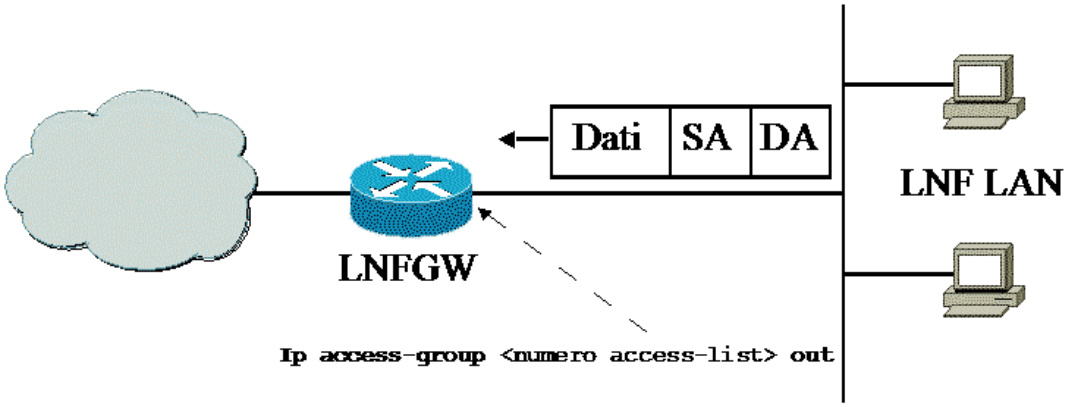



**Figura 6-3: ACL in uscita sul router d'accesso**

Questa ACL serve ad imporre che dalla LAN escano esclusivamente pacchetti con i indirizzi IP delle reti pubbliche appartenenti alla LAN. Questa, che potrebbe sembrare una misura di sicurezza superflua, è in realtà necessaria per evitare il fenomeno dell'*IP Spoofing*. Questo fenomeno si ha quando un hacker prende il controllo di una macchina interna alla LAN, e genera traffico con pacchetti aventi un indirizzo IP sorgente modificato. Impostando questa ACL eventuali pacchetti "contraffati" vengono bloccati dal router impedendone così l'uscita dalla LAN (e quindi prevenendo l'*IP spoofing* dall'interno verso l'esterno).

#### 6.5.1.2 Politiche di sicurezza: packet filtering in entrata

Il controllo del traffico in ingresso è più complesso del controllo del traffico in uscita. La politica di base impostata è di tipo *deny all but*, ovvero i filtri impostati vietano l'accesso a tutti i pacchetti IP tranne a quelli specificati nelle regole dell'ACL in ingresso.

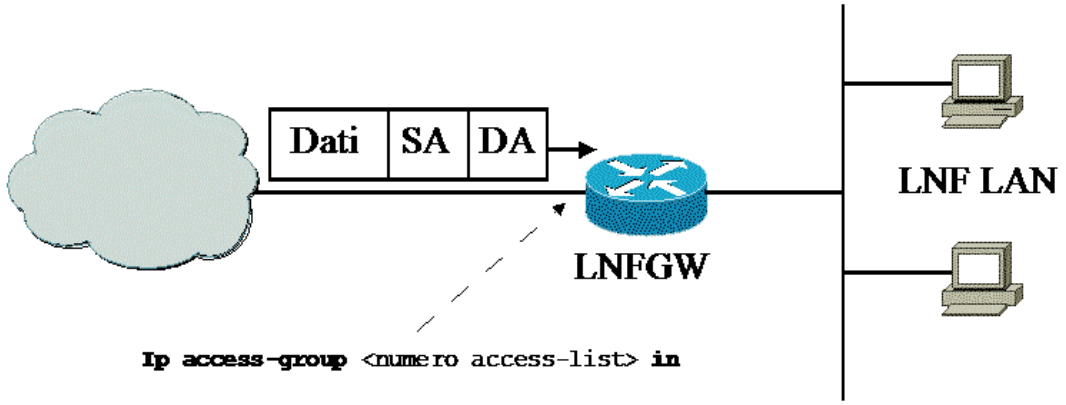



**Figura 6-4: ACL in entrata sul router d'accesso**

Questo tipo di filtro impedirebbe anche il traffico di ritorno per le connessioni aperte dall'interno della LAN verso l'esterno. Per ovviare a questo inconveniente le prime regole impostate nell'ACL d'ingresso prevedono di accettare il traffico relativo alle connessioni già stabilite. Quindi il traffico IP derivante da connessioni instaurate a partire dall'interno della LAN è consentito. In questo modo gli utenti interni che aprono connessioni TCP/IP verso l'esterno della LAN, non risentono dei filtri impostati in ingresso sul router di frontiera.

Le altre regole dell'ACL permettono alcune connessioni dall'esterno verso l'interno, ed in particolare verso server gestiti dal Servizio di Calcolo, o comunque sotto il controllo del Servizio. Senza entrare nel dettaglio, si fornisco alcune regole generiche a titolo di esempio:

- Accesso UDP ai DNS Server (porta 53)
- Accesso UDP ai server di autenticazione Kerberos (porte 88 e 750)
- Accesso UDP ai server NTP (porta 123)
- Accesso TCP ai server e ai client AFS (porte da 7000 a 7009)
- Accesso TCP ai Web server (protocollo http, porte 80 e 443)
- Accesso TCP agli SMTP server (porte 25, 587 e 465)
- Accesso TCP agli IMAP server (porta 993 protocollo cifrato SSL)
- Accesso TCP al bastion host per il login interattivo (protocollo SSH, porta 22)
- Accesso TCP ai Computing element e agli storage element della griglia computazionale (varie porte)
- Accesso TCP/UDP al VPN Concentrator dei LNF per il protocollo IPSEC (varie porte); rif. paragrafo 6.5.3.

#### 6.5.1.3 Difesa perimetrale dai malware e dai mail spam

Il servizio di posta elettronica è uno dei veicoli più utilizzati dagli hacker per diffondere malware o inviare spam. Occorre mettere in evidenza che il traffico mail dei Laboratori Nazionali di Frascati, sia in ingresso che in uscita, passa necessariamente per i mail server gestiti dal Servizio di Calcolo, grazie ad apposite ACL impostate sul router di frontiera. Per arginare questo tipo di attacchi è quindi sufficiente impostare le opportune protezioni sui mail server. Nel paragrafo 4.3.2.1, nella sezione dedicata al milter, è già stato trattato l'argomento relativo alla difesa perimetrale da questa tipologia di attacchi.

### 6.5.2 Difesa sui nodi della rete

Come già accennato, con i sistemi di difesa perimetrale, si può raggiungere un discreto livello di sicurezza in relazione alle tipologie di attacco che provengono dall'esterno della LAN. Queste difese tuttavia sono assolutamente inefficaci per le tipologie di attacchi che provengono dall'interno.

Per questo motivo anche i singoli sistemi devono implementare degli adeguati livelli di protezione. Le protezioni sui singoli sistemi sono diversificate in funzione del sistema operativo e delle funzioni che svolgono.

#### 6.5.2.1 Server basati su Scientific Linux

La maggior parte dei servizi gestiti dal Servizio di Calcolo è basata su sistema operativo Scientific Linux 5. La protezione di questo tipo di server viene realizzata attraverso due strumenti diversi.

È fondamentale infatti mantenere la versione del sistema operativo, e dei servizi installati al contorno, al livello massimo di *patch* possibile. Chi fornisce le distribuzioni di s/w opensource, in genere si occupa anche di fornire gli update a release (o subrelease) successive, al fine di migliorare le funzionalità, ma anche di eliminare errori s/w o buchi di sicurezza. Questa operazione viene effettuata tutte le notti tramite *update* automatici.

Il secondo strumento è fornito da un modulo del kernel di Linux (*netfilter*) e svolge le funzioni di un vero e proprio firewall sul nodo. Sui server dei LNF viene impostato in modo da rifiutare tutti i pacchetti tranne quelli relativi al servizio che il server implementa. Ad esempio, nel caso dei server web, il modulo netfilter viene configurato in modo tale da accettare tutte le connessioni TCP sulla porte 80 e 443 (protocollo http e https) provenienti da qualunque nodo di internet, e le connessioni TCP sulla porta 22 (protocollo SSH) provenienti esclusivamente dall'interno della LAN per il login interattivo dei sistemisti. Inoltre il modulo netfilter viene configurato anche per proteggersi da attacchi di tipo *Denyal of Service* o *Distributed Denyal of Service*, respingendo *ping flood* e *syn flood*.

#### 6.5.2.2 Server e client basati su Microsoft Windows

La politica adottata sui server e sui client Windows è uguale a quella adottata per i server Linux, con la fondamentale ulteriore attenzione al problema dei virus e dei malware in genere.

Anche in questo caso si operano update automatici giornalieri del sistema operativo e dell'*office* forniti dalla Microsoft; vengono anche impostati gli update automatici degli applicativi al contorno, quali s/w di terze parti, s/w freeware e opensource eventualmente installati sui sistemi.

Inoltre, anche nei sistemi Windows si può attivare un Firewall nativo, con le stesse politiche adottate sui server Linux. Ad esempio per quanto riguarda il printer server, vengono accettate esclusivamente le connessioni TCP sulle porte utilizzate dal servizio (tipicamente 139 netbios, 515 LPR/LPD, 9100 *Standard TCP/IP Port Protocol*). Tutti gli altri pacchetti vengono scartati.

Infine i sistemi Windows vengono protetti da opportuno s/w antivirus. L'INFN ha sottoscritto un contratto con la Sophos che fornisce un antivirus efficace, costantemente aggiornato. Sui server Windows viene installato e configurato Sophos antivirus, che prevede update automatici delle impronte virali ogni quattro ore, mantenendo i sistemi ad un buon livello di protezione.

### 6.5.3 Accesso dall'esterno verso i nodi interni

Le politiche di sicurezza impostate, tramite le ACL sul router di frontiera e la configurazione dei servizi, impongono che le connessioni in ingresso siano consentite esclusivamente verso sistemi e servizi gestiti e controllati dal Servizio di Calcolo. Inoltre impongono che la comunicazione avvenga tramite protocolli che prevedono la cifratura dei dati, soprattutto per i dati più critici o sensibili tra i quali, ad esempio, le credenziali per l'autenticazione.

Tali politiche tuttavia, pur aumentando il livello di sicurezza complessivo, spesso impediscono o rendono difficoltoso l'accesso ai servizi specifici dei gruppi sperimentali, o più in generale degli utenti.

Ad esempio, il login interattivo è possibile solamente su un *bastion host* gestito dal Servizio di Calcolo attraverso il protocollo cifrato SSH. Dal bastion host è possibile utilizzare il comando *ssh* per effettuare un nuovo login interattivo verso un altro host qualunque della LAN. Il protocollo SSH è in grado anche di trasportare, in un tunnel cifrato, il protocollo di gestione grafica X11 (ed anche altre comunicazioni), indipendentemente dal numero dei salti effettuati per arrivare al nodo target. Spesso questa modalità di accesso è sufficiente per arrivare ai servizi più interni della LAN.

I filtri impostati permettono inoltre agli utenti esterni di accedere ai servizi web centrali, di leggere e inviare la posta, di accedere alle risorse di storage AFS.

Tuttavia a volte gli utenti hanno necessità di accedere ad altri servizi, eludendo i filtri impostati sul router di frontiera. A tale scopo ai LNF è stato installato un VPN Concentrator della Cisco.

#### 6.5.3.1 Accesso tramite VPN

Si utilizza la tecnologia delle Virtual Private Network per connettere in sicurezza uffici distaccati o utenti remoti, in modo da fornire le stesse modalità e gli stessi diritti di accesso ai nodi interni che si avrebbero stando direttamente connessi sulla rete locale.

Vi sono infatti due usi principali delle VPN: il primo consente di connettere due o più reti geografiche distinte, come ad esempio la sede centrale e le filiali, per farle comunicare attraverso un'unica rete privata virtuale; il secondo consente ad utenti autorizzati di accedere alla rete aziendale dal proprio PC, connesso a internet remotamente.

Per il primo tipo di connessione si utilizzano due o più gateway in grado di cifrare e decifrare tutto il traffico di dati scambiati tra le varie sedi. Per il secondo tipo di connessione serve un gateway centrale (che fa da VPN Server) e un software (VPN client) da installare sul PC.

Il compito dei gateway e del software è quello di cifrare i dati inoltrati e decifrare quelli ricevuti. Per poter essere protetti, i dati scambiati sulla rete VPN devono essere incapsulati tramite un processo chiamato *tunnelling*, che ha lo scopo di

collocare i dati in buste digitalizzate. Il tunnel è il *link* virtuale che si instaura tra client e server, visto dal client come una connessione diretta al gateway (dal quale ottiene un nuovo indirizzo IP tra quelli a disposizione per la LAN), nel quale viaggia tutto il traffico opportunamente cifrato e talvolta anche compresso.

Per la cifratura si usa lo standard *IPsec* (*IP Security*), che è una suite di protocolli che garantiscono la sicurezza. IPsec è composto da 3 protocolli principali:

- *IKE* (*Internet Key Exchange*): gestisce lo scambio delle chiavi di cifratura
- *AH* (*Authentication Header*): fornisce servizi di autenticazione ed integrità
- *ESP* (*Encapsulating Security Payload*): fornisce servizi di riservatezza tramite la cifratura

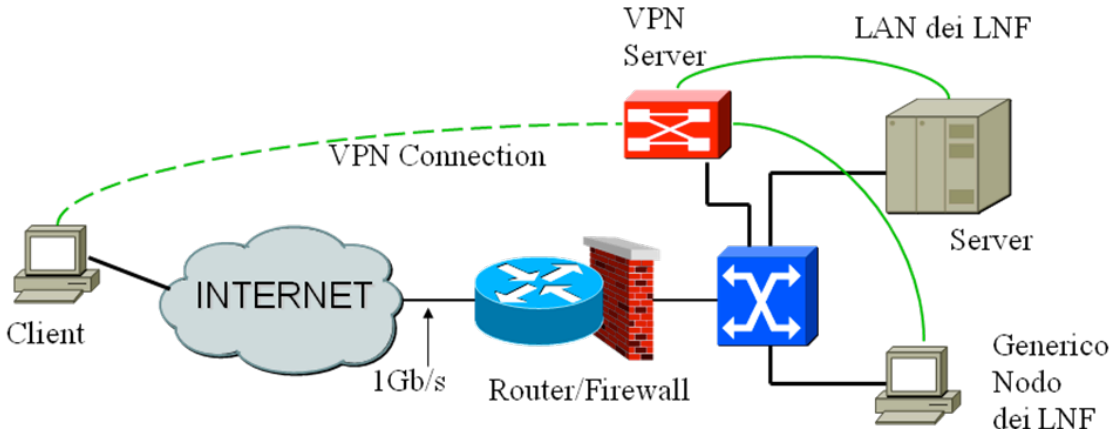



**Figura 6-5: Connessione ai LNF tramite VPN**

Ai LNF la soluzione VPN è implementata attraverso un VPN Concentrator Cisco 3030. Tale soluzione fornisce il software client per moltissime piattaforme (Windows, Mac OS X, Linux, etc.). Gli utenti che hanno necessità di connettersi tramite VPN devono preventivamente installare il software VPN client sul proprio PC.

Operativamente l'utente può stabilire un tunnel cifrato con il VPN server semplicemente eseguendo l'applicazione client ed effettuando l'autenticazione. Il VPN server dei LNF è configurato per accettare autenticazioni tramite certificato Digitale X.509 rilasciato dalla Certification Authority dell'INFN, oppure tramite inserimento delle proprie credenziali d'accesso. In questo secondo caso l'autenticazione verrà validata dal server RADIUS, che a sua volta la demanderà al sistema Kerberos 5.

Una volta autenticato, il client verrà proiettato sulla LAN e potrà accedere a tutti i servizi della rete, come se fosse direttamente connesso all'interno della rete locale dei LNF.

### 6.5.4 Log degli eventi

Il *logbook* degli eventi è una sorta di giornale di bordo, o semplicemente giornale, su cui vengono registrati gli eventi significativi di un sistema o di un servizio, in ordine cronologico.

Il *log* più semplice è un file sequenziale sempre aperto in scrittura, che viene chiuso e conservato a cadenze regolari e reso disponibile per:

- analisi delle segnalazioni di errore
- produzione di statistiche di esercizio, come ad esempio quelle del traffico nei servizi web
- ripristino di situazioni precedenti
- analisi delle modifiche fatte nella base dati
- analisi delle operazioni fatte e dei responsabili di tali operazioni.

Ai fini della sicurezza informatica è fondamentale raccogliere gli *event log* per utilizzarli come strumento di analisi di ciò che avviene o che è avvenuto su un sistema o su un servizio. In effetti dall'analisi dei log si può verificare un eventuale uso improprio degli strumenti informatici e risalire spesso all'autore o al nodo di provenienza. La legislazione corrente prevede che i log vengano ben conservati per un periodo di almeno 6 mesi.

Un *hacker* esperto tende a non lasciare traccia delle operazioni effettuate e quasi sempre, a lavoro ultimato, distrugge i file di log del sistema violato.

Per ovviare a questo problema ai LNF si utilizza il protocollo *syslog* (*System Log*). Syslog è un protocollo appartenente alla Suite di protocolli Internet utilizzato per trasmettere attraverso la rete tutte le informazioni di log.

Tutti i sistemi gestiti centralmente sono configurati per registrare i log su file locali e contemporaneamente per inviarli via rete ad un *Syslog Server* centrale. La semplicità del protocollo fa si che il server possa gestire messaggi provenienti da una variegata tipologia di macchine, da computer, stampanti, dispositivi di rete, etc.. Il server può limitarsi a registrare l'evento, per avere un archivio centrale degli avvenimenti, oppure reagire a particolari livelli di severità chiamando programmi, inviando e-mail di allarme, etc..

Il Syslog Server dei LNF è configurato con il software opensource *syslog-ng* che è in grado di distinguere il nodo o il servizio di provenienza dell'evento e classificare i log su file diversi e specifici.

Il Servizio di Calcolo ha configurato il Syslog Server in modo tale da salvare un file specifico per gli eventi di sistema o per gli eventi generici di ogni host o apparato di rete. Però salva un file unico per servizio, anche se il servizio è svolto in load balancing da più host (come il mailing, il web, etc.).

Ogni notte i file di log vengono chiusi, compressi e archiviati su libreria a nastri tramite Tivoli Storage Manager a tempo indeterminato.

### 6.5.5 Monitoring

È fondamentale per chi gestisce servizi informatici disporre di uno strumento di monitoring che tenga sotto controllo tutti i sistemi, i servizi e gli apparati di rete, allo scopo di elevare il livello di affidabilità e di disponibilità dei servizi stessi.

Come già accennato in occasione delle griglie computazionali, uno strumento molto utilizzato per il monitoring è Nagios. Nagios è un software open-source per il monitoraggio ed il controllo costante di Server e di Servizi molto usato nel mondo Linux. È in grado di eseguire controlli su un'ampia serie di servizi quali HTTP, FTP, SSH, Numero di Processi attivi, Carico del Server, Numero di Utenti collegati, Risposta del Server ai Ping, Controllo dei DNS, Controllo del Demone di MySql, fare query con protocollo SNMP, e molto altro ancora.

Nagios fornisce l'output su web, e ciò consente con un colpo d'occhio di tenere sotto controllo la funzionalità di tutti gli apparati informatici.

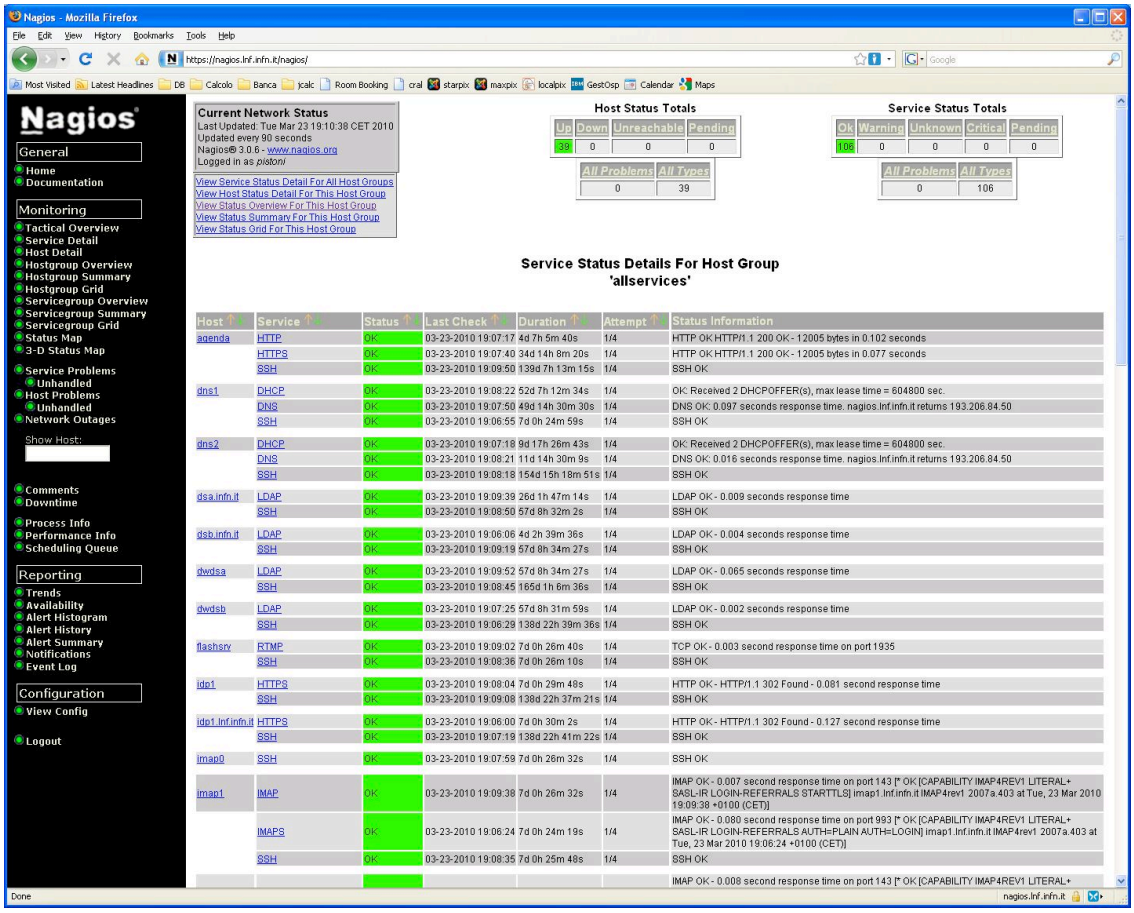
Il Servizio di Calcolo ha configurato un server Nagios per monitorare tutti gli apparati della rete locale e geografica e le relative interfacce di connessione, tutti gli apparati di storage, tutti i sistemi e tutti i servizi gestiti centralmente. Nagios infatti fornisce una serie di plugin per il monitoring di molti servizi tramite interrogazioni di funzionalità via rete; inoltre il Calcolo ha realizzato altri plugin per il monitoring di servizi specifici non forniti con il kit d'installazione base (ad esempio per conoscere lo stato dei file server AFS, o delle stampanti dipartimentali).

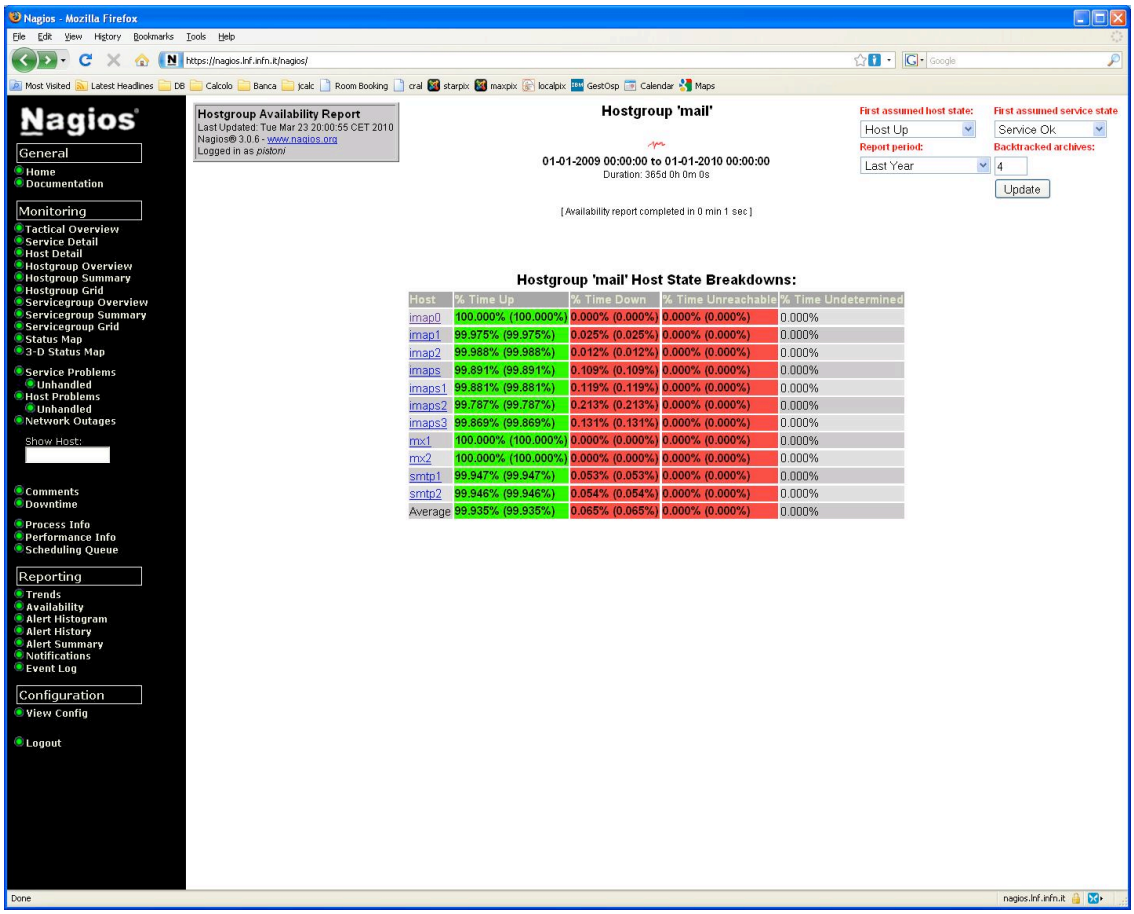
Nagios ha inoltre la peculiarità di inviare un'e-mail per ogni evento critico, quali possono essere ad esempio il fallimento e/o il ripristino di linee o apparati di rete, di sistemi o di servizi. Grazie a questa caratteristica il Servizio di Calcolo è immediatamente allarmato all'incorrere dell'evento e può avviare subito le procedure di ripristino, a volte prima ancora che gli utenti possano accorgersi del disservizio.

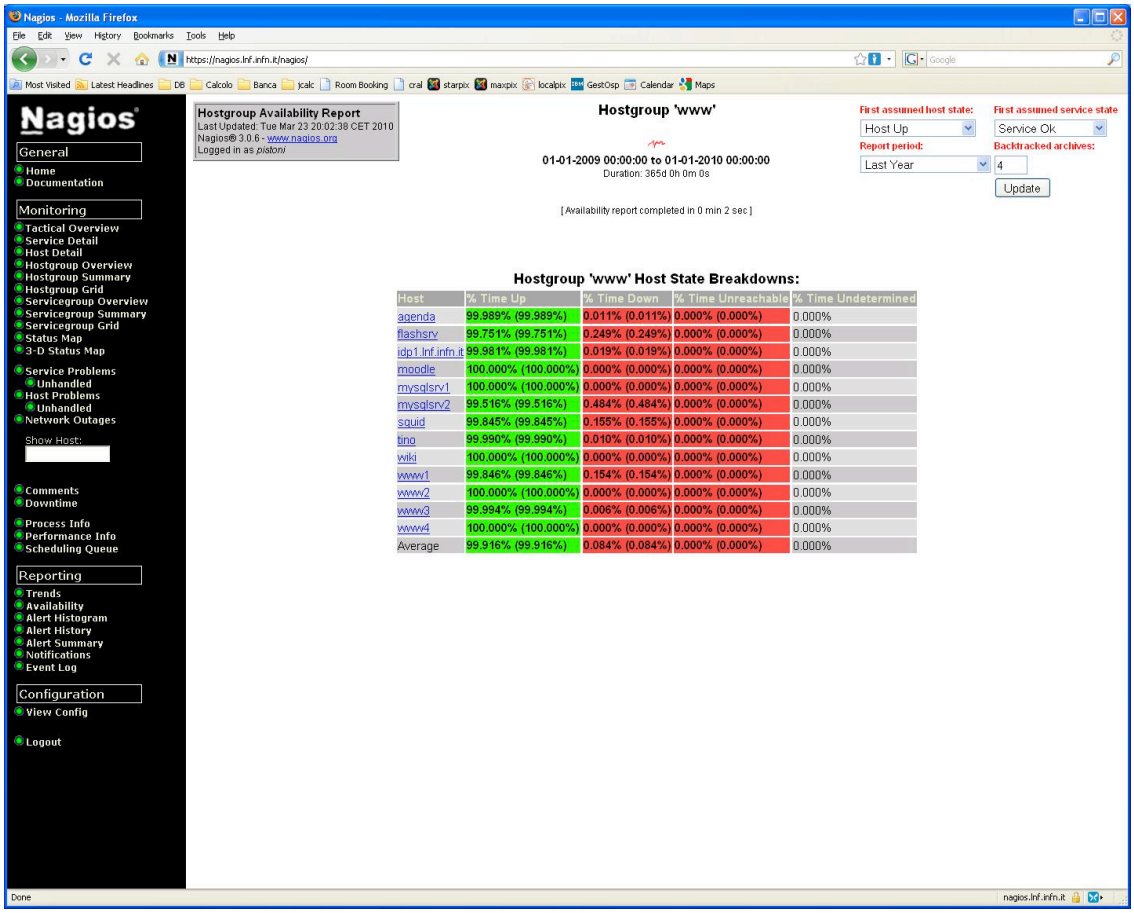
Ovviamente lo strumento del mail ha anch'esso le sue criticità. I sistemisti non verrebbero avvertiti nel caso in cui a fallire sia proprio il servizio di mail o un servizio infrastrutturale al servizio di mail (rete locale, storage area network, Linux Cluster, etc.). Tuttavia questa tipologia di eventi è una delle più gravi che possano avvenire e generalmente producono una serie di inconvenienti tali, da essere comunque immediatamente rilevata dal personale sistemista del Servizio. Inoltre tali eventi sono, fortunatamente, estremamente rari.

Infine Nagios è in grado di salvare gli eventi in un suo database interno, e di fornire la storia della disponibilità di un determinato nodo o servizio attraverso dei report testuali o attraverso grafici di immediata lettura. Grazie all'analisi di tali dati si possono estrapolare informazioni sull'affidabilità complessiva dei servizi informatici.

A titolo di esempio, seguono alcune visualizzazioni del s/w Nagios installato ai LNF. La prima visualizza lo stato di tutti i servizi. La seconda e la terza visualizzano la percentuale di disponibilità (*availability*) dei due gruppi di host che forniscono rispettivamente i servizi mail e web.


Nagios - Mozilla Firefox
https://nagios.lnf.infn.it/nagios/
Nagios
General
Home
Documentation
Monitoring
Tactical Overview
Service Detail
Host Detail
Hostgroup Overview
Hostgroup Summary
Hostgroup Grid
Servicegroup Overview
Servicegroup Summary
Servicegroup Grid
Status Map
3-D Status Map
Service Problems
Unhandled
Host Problems
Unhandled
Network Outages
Show Host:
Comments
Downtime
Process Info
Performance Info
Scheduling Queue
Reporting
Trends
Availability
Alert Histogram
Alert History
Alert Summary
Notifications
Event Log
Configuration
View Config
Logout
Current Network Status
Last Updated: Tue Mar 23 19:10:38 CET 2010
Updated every 90 seconds
Nagios® 3.0.6 - www.nagios.org
Logged in as pistoni
View Service Status Detail For All Host Groups
View Host Status Detail For This Host Group
View Status Overview For This Host Group
View Status Summary For This Host Group
View Status Grid For This Host Group
Host Status Totals
Up Down Unreachable Pending
39 0 0 0
All Problems All Types
0 39
Service Status Totals
Ok Warning Unknown Critical Pending
106 0 0 0 0
All Problems All Types
0 106
Service Status Details For Host Group 'allservices'
Host Service Status Last Check Duration Attempt Status Information
agenda HTTP OK 03-23-2010 19:07:17 4d 7h 5m 40s 1/4 HTTP OK HTTP/1.1 200 OK - 12005 bytes in 0.102 seconds
HTTPS OK 03-23-2010 19:07:40 34d 14h 8m 20s 1/4 HTTP OK HTTP/1.1 200 OK - 12005 bytes in 0.077 seconds
SSH OK 03-23-2010 19:09:50 139d 7h 13m 15s 1/4 SSH OK
dns1 DHCP OK 03-23-2010 19:08:22 52d 7h 12m 34s 1/4 OK: Received 2 DHCPOFFER(s), max lease time = 604800 sec.
DNS OK 03-23-2010 19:07:50 49d 14h 30m 30s 1/4 DNS OK: 0.097 seconds response time. nagios.lnf.infn.it returns 193.206.84.50
SSH OK 03-23-2010 19:06:55 7d 0h 24m 59s 1/4 SSH OK
dns2 DHCP OK 03-23-2010 19:07:18 9d 17h 26m 43s 1/4 OK: Received 2 DHCPOFFER(s), max lease time = 604800 sec.
DNS OK 03-23-2010 19:08:21 11d 14h 30m 9s 1/4 DNS OK: 0.016 seconds response time. nagios.lnf.infn.it returns 193.206.84.50
SSH OK 03-23-2010 19:08:18 154d 15h 18m 51s 1/4 SSH OK
dsa.infn.it LDAP OK 03-23-2010 19:09:39 26d 1h 47m 14s 1/4 LDAP OK - 0.009 seconds response time
SSH OK 03-23-2010 19:08:50 57d 8h 32m 2s 1/4 SSH OK
dsb.infn.it LDAP OK 03-23-2010 19:06:06 4d 2h 39m 36s 1/4 LDAP OK - 0.004 seconds response time
SSH OK 03-23-2010 19:09:19 57d 8h 34m 27s 1/4 SSH OK
dwdsa LDAP OK 03-23-2010 19:09:52 57d 8h 34m 27s 1/4 LDAP OK - 0.065 seconds response time
SSH OK 03-23-2010 19:08:45 165d 1h 6m 36s 1/4 SSH OK
dwdsb LDAP OK 03-23-2010 19:07:25 57d 8h 31m 59s 1/4 LDAP OK - 0.002 seconds response time
SSH OK 03-23-2010 19:06:29 138d 22h 39m 36s 1/4 SSH OK
flashsrv RTMP OK 03-23-2010 19:09:02 7d 0h 26m 40s 1/4 TCP OK - 0.003 second response time on port 1935
SSH OK 03-23-2010 19:08:36 7d 0h 26m 10s 1/4 SSH OK
idp1 HTTPS OK 03-23-2010 19:08:04 7d 0h 29m 48s 1/4 HTTP OK - HTTP/1.1 302 Found - 0.081 second response time
SSH OK 03-23-2010 19:09:08 138d 22h 37m 21s 1/4 SSH OK
idp1.lnf.infn.it HTTPS OK 03-23-2010 19:06:00 7d 0h 30m 2s 1/4 HTTP OK - HTTP/1.1 302 Found - 0.127 second response time
SSH OK 03-23-2010 19:07:19 138d 22h 41m 22s 1/4 SSH OK
imap0 SSH OK 03-23-2010 19:07:59 7d 0h 26m 32s 1/4 SSH OK
imap1 IMAP OK 03-23-2010 19:09:38 7d 0h 26m 32s 1/4 IMAP OK - 0.007 second response time on port 143 [* OK [CAPABILITY IMAP4REV1 LITERAL+ SASL-IR LOGIN-REFERRALS STARTTLS] imap1.lnf.infn.it IMAP4rev1 2007a.403 at Tue, 23 Mar 2010 19:09:38 +0100 (CET)]
IMAPS OK 03-23-2010 19:06:24 7d 0h 24m 19s 1/4 IMAP OK - 0.080 second response time on port 993 [* OK [CAPABILITY IMAP4REV1 LITERAL+ SASL-IR LOGIN-REFERRALS AUTH=PLAIN AUTH=LOGIN] imap1.lnf.infn.it IMAP4rev1 2007a.403 at Tue, 23 Mar 2010 19:06:24 +0100 (CET)]
SSH OK 03-23-2010 19:08:35 7d 0h 25m 48s 1/4 SSH OK
IMAP OK - 0.008 second response time on port 143 [* OK [CAPABILITY IMAP4REV1 LITERAL+
Done
nagios.lnf.infn.it


Nagios - Mozilla Firefox
https://nagios.lnf.infn.it/nagios/
Nagios
General
Home
Documentation
Monitoring
Tactical Overview
Service Detail
Host Detail
Hostgroup Overview
Hostgroup Summary
Hostgroup Grid
Servicegroup Overview
Servicegroup Summary
Servicegroup Grid
Status Map
3-D Status Map
Service Problems
Unhandled
Host Problems
Unhandled
Network Outages
Show Host:
Comments
Downtime
Process Info
Performance Info
Scheduling Queue
Reporting
Trends
Availability
Alert Histogram
Alert History
Alert Summary
Notifications
Event Log
Configuration
View Config
Logout
Hostgroup Availability Report
Last Updated: Tue Mar 23 20:00:55 CET 2010
Nagios® 3.0.6 - www.nagios.org
Logged in as pistoni
Hostgroup 'mail'
01-01-2009 00:00:00 to 01-01-2010 00:00:00
Duration: 365d 0h 0m 0s
[ Availability report completed in 0 min 1 sec ]
First assumed host state:
Host Up
First assumed service state
Service Ok
Report period:
Last Year
Backtracked archives:
4
Update
Hostgroup 'mail' Host State Breakdowns:
Host % Time Up % Time Down % Time Unreachable % Time Undetermined
imap0 100.000% (100.000%) 0.000% (0.000%) 0.000% (0.000%) 0.000%
imap1 99.975% (99.975%) 0.025% (0.025%) 0.000% (0.000%) 0.000%
imap2 99.988% (99.988%) 0.012% (0.012%) 0.000% (0.000%) 0.000%
imaps 99.891% (99.891%) 0.109% (0.109%) 0.000% (0.000%) 0.000%
imaps1 99.881% (99.881%) 0.119% (0.119%) 0.000% (0.000%) 0.000%
imaps2 99.787% (99.787%) 0.213% (0.213%) 0.000% (0.000%) 0.000%
imaps3 99.869% (99.869%) 0.131% (0.131%) 0.000% (0.000%) 0.000%
mx1 100.000% (100.000%) 0.000% (0.000%) 0.000% (0.000%) 0.000%
mx2 100.000% (100.000%) 0.000% (0.000%) 0.000% (0.000%) 0.000%
smtp1 99.947% (99.947%) 0.053% (0.053%) 0.000% (0.000%) 0.000%
smtp2 99.946% (99.946%) 0.054% (0.054%) 0.000% (0.000%) 0.000%
Average 99.935% (99.935%) 0.065% (0.065%) 0.000% (0.000%) 0.000%
Done
nagios.lnf.infn.it




### 6.5.6 Conclusioni

Gli obiettivi principali del Servizio di Calcolo dei Laboratori Nazionali di Frascati nella predisposizione dei servizi informatici forniti all'utenza sono:

- Semplicità d'uso da parte dell'utenza
- Alta affidabilità
- Alta disponibilità

Si ritiene che tutti gli obiettivi siano stati raggiunti. Il primo, grazie alle scelte tecnologiche e implementative, che permettono a tutta l'utenza di fruire semplicemente di tutti i servizi erogati.

Gli ultimi due obiettivi, si ritiene che siano stati largamente raggiunti grazie ai diffusi sistemi di ridondanza applicati a tutti i livelli architetturali.

Statisticamente i servizi più critici hanno un tempo di downtime annuale (programmato e non) che varia da qualche minuto a qualche ora. Ovvero in termini percentuali il disservizio più lungo sugli host che erogano i servizi fondamentali è dell'ordine di qualche unità "per mille" ovvero la disponibilità minima di tali host (presi singolarmente), misurata nell'ultimo anno di attività, è stata circa del 99,8%. Tuttavia, grazie alle ridondanze, il servizio apparente all'utenza è stato anche superiore, ovvero dell'ordine del 99,9%.

# Appendice

## A.1 Cenni al modello standard ISO/OSI

Lo standard OSI tratta lo scambio di informazioni tra i sistemi, con particolare riferimento ai protocolli di comunicazione e il mezzo fisico tramite cui avviene tale scambio.

Tale trasferimento di informazioni avviene su mezzi fisici (physical media) secondo lo schema di Figura A-6-6 L'architettura del modello di riferimento OSI è stata progettata pensando a tre componenti principali:

- il processo applicativo che deve scambiare le informazioni;
- la connessione che permette lo scambio delle informazioni;
- i sistemi.

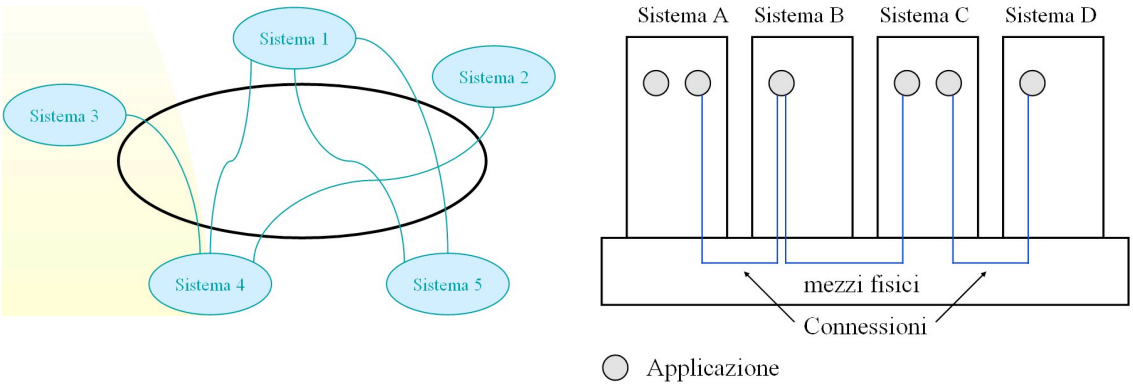



**Figura A-6-6: Sistemi interconnessi da mezzi fisici**

Per ridurre la complessità progettuale, OSI introduce un'architettura a livelli (*layered architecture*) i cui componenti principali sono:

- i livelli (*layers*);
- le entità (*entities*);
- i punti di accesso al servizio (*SAP: Service Access Points*);
- le connessioni (*connections*).

Livelli adiacenti comunicano tramite la loro interfaccia (*interface*). Ogni livello è poi composto da una o più entità. Entità appartenenti allo stesso livello, su sistemi diversi, vengono dette *peer-entities*.

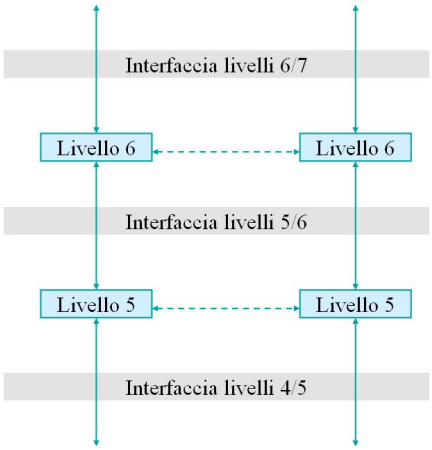
Tale approccio di progettazione a livelli è comune a tutte le moderne architetture di rete; ciò che varia dall'una all'altra è il numero dei livelli, il loro nome e le entità contenute.

Lo scopo di ciascun livello è quello di fornire servizi alle entità del livello superiore, mascherando il modo in cui questi sono implementati.

Ad eccezione del livello più alto, un livello N fornisce servizi di livello N alle entità di livello N+1.

Le entità di livello N, eccetto il livello 1, per comunicare usano servizi di livello N-1. Le entità di livello 1 comunicano direttamente tramite i mezzi trasmissivi che le interconnettono.

Livelli N comunicano attraverso un protocollo di livello N: ogni livello deve quindi mostrare un'interfaccia ben definita a quello immediatamente superiore.




Anche se è definito un protocollo di livello N, nessun dato è trasferito direttamente da un livello N all'altro; infatti ogni livello passa dati e informazioni di controllo a quello sottostante, sino a quando si giunge al livello Fisico, che effettua la trasmissione.

L'interfaccia definisce quali operazioni primitive e quali servizi sono forniti da un livello ai livelli superiori.

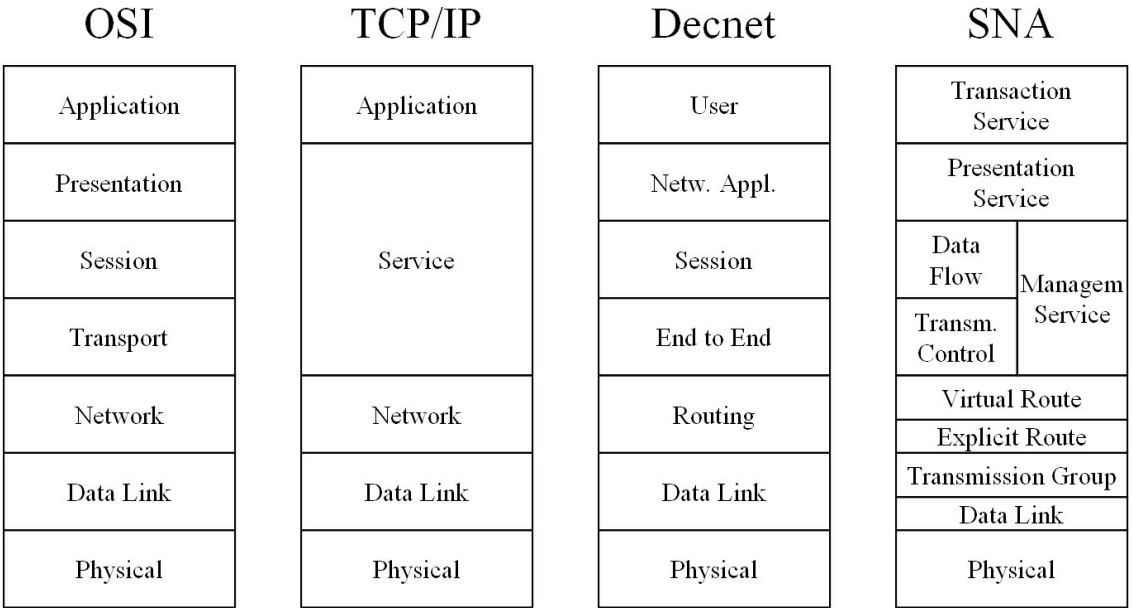



**Figura A-6-7: Principali architetture di rete**

Non sempre lo scambio di informazione avviene direttamente tra i due sistemi finali che contengono le applicazioni (ES: *End Systems*). Può anche implicare l'attraversamento di sistemi intermedi (IS: *Intermediate Systems*).

In essi esistono delle entità che assumono la funzionalità di *relaying*, cioè di inoltratrici di informazione.

Tali entità possono essere collocate a vari livelli del modello OSI e gli IS assumono nomi diversi in funzione del livello a cui avviene il *relaying*:

- *Repeater* (livello 1)
- *bridge* (livello 2)
- *router* (livello 3)
- *gateway* (livello 7)

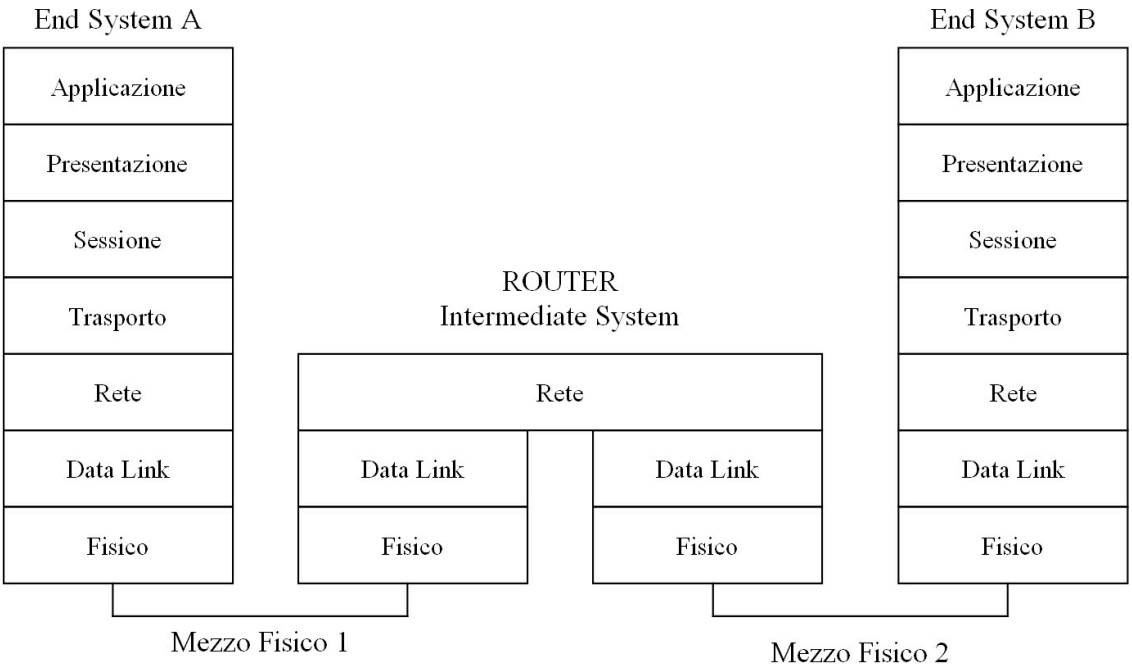



**Figura A-6-8: Sistemi intermedi**

Ogni livello N aggiunge ai dati ricevuti dal livello superiore alcune informazioni di controllo del protocollo N, dette comunemente "busta di livello N". Il tutto rappresenta i dati che verranno passati al livello inferiore che opererà in modo analogo.

I dati generati da un protocollo di livello N sono detti N–PDU (*Protocol Data Unit*).

Una volta attraversata l'interfaccia tra il livello N e il livello N-1, essi diventano una (N-1)-SDU (*Service Data Unit*).

La PDU di livello N-1 viene quindi costruita preponendo alla (N-1)-SDU una (N-1)-PCI (*Protocol Control Information*).

Scopo della PCI è quello di contenere le informazioni di controllo del protocollo.

Molto spesso al termine PDU vengono sostituiti quelli meno precisi, ma di uso comune, di pacchetto o trama.

Nell'ambito di un pacchetto il PCI rappresenta l'*header* del pacchetto stesso, già definito busta.

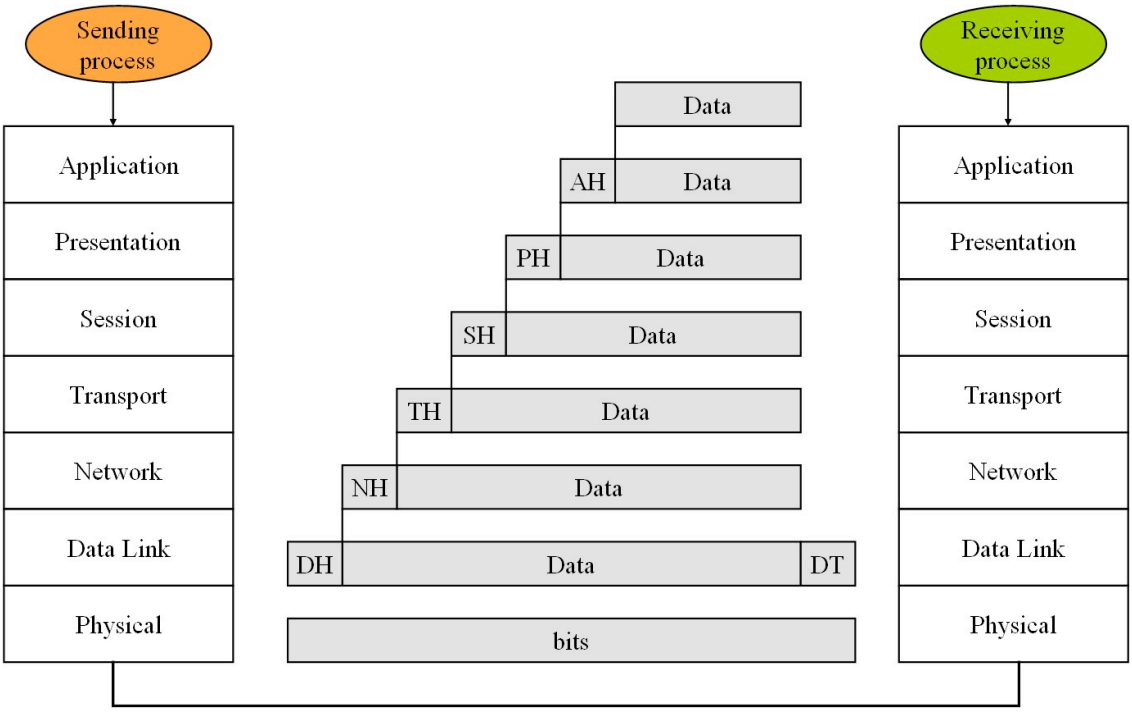



**Figura A-6-9: Imbustamento multiplo**

- **Livello 7: Applicazione**

Il livello 7 è il livello Applicazione, cioè dei programmi applicativi (facenti parte del sistema operativo o scritti dagli utenti) attraverso i quali l'utente finale utilizza la rete.

- **Livello 6: Presentazione**

Il livello 6 è il livello Presentazione, che gestisce la sintassi dell'informazione da trasferire (ad esempio codifica ASCII o EBCDIC); a questo livello sono previste tre diverse sintassi:

  - astratta (definizione formale dei dati che gli applicativi si scambiano)
  - concreta locale (come i dati sono rappresentati localmente)
  - di trasferimento (come i dati sono codificati durante il trasferimento).

- **Livello 5: Sessione**

Il livello 5 è il livello Sessione, responsabile dell'organizzazione del dialogo tra due programmi applicativi e del conseguente scambio di dati. Esso consente di aggiungere a connessioni *end-to-end* (cioè tra due entità collocate in ES) servizi più avanzati, quali la gestione del dialogo (mono o bidirezionale), la gestione del *token* (per effettuare mutua esclusione nell'utilizzo di una risorsa condivisa) o la sincronizzazione (inserendo dei *checkpoint* in modo da ridurre la quantità di dati da ritrasmettere in caso di gravi malfunzionamenti).

- **Livello 4: Trasporto**

Il livello 4 è il livello Trasporto, e fornisce trasferimento trasparente di informazione tra entità del livello sessione. In particolare, si occupa di fornire un trasferimento dati affidabile e di ottimizzare l'uso delle risorse di rete. Compiti del livello 4 saranno quindi tipicamente la frammentazione, la correzione degli errori e la prevenzione della congestione della rete. Il livello 4 è il più basso livello a trascurare la

topologia della rete e la presenza di sistemi intermedi (IS) e quindi è il primo livello detto *end-to-end*.

- **Livello 3: Rete**

Il livello 3 è il livello Network, che gestisce l'instradamento dei messaggi. Esso determina se e quali sistemi intermedi devono essere attraversati dal messaggio per giungere a destinazione. Quindi deve gestire delle tabelle di instradamento e provvedere ad instradamenti alternativi in caso di guasti (*fault tolerance*).

- **Livello 2: Data Link**

Il livello 2 è il livello Data Link, che ha come scopo la trasmissione sufficientemente affidabile di trame (*frame*). Accetta come *input* dei pacchetti di livello 3 (tipicamente poche centinaia di bit) e li trasmette sequenzialmente. Esso verifica la presenza di errori aggiungendo delle FCS (*Frame Control Sequence*) e può gestire meccanismi di correzione di tali errori tramite ritrasmissione.

- **Livello 1: Fisico**

Il livello 1 del modello OSI è il livello Fisico, che si occupa di trasmettere sequenze binarie sul canale di comunicazione. A questo livello si specificano, ad esempio, le tensioni che rappresentano 0 e 1 e le caratteristiche dei cavi e dei connettori.

Per tutti i livelli superiori al livello fisico sono definite due modalità operative:

- **connessa** (CONS: *Connection Oriented Network Service*)
- **non connessa** (CLNS: *ConnectionLess Network Service*)

Un dato livello può fornire al livello superiore servizi di tipo connesso, non-connesso o entrambi. Questa è una scelta progettuale che varia per ogni livello, da architettura ad architettura.

In un servizio **non connesso** la spedizione di un pacchetto è simile alla spedizione di una lettera ordinaria con il sistema postale. Tutto avviene in una sola fase lasciando cadere la lettera nella buca delle lettere. La lettera deve contenere sulla busta l'indirizzo completo del destinatario. Non vi è alcun riscontro diretto che la lettera giunga a destinazione correttamente.

In un servizio **connesso** lo scambio di dati tramite pacchetti ricorda le frasi scambiate tra due interlocutori al telefono. Vi sono tre momenti principali:

- creazione della connessione (il comporre il numero telefonico e il "pronto" alla risposta)
- trasferimento dei dati (la conversazione telefonica)
- chiusura della connessione (i saluti e il posare il microtelefono)

Nella modalità **connessa**, durante la fase di creazione della connessione (*initial setup*) due *peer-entities* concordano che trasferiranno delle PDU. Solo durante tale fase devono essere specificati gli indirizzi completi del mittente e del destinatario: successivamente le entità coinvolte specificheranno soltanto l'identificativo della connessione stabilito durante la prima fase.

Un servizio connesso fornisce una modalità di trasferimento delle PDU affidabile e sequenziale. Per tutta la durata della connessione le PDU inviate sono ricevute correttamente nello stesso ordine. Se qualcosa non funziona correttamente, la connessione può essere riavviata (*reset*) o terminata (*released*).

Per verificare che tutte le PDU inviate giungano a destinazione correttamente un servizio connesso utilizza degli schemi di numerazione dei pacchetti e di verifica dell'avvenuta corretta ricezione (ACK: *acknowledgement*). Quindi un protocollo connesso è in generale in grado non solo di rilevare la presenza di errori, ma anche di correggerli tramite ritrasmissioni.

Con una modalità **non connessa** la comunicazione ha luogo in una fase singola: il pacchetto è inviato e deve contenere l'indirizzo completo del destinatario. Non essendo i pacchetti organizzati in una connessione, un pacchetto non può fare riferimento ad altri pacchetti trasmessi precedentemente o in seguito. Quindi un protocollo non connesso può solo rilevare la presenza di errori (scartando quindi le PDU errate), ma non correggerli in quanto non si possono realizzare meccanismi di ritrasmissione (in un pacchetto non è possibile fare riferimento ad altri pacchetti).

Un protocollo non connesso è in generale più efficiente di un protocollo connesso, specialmente se bisogna trasferire piccole quantità di dati: in quest'ultimo caso infatti l'*overhead* della creazione e distruzione della connessione è rilevante. Un protocollo non connesso (detto anche *datagram*), non potendo garantire l'affidabilità del trasferimento dati, necessita che almeno un protocollo di livello superiore sia di tipo connesso.

| Caratteristica | Connection-Oriented | Connectionless |
|---|---|---|
| **Initial setup** | Richiesto | Impossibile |
| **Indirizzo di destinazione** | Durante il setup | In ogni pacchetto |
| **Ordine dei pacchetti** | Garantito | Non garantito |
| **Controllo degli errori** | Si | No |
| **Controllo di flusso** | Si | No |
| **Negoziazione di opzioni** | Si | No |
| **Identificatore di connessione** | Si | No |

**Tabella A-6-2: Differenze tra le modalità di connessione CONS e CNLS**

## *A.2 Cenni al modello standard IEEE 802*

Quando le prime LAN cominciarono a diffondersi (Ethernet, Token Ring, etc.), l'IEEE decise di costituire sei comitati per studiare il problema della standardizzazione delle LAN e delle MAN, complessivamente raccolti nel progetto IEEE 802. Tali comitati sono:

- 802.1 Overview, Architecture, Bridging and Management;
- 802.2 Logical Link Control;
- 802.3 CSMA/CD (Carrier Sense, Multiple Access with Collision Detection) ovvero Ethernet (10 Mbit/s);

- 802.4 Token Bus;
- 802.5 Token Ring;
- 802.6 Metropolitan Area Networks - DQDB (Distributed Queue, Dual Bus).

A tali comitati in seguito se ne sono aggiunti molti altri, tra cui alcuni dei più rilevanti sono:

- 802.3u Fast Ethernet (100 Mbit/s);
- 802.3z Gigabit Ethernet su fibra ottica (1000 Mbit/s);
- 802.3ae Gigabit Ethernet su doppino (1000 Mbit/s);
- 802.11 Wireless network;

Il lavoro di tali comitati si svolse in armonia con il modello di riferimento OSI.

- **IEEE 802.1** ***Higher Layer and Management***

È lo standard contenente le specifiche generali del progetto 802; esso è composto da molte parti, tra cui:

- 802.1 Part A (*Overview and Architecture*);
- 802.1 Part B (*Addressing Internetworking and Network Management*);
- 802.1 Part D (*MAC Bridges*).

IEEE 802 introduce l'idea che le LAN e le MAN devono fornire un'interfaccia unificata verso il livello Network (livello rete), pur utilizzando tecnologie trasmissive differenziate. Per ottenere tale risultato, il progetto IEEE 802 suddivide il livello Data Link in due sottolivelli:

- LLC (Logical Link Control);
- MAC (Media Access Control).

Il sottolivello LLC è comune a tutte le LAN, mentre il MAC è peculiare di ciascuna LAN, così come il livello fisico al quale è strettamente associato. Il sottolivello LLC è l'interfaccia unificata verso il livello Network ed è descritto nell'apposito standard IEEE 802.2, mentre i vari MAC sono descritti negli standard specifici di ogni rete locale (ad esempio il MAC CSMA/CD è descritto nello standard IEEE 802.3). Nel seguito, per facilità di lettura, si parlerà solo di reti locali (LAN), ma quanto detto vale ovviamente anche per le reti metropolitane (MAN), comprese anch'esse nel progetto IEEE 802.

- **MAC**

Il sottolivello MAC è specifico di ogni LAN e risolve il problema della condivisione del mezzo trasmissivo. Esistono vari tipi di MAC, basati su principi diversi, quali la contesa, il token, la prenotazione e il round-robin. Il MAC è indispensabile in quanto a livello 2 (Data Link) le LAN implementano sempre una sottorete trasmissiva di tipo broadcast in cui ogni sistema riceve tutti i frame inviati dagli altri.

Trasmettere in broadcast, cioè far condividere un unico canale trasmissivo a tutti i sistemi, implica la soluzione di due problemi:

- in trasmissione, verificare che il canale sia libero prima di trasmettere e risolvere eventuali conflitti di più sistemi che vogliano utilizzare contemporaneamente il canale;
- in ricezione, determinare a quali sistemi è effettivamente destinato il messaggio e quale sistema lo ha generato.

La soluzione del primo problema è data dai vari algoritmi di MAC che, per poter soddisfare il requisito di "apparecchiature indipendenti", devono essere algoritmi distribuiti su vari sistemi e non necessitare di un sistema master.

La soluzione del secondo problema implica la presenza di indirizzi a livello MAC (quindi nella MAC-PDU) che trasformino trasmissioni broadcast in:

- trasmissioni punto-punto, se l'indirizzo di destinazione indica un singolo sistema;
- trasmissioni punto-gruppo, se l'indirizzo di destinazione indica un gruppo di sistemi;
- trasmissioni effettivamente broadcast, se l'indirizzo di destinazione indica tutti i sistemi.

Il MAC deve anche tener conto della topologia della LAN, che implica leggere variazioni sulle possibili modalità di realizzazione del broadcast: con topologie a bus, è un broadcast a livello fisico (elettrico), mentre con topologie utilizzanti canali punto-punto, quali l'anello, è un broadcast di tipo logico.

Le reti locali hanno canali sufficientemente affidabili, quindi non è in genere necessario effettuare correzione degli errori. Se ciò fosse richiesto, sarebbe il sottolivello LLC ad occuparsene essendo il MAC sempre *connectionless*.

- **IEEE 802.3 (*CSMA/CD*)**

IEEE 802.3 è l'evoluzione della rete Ethernet proposta da Digital, Intel e Xerox (DIX). Utilizza un MAC di tipo CSMA/CD (*Carrier Sense Multiple Access – Collision Detection*) in cui l'arbitraggio del canale trasmissivo avviene tramite un meccanismo di contesa non deterministico, che non garantisce un tempo di attesa limitato superiormente.

IEEE 802.3 prevede una topologia logica a bus, con cablaggio a bus o a stella. La velocità trasmissiva è di 10 Mb/s e il throughput massimo di circa 4 Mb/s.

- **MAC-PDU**

Nelle reti locali, al livello 2 OSI, sono presenti due tipi di PDU corrispondenti ai due sottolivelli LLC e MAC. Il formato della LLC-PDU è comune a tutte le reti locali, mentre quello della MAC-PDU è peculiare di ogni singolo MAC. Tuttavia alcuni campi principali sono presenti in tutte le MAC-PDU. In particolare una MAC-PDU contiene due indirizzi (SAP), uno di mittente (MAC-SSAP) e uno di destinatario (MAC-DSAP), un campo INFO contenente la LLC-PDU (cioè il pacchetto di livello LLC) e una FCS (Frame Control Sequence) su 32 bit, cioè un codice a ridondanza ciclica (CRC) per l'identificazione di errori di trasmissione.

| MAC-DSAP | MAC-SSAP | INFO | |
|---|---|---|---|
| Indirizzo di destinazione | Indirizzo del mittente | LLC PDU | FCS |

- **Indirizzi MAC**

Gli indirizzi MAC sono lunghi 6 byte, si scrivono per convenzione in esadecimale e sono univoci a livello mondiale. Essi sono scritti in una ROM dal costruttore della scheda di rete e possono essere eventualmente sostituiti via software da indirizzi scritti in un apposito buffer. Essi si compongono di due parti di 3 byte ciascuna:

- i 3 byte più significativi indicano il lotto di indirizzi assegnato al costruttore della scheda di rete locale o all'organizzazione che ha progettato una data architettura di rete; essi vengono detti OUI (*Organization Unique Identifier*);
- i 3 byte meno significativi sono una numerazione interna progressiva decisa dal costruttore stesso.

I primi due bit trasmessi sul canale hanno un'importanza particolare: il primo si chiama I/G (*Individual/Group*) ed indica se l'indirizzo è di un singolo sistema o di un gruppo di sistemi, il secondo U/L (*Universal/Local*) indica se l'indirizzo è stato assegnato ufficialmente o è stato deciso su base locale.

Gli indirizzi MAC possono essere di tre tipi:
- *single*, se riferito ad un singolo sistema;
- *multicast*, se riferito ad un gruppo di sistemi;
- *broadcast*, se riferito a tutti i sistemi.
  Il *broadcast* è un tipo particolare di multicast con codifica esadecimale ff-ff-ff-ff-ff-ff.

Quando una scheda di rete locale riceve un pacchetto, non lo passa automaticamente al livello superiore (LLC), ma effettua una serie di controlli. Per prima cosa verifica che il pacchetto sia integro (cioè abbia una FCS corretta) e di dimensioni ammesse. Quindi analizza l'indirizzo di destinazione (MAC-DSAP). Si possono porre tre casi:
- se il MAC-DSAP è broadcast, il pacchetto viene sempre passato al LLC;
- se il MAC-DSAP è single, il pacchetto viene passato al LLC solo se il MAC-DSAP è uguale all'indirizzo hardware della scheda o a quello caricato da software nell'apposito buffer;
- se il MAC-DSAP è multicast, si verifica se la ricezione di quel multicast è stata abilitata dal software di livello superiore, cioè se la scheda appartiene al gruppo indirizzato. Poiché non è noto a priori a quanti gruppi possa appartenere una scheda, si usano delle tecniche di hash per mantenere la lista dei gruppi abilitati.

Gli indirizzi di gruppo servono principalmente per scoprire quali altri sistemi sono collegati alla rete locale, quali servizi questi mettono a disposizione e le relazioni esistenti tra gli indirizzi MAC e gli indirizzi di livello 3. La trasmissione in multicast ha due diverse modalità d'impiego:
- *Solicitation*, un sistema che necessiti di accedere ad un dato servizio richiede, trasmettendo un pacchetto all'indirizzo di multicast di tale servizio, quali sistemi siano in grado di offrirlo. I sistemi che offrono il servizio rispondono alla richiesta;
- *Advertisement*, i sistemi che offrono un servizio trasmettono periodicamente tale informazione in multicast. Un esempio semplice è il messaggio di "Hello" con cui ogni sistema comunica periodicamente la sua esistenza e quindi la sua raggiungibilità sulla rete locale.

- **Relazioni tra L3, LLC e MAC**

In Figura A-6-10 sono rappresentate le relazioni tra le PDU di livello 3 (Network), le LLC-PDU e le MAC-PDU.

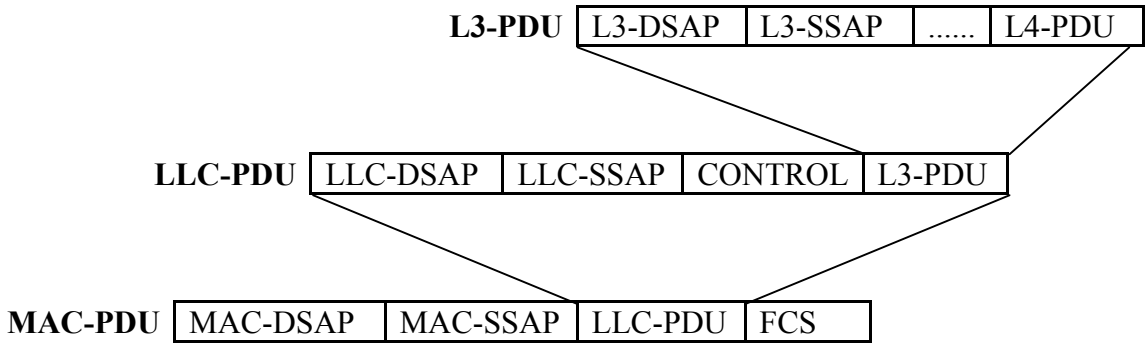



**Figura A-6-10: Relazione tra MAC-PDU e LLC-PDU.**

Ogni interfaccia di rete locale è gestita da un suo livello MAC. Su tale livello MAC si appoggia un livello LLC. Il livello MAC è implementato nell'hardware della scheda di rete locale, mentre il livello LLC è di solito realizzato in software. Ogni livello LLC può gestire un solo livello MAC: questo significa che un livello LLC non può avere funzionalità di "relaying" (non può inoltrare pacchetti) tra più MAC. Tale funzionalità di instradamento dei pacchetti è delegata al livello 3.

## *A.3 Sala macchine e componenti infrastrutturali*

Gli aspetti infrastrutturali di una sala macchine sono molto importanti ai fini del conseguimento di un sistema informatico ad alta affidabilità e ad alta disponibilità.

È fondamentale che tutti gli apparati informatici abbiano a disposizione un'alimentazione elettrica stabile, ovvero senza variazioni di tensione (o al limite variazioni contenute entro i margini di tolleranza), senza interruzioni e senza microinterruzioni.

Per raggiungere l'obiettivo della continuità elettrica è opportuno installare un Gruppo di Continuità (UPS: *Unattended Power Supply*) dedicato alle apparecchiature informatiche in modalità esclusiva. L'UPS svolge la funzione di stabilizzatore dell'energia elettrica e inoltre è dotato di accumulatori di energia per sopperire a brevi interruzioni di erogazione.

I due parametri più importanti di cui tenere conto in fase di dimensionamento dell'UPS per la sala macchine, sono il carico complessivo (in Ampere o in KWatt) degli apparati che dovrà alimentare e la durata della carica in mancanza di energia elettrica (per il dimensionamento degli accumulatori).

Di fatto l'UPS trasforma la tensione alternata (trifase 380V 50Hz) erogata dal fornitore di elettricità in tensione continua, con la quale tiene in carica gli accumulatori. Poi, tramite un circuito detto *inverter*, trasforma la tensione continua nuovamente in tensione alternata stabile.

La continuità è garantita dalla carica degli accumulatori, anche a fronte di interruzioni di erogazione da parte della società che fornisce l'energia elettrica. In genere la carica accumulata è in grado di sopperire ad interruzioni la cui durata può

variare da alcuni minuti a qualche ora (parametri fortemente dipendenti dal consumo complessivo degli apparati informatici e dal dimensionamento degli accumulatori).

Per sopperire ad interruzioni più lunghe, a monte dell'UPS è opportuno installare un gruppo elettrogeno. Consiste in un motore a scoppio (generalmente Diesel) con un alternatore in grado di convertire l'energia meccanica prodotta dal motore in energia elettrica, quindi di erogare tensione alternata (trifase 380V 50Hz). Il gruppo elettrogeno ha un sensore che avverte la mancanza di energia elettrica, in grado di dare il consenso per avviare il motore; i tempi di intervento sono dell'ordine di decine di secondi.

I sistemi UPS di ultima generazione hanno dei circuiti elettronici di autodiagnosi e monitoraggio, connessi direttamente sulla rete LAN di trasmissione dati (Ethernet), in grado di mandare allarmi in caso di fallimenti di erogazione di energia elettrica, ma anche in caso di guasto dello stesso UPS.

In effetti l'UPS, come tutti gli apparati elettronici attivi, ha, a sua volta, per quanto piccola, una sua probabilità di fallimento. Il fallimento dell'UPS, potrebbe generare una sospensione repentina e non programmata nell'erogazione dell'alimentazione elettrica della sala macchine. Questo è di per se uno degli eventi più traumatici per i sistemi informatici, che può provocare danni ai sistemi e alla consistenza dei dati immagazzinati nello storage.

Per evitare tale evenienza l'ideale sarebbe ridondare l'UPS mettendo un secondo UPS gemello, e adottare la politica di acquistare apparati informatici con alimentatori duali e ridondati in modo da alimentare ogni singolo apparato da entrambi gli UPS. In mancanza di un UPS di ridondanza si può comunque ottenere un buon livello di affidabilità complessiva, alimentando il secondo alimentatore di ogni sistema informatico con l'energia proveniente direttamente dalla società fornitrice.

Inoltre gli apparati informatici producono calore che provoca l'innalzamento della temperatura nell'ambiente in cui sono situati. D'altra parte l'elettronica è estremamente sensibile alla temperatura; gli stessi apparati non riuscirebbero a funzionare se la temperatura superasse le soglie critiche. Per questo motivo è fondamentale avere un adeguato sistema di condizionamento della sala, idoneo ad estrarre il calore prodotto da tutti gli apparati informatici in funzione.

Gli aspetti ingegneristici relativi ai sistemi di raffreddamento sono estremamente complessi ed esulano da questa trattazione, tuttavia occorre ricordare che due tra i parametri più importanti di cui tenere conto in fase di dimensionamento dell'impianto di raffreddamento sono la capacità di raffreddamento (in chilofrigorie), ovvero la quantità di calore che il sistema deve essere in grado di estrarre nell'unità di tempo e il flusso dell'aria, che deve essere idoneo al volume e alla forma della sala macchine per garantire un'isotropia di distribuzione del calore ed una temperatura costante nello spazio e nel tempo.

A tale scopo, dato che il flusso dell'aria condizionata è molto spesso convogliato attraverso il pavimento rialzato, è opportuno prendere in considerazione anche l'altezza da terra del pavimento rialzato, quale parametro importante per garantire un flusso adeguato; in effetti occorre considerare che sotto al pavimento

rialzato di una sala macchine si accumulano col tempo una grande quantità di cavi e canaline che possono ostruire il passaggio dell'aria.

Quando si dimensiona l'impianto di distribuzione elettrica per una sala macchine occorre tenere conto anche del consumo dell'impianto di condizionamento. Un impianto di condizionamento efficiente consuma dal 30% al 40% del carico dei sistemi, per tenerli a temperatura costante. Un sistema di infrastrutture ben progettato deve prevedere che il sistema di condizionamento rimanga alimentato anche in mancanza di energia elettrica da parte della società erogatrice. Ciò implica che esso debba essere alimentato quanto meno dal gruppo elettrogeno (se non anche dall'UPS).

Infine una sala macchine deve essere dotata di un adeguato impianto antincendio e antiallagamento. L'incendio e l'allagamento sono due eventi in grado di distruggere in poco tempo tutti i sistemi informatici e, ciò che è peggio, le informazioni accumulate ed elaborate in decenni di lavoro.

Questo tipo di impianti hanno dei sensori in grado di rilevare acqua sul pavimento o fumi nell'ambiente. Nel primo caso possono lanciare un allarme tramite messaggio e-mail o SMS oppure possono lanciare un allarme acustico, con ripetizione in ambienti presidiati 24 ore al giorno. Nel secondo caso, oltre a mandare l'allarme, sono in grado di estinguere l'incendio emettendo una scarica di un particolare gas capace di inibire la reazione chimica di combustione.

### A.3.1 Soluzioni tecniche implementate ai LNF

Nella Figura A-6-11 è rappresentata la pianta del piano terra dell'edificio Calcolo dei LNF, ove è situata la sala macchine.

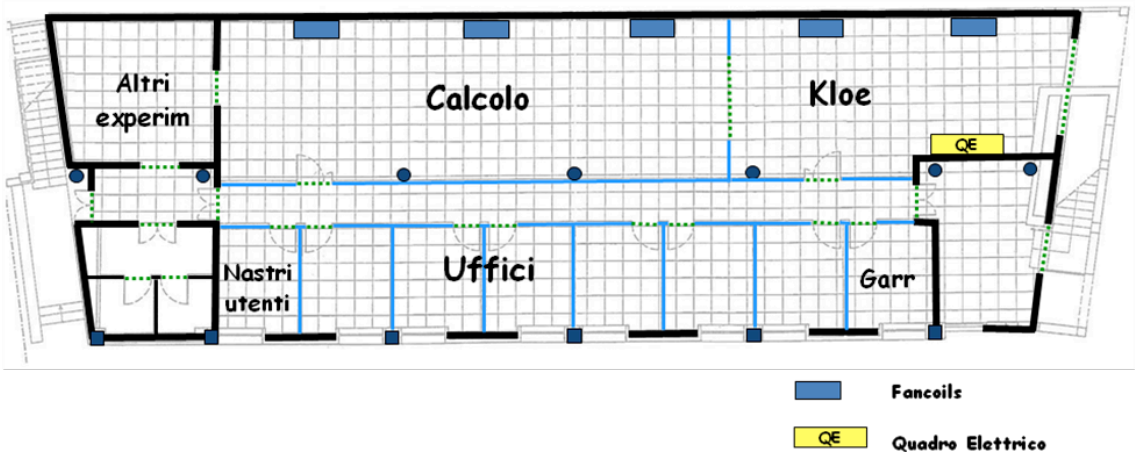



**Figura A-6-11: LNF: Edificio Calcolo – Piano Terra**

Ogni quadretto rappresenta una mattonella quadrata di pavimento rialzato avente lato di 60 centimetri. Lo spazio complessivo è di circa 300m$^2$ di cui circa 100m$^2$ sono dedicati alla sala macchine del Servizio di Calcolo.

La distribuzione elettrica avviene sotto al pavimento rialzato attraverso una serie di blindosbarre. Le blindosbarre sono alimentate da un UPS dedicato al Servizio

di Calcolo da 160KVA, avente un'autonomia a pieno carico di circa 15 minuti. L'UPS, a sua volta, è alimentato da un gruppo elettrogeno che serve tutte le utenze critiche dei Laboratori Nazionali di Frascati. Il consumo complessivo attuale della sala macchine è di circa 50KWatt (escluso l'impianto di condizionamento).

Il sistema di condizionamento è a flusso di aria canalizzata sotto al pavimento rialzato attraverso armadi che contengono dei radiatori per lo scambio termico (detti *fancoil*) in cui arriva acqua fredda (circa 7 gradi) dall'unità centrale esterna. L'unità centrale esterna si è guastata irreversibilmente qualche anno fa; in tale occasione, per un ripristino celere dei servizi, fu adottata la soluzione di far ricevere ai fancoil l'acqua refrigerata dalla centrale frigorifera della macchina acceleratrice, poco distante, tramite la stesura di una coppia di tubi idraulici.

Tale soluzione tuttavia, pur essendo efficiente ed affidabile, presenta il problema di non prevedere un'alimentazione dal gruppo elettrogeno per via dell'alto assorbimento della centrale frigorifera della macchina acceleratrice.

Lo schema degli impianti è rappresentato nella Figura A-6-12.

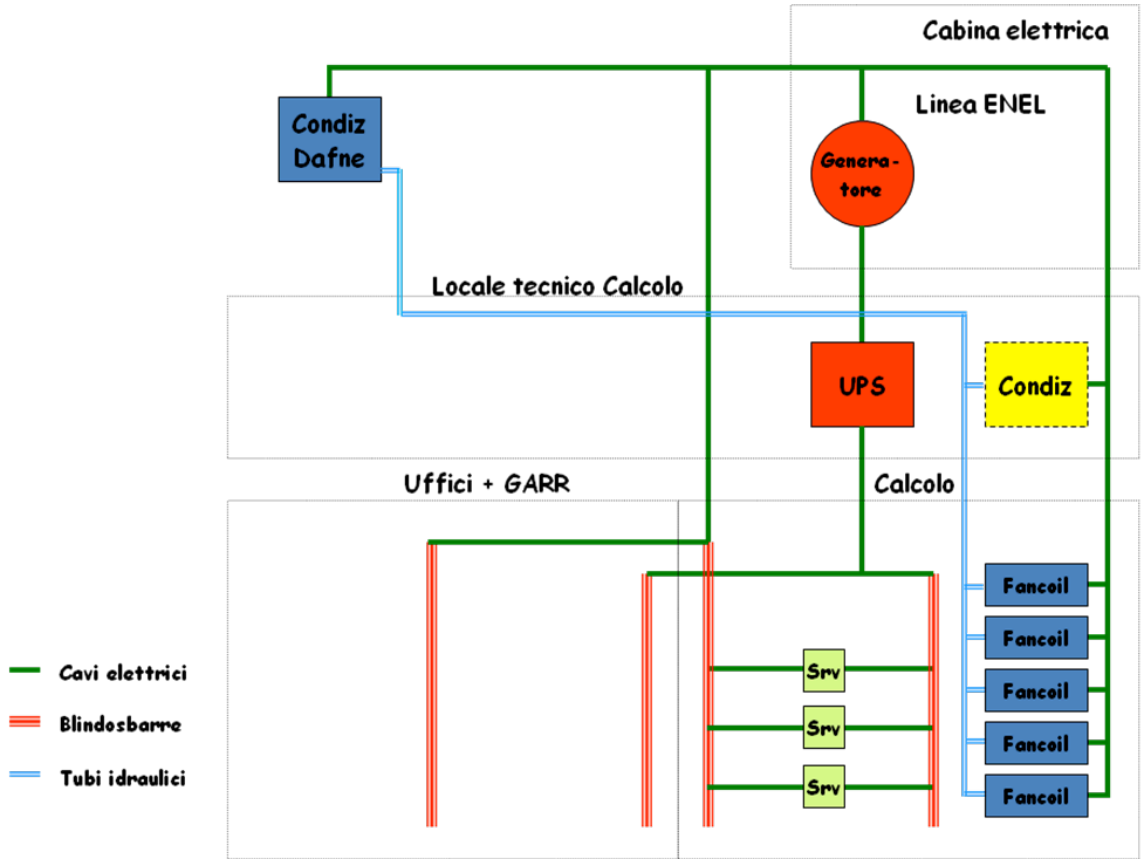



**Figura A-6-12: Schema degli impianti**

Nell'ambito del progetto di espansione del sistema di calcolo scientifico dei LNF è previsto che l'edificio Calcolo alloggi anche i sistemi informatici degli esperimenti. In particolare, gli esperimenti ad LHC, che avranno bisogno di elaborare una grande quantità di informazioni sono Alice e Atlas. Ma anche esperimenti più

piccoli e i fisici teorici, con le loro necessità di calcolo, contribuiranno ad aumentare i sistemi installati nell'edificio (rif. Tabella A-6-3).

| | CPU (KSI2K) | Disk (TB) | Power (KW) |
|---|---|---|---|
| **CSN-I** | 1640 | 628 | 76 |
| **CSN-II** | 205 | 35 | 4 |
| **CSN-III** | 911 | 261 | 31 |
| **CSN-IV** | 380 | 6 | 14 |
| **Totale** | **3136** | **1130** | **125** |

**Tabella A-6-3: Risorse di calcolo richieste**

Le attuali dimensioni della sala macchine sono tuttavia insufficienti per alloggiare tutte le risorse informatiche necessarie (server di calcolo e storage), e anche gli aspetti infrastrutturali legati al consumo di energia elettrica e al condizionamento andranno adeguati alle necessità di progetto.

Già si è avviata un procedura amministrativa per eseguire i lavori di ampliamento dei locali destinati alle apparecchiature informatiche. La nuova sala macchine è rappresentata in Figura A-6-13.

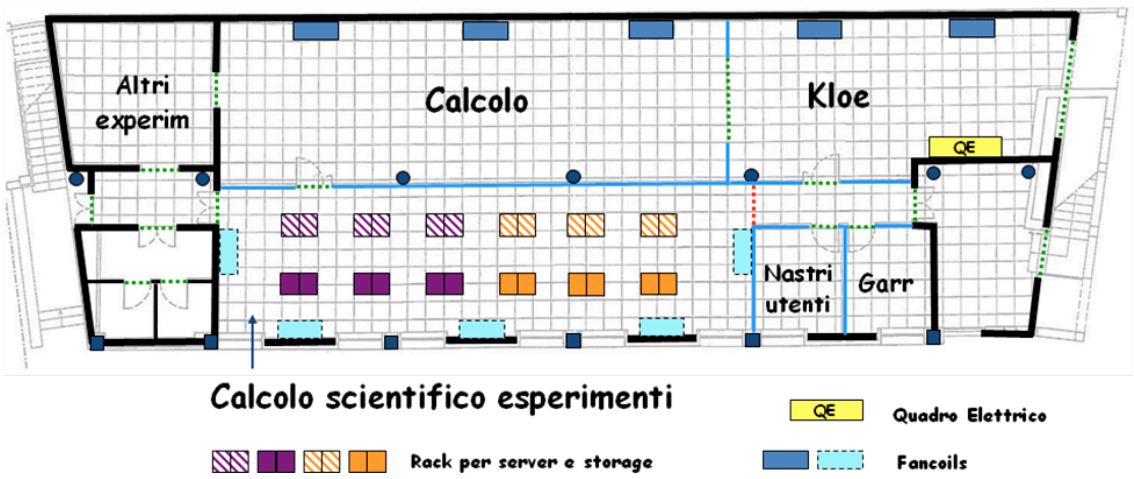



**Figura A-6-13: LNF: Edificio Calcolo – nuova sala macchine**

Ai fini del dimensionamento degli impianti infrastrutturali occorre tener presente che i valori della Tabella A-6-3 si aggiungono a quelli già utilizzati dalle attuali risorse del Servizio di Calcolo; inoltre anche queste ultime è previsto che aumentino. Considerando gli opportuni margini di errore, il parametro di consumo elettrico di riferimento a regime è quindi di circa 200 KW.

Per quanto riguarda gli impianti infrastrutturali siamo ancora in fase di progettazione ingegneristica ed è previsto che vengano realizzati entro l'estate del 2011 secondo lo schema di massima rappresentato in Figura A-6-14.

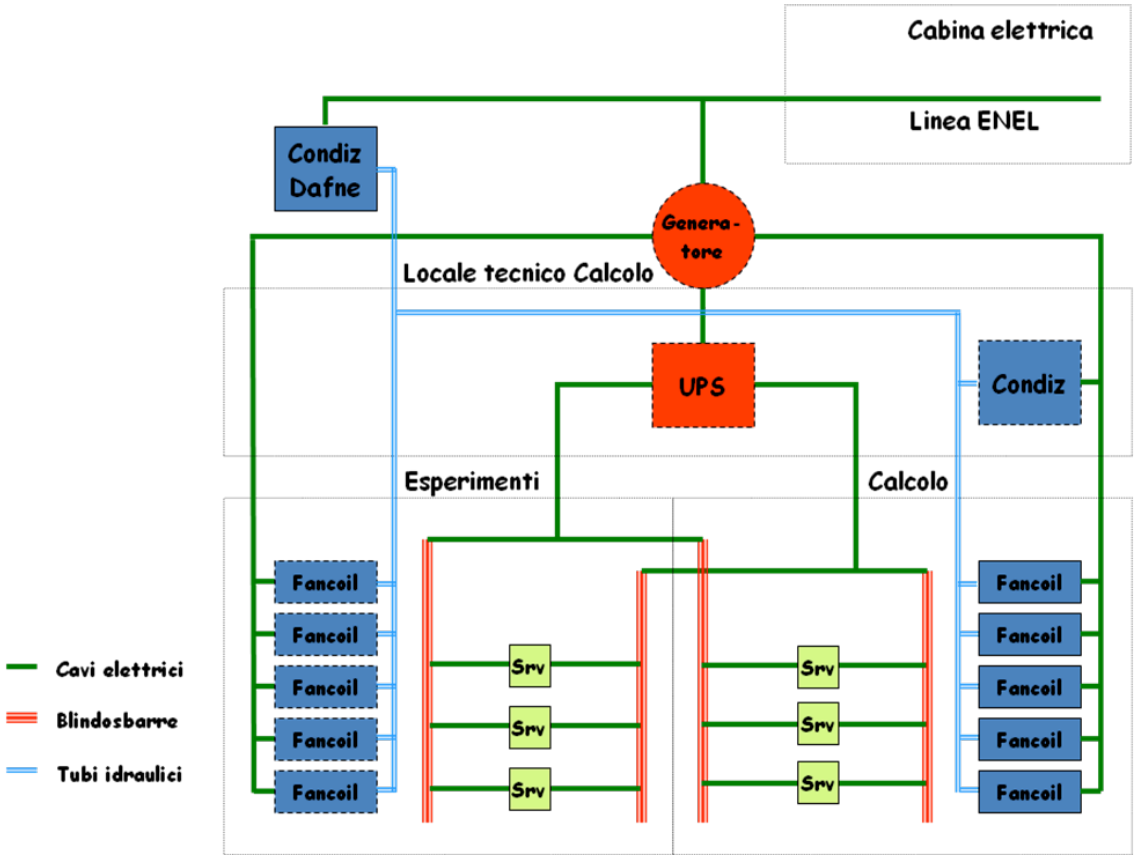



**Figura A-6-14: Schema dei nuovi impianti**

Lo schema prevede l'utilizzo dell'attuale UPS, finché risulterà sufficiente; dopo verrà sostituito con uno in grado di erogare maggiore potenza, oppure ridondato.

Si installeranno nuovi Fancoil nella sala macchine dedicata al calcolo scientifico degli esperimenti. Si continuerà ad utilizzare la centrale frigorifera esterna della macchina acceleratrice che ha ancora molto margine di capacità frigorifera da poter utilizzare. Tuttavia è prevista l'installazione di una centrale frigorifera di backup alimentata dal gruppo elettrogeno.

Questa soluzione garantisce la continuità di servizio sia durante le fermate del sistema centrale di condizionamento della macchina frigorifera, sia durante la sospensione di erogazione dell'energia elettrica da parte della società fornitrice.

Il gruppo elettrogeno dei LNF è idoneo anche per l'aumento dei consumi previsto per la realizzazione del progetto.

# Glossario

**802.1x**: vedi IEEE 802.1x

**802.2**: vedi IEEE 802.2

**802.3**: vedi IEEE 802.2

**3DES**: il Triple DES (DES triplo) è un cifrario a blocchi basato sulla ripetizione del Data Encryption Standard (DES) per tre volte.

**Access Point**: dispositivo che permette all'utente mobile di collegarsi ad una rete wireless. L'access point, collegato fisicamente ad una rete cablata (oppure via radio ad un altro access point), riceve ed invia un segnale radio all'utente, permettendo così la connessione.

**Account**: iscrizione registrata su di un server che, tramite l'inserimento di un username e di una password, consente l'accesso alla rete e/o ai servizi.

**ACL**: *Access Control List*, un meccanismo generalmente usato in informatica per esprimere regole complesse. Tra le sue applicazioni principali, la configurazione di Firewall e dei diritti di accesso a file e directory.

**AES**: Advanced Encryption Standard (AES), è un algoritmo di cifratura a blocchi utilizzato come standard dal governo degli Stati Uniti d'America. Data la sua sicurezza e le sue specifiche pubbliche si presume che in un prossimo futuro venga utilizzato in tutto il mondo come è successo al suo predecessore, il Data Encryption Standard (DES).

**AFS**: *Andrew File System*, è un file system distribuito sviluppato dalla Carnegie Mellon University, all'interno del progetto Andrew. Il nome è stato dato in onore di Andrew Carnegie e Andrew Mellon. L'utilizzo principale di questo file system è nell'elaborazione distribuita.

**ANSI**: *American National Standards Institute*, è il rappresentante USA nell'ISO.

**Anti-Spamming**: funzione dei server di posta elettronica che blocca l'inoltro e la ricezione indiscriminata di grosse quantità di messaggi di posta elettronica, spesso inviati a scopo pubblicitario e commerciale.

**Antivirus**: Un antivirus è un software atto a rilevare ed eliminare virus informatici o altri programmi dannosi (malware) come worm, trojan e dialer.

**API**: *Application Programming Interface*, termine generico che indica un set di librerie che permettono ad un programmatore di inserire alcune funzionalità nei propri programmi.

**AppleTalk**: era un protocollo molto usato nelle reti Macintosh. Il suo principale scopo era quello di condividere stampati e/o file, e quindi metterli a disposizioni degli utenti che fanno parte della rete.

**BIOS**: Basic Input Output System, parte del firmware residente nella ROM e che contiene i comandi essenziali per l'avvio del computer. È residente permanentemente in chip EEPROM o flash ROM sulla scheda madre e non è modificabile dall'utente.

**BootP**: Boot Protocol, è uno dei protocolli standard che utilizzano un'architettura di tipo client/server per l'assegnazione dinamica di indirizzi IP agli host di una LAN.

**Broadcast**: modalità di instradamento per la quale un pacchetto inviato ad un indirizzo particolare (detto appunto di broadcast) verrà consegnato a tutti i computer collegati alla rete (ad esempio, tutti quelli su un segmento di rete ethernet, o tutti quelli di una sottorete IP).

**Browser**: è un programma che consente di visualizzare i contenuti delle pagine dei siti web e di interagire con essi, permettendo così all'utente di navigare in internet. Il browser è infatti in grado di interpretare l'HTML e di visualizzarlo in forma di ipertesto.

**Buffer overflow**: vulnerabilità di sicurezza che può affliggere un programma software. Consiste nel fatto che tale programma non controlla in anticipo la lunghezza dei dati in arrivo, ma si limita a scrivere il loro valore in un buffer di lunghezza prestabilita, confidando che l'utente (o il mittente) non immetta più dati di quanti esso ne possa contenere: questo può accadere se il programma è stato scritto usando funzioni di libreria di input/output che non fanno controlli sulle dimensioni dei dati trasferiti.

**Bug**: un errore di programmazione che porta a risultati inattesi, o più comunemente al blocco, in un programma o nell'intero sistema del computer.

**Building distributor**: armadio di edificio per la distribuzione della rete verso i floor distributor.

**CA**: vedi *Certification Authority*.

**Cache**: memoria temporanea, non visibile al software, che memorizza un insieme di dati che possano successivamente essere velocemente recuperati su richiesta; la cache è un'area di memoria, spesso notevolmente più veloce delle altre, in cui vengono registrate le informazioni che si intendono utilizzare più di frequente.

**Campus distributor**: armadio di Campus per la distribuzione della rete verso i building distributor.

**CCITT** : *Comitè Consultatif International de Telegraphie et Telephonie*, è l'organismo internazionale che emette le specifiche tecniche per la trasmissione dati su linee telefoniche (adottate dalle PTT).

**CERN**: Organizzazione Europea per la Ricerca Nucleare, è il più grande laboratorio al mondo di fisica delle particelle. Si trova al confine tra Svizzera e Francia alla periferia

ovest della città di Ginevra. La convenzione che istituiva il CERN fu firmata il 29 settembre 1954 da 12 stati membri. Oggi ne fanno parte 20 stati membri più alcuni osservatori, compresi stati extraeuropei.

**Certification Authority (CA)**: *Certification Authority*, ente pubblico o privato, abilitato a rilasciare un certificato digitale tramite procedura di certificazione che segue standard internazionali e conforme alla normativa europea e nazionale in materia. Il sistema in oggetto utilizza la crittografia a doppia chiave, o asimmetrica, in cui una delle due chiavi viene resa pubblica all'interno del certificato (chiave pubblica), mentre la seconda, univocamente correlata con la prima, rimane segreta e associata al titolare (chiave privata).

**Certificato digitale**: conosciuto anche come Digital ID, è l'equivalente elettronico di un passaporto o di una licenza di commercio. Viene emesso da un'Autorità Certificativa (Certification Authority o CA) e certifica ufficialmente l'identità del suo possessore. Un Digital ID è composto da due chiavi complementari.

**Cifratura o crittografia**: sistema che permette di codificare messaggi testuali in simboli non comprensibili a prima vista, in modo che non possano essere interpretati da chi non possiede la corretta chiave di lettura. La crittografia è nata prima dei computer, ed è stata spesso utilizzata per cifrare i messaggi militari. In pratica le lettere del testo vengono trasformate con un determinato algoritmo ed è sufficiente conoscere o scoprire tale algoritmo per decifrare (decryption) il messaggio. Oggi, invece, l'algoritmo è ben noto, ma per la decodifica è necessario conoscere una o più parole chiave. In inglese è chiamata “encryption”.

**Cluster**: Termine che indica un gruppo di computer che lavorano congiuntamente per eseguire un unico compito. Si parla, in genere, di cluster di calcolatori per indicare un insieme di calcolatori che usano le proprie risorse in un sistema di calcolo parallelo o per aumentare la disponibilità dei servizi offerti. Spesso la parola cluster è usata erroneamente come sinonimo di Farm.

**Computing Element (CE)**: è una risorsa grid in grado di fornire cicli di CPU per l’esecuzione di job. Un CE può anche essere il gateway di un cluster di PC, un supercomputer per l’esecuzione di job paralleli, o una postazione standard di calcolo interattivo in grado di gestire applicazioni grafiche e I/O verso dispositivi di storage.

**CPU**: *Central Processing Unit*, unità di elaborazione centrale o processore. È la parte "pensante" di ogni computer, costituita da un sottile cristallo di silicio ottenuto dal wafer. Il processore esegue le istruzioni e i dati dei programmi presenti nella memoria RAM e nei registri interni al processore. I registri sono delle memorie particolari e piú veloci, in termini di accesso, ad ogni altro tipo di memoria.

**Cracker**: vedi hacker.

**Datalink**: connessione tra una stazione e un'altra al fine di trasmettere e ricevere segnali digitali (livello 2 della pila OSI).

**DES**: *Data Encryption Standard*, algoritmo introdotto negli USA nella metà degli anni ’70. Il DES è un sistema crittografico che sfrutta le chiavi di lunghezza pari a 56 bit.

**DHCP**: *Dynamic Host Configuration Protocol*, è uno dei protocolli standard che utilizzano un'architettura di tipo client/server per l'assegnazione dinamica delle impostazioni IP agli host di una LAN.

**DNS**: *Domain Name System*, sistema utilizzato per la risoluzione di nomi di host in indirizzi IP e viceversa. Il servizio è realizzato tramite un database distribuito a livello mondiale, costituito dai server DNS.

**DWDM**: vedi WDM

**EAP**: *Extensible Autentication Potocol*, protocollo di autenticazione utilizzato spesso sugli access point e nelle connessioni PPP. L'utilizzo di EAP, all'interno di una rete wireless, prevede che non sia l'access point ad autenticare il client: esso ridirige la richiesta di autenticazione avanzata dal client ad uno specifico server, configurato per questo scopo come un RADIUS.

**E-mail**: servizio Internet grazie al quale ogni utente può inviare o ricevere dei messaggi. È tra le applicazioni Internet più conosciute e più utilizzate attualmente.

**Encryption**: trasformazione, mediante un algoritmo matematico ed una chiave, di un messaggio leggibile in un altro messaggio non facilmente interpretabile. L'obiettivo è quello di nascondere la natura dell'informazione agli occhi di persone non autorizzate a venirne in possesso.

**Ethernet**: la più diffusa tecnologia LAN inventata dalla Xerox Corporation e sviluppata successivamente dalla stessa Xerox insieme ad Intel e Digital Equipment Corporation. La tecnologia Ethernet utilizza il protocollo CSMA/CD (Collision Detection) per trasferire i pacchetti tra computer. Opera su vari tipi di cavi (coassiali o doppini telefonici) ad una velocità di 10 Mbps, è simile alle serie standard IEEE 802.3.

**Exploit**: codice che, sfruttando un bug o una vulnerabilità, porta all'acquisizione di privilegi o al denial of service di un computer.

**Extranet**: una rete simile ad Internet ma limitata nell'accesso a partner, fornitori o clienti di un'azienda, cui è stata fornita un'apposita password. Permette di condividere in modo semplice e conveniente informazioni e risorse.

**Farm**: Questo termine indica un gruppo di computer adibito ad una produzione specifica, come ad esempio il lavoro di simulazione o analisi dati per un esperimento. Una farm può essere gestita in modi più o meno complicati dal software che si occupa del job management.

**Fast Ethernet**: termine collettivo per indicare un numero di standard Ethernet che trasportano il traffico alla velocità di 100 Mbps rispetto alla velocità originale Ethernet di 10 Mbps. Tra gli standard ethernet a 100 megabit, 100baseTX è il più comune e supportato dalla grande maggioranza dell'hardware prodotto.

**FC**: *Fibre Channel*, tecnologia per reti dati, usata principalmente per implementazioni in Storage Area Network. Fibre Channel è uno standard nato principalmente per un

utilizzo nel campo dei supercomputer, ma è diventato il tipo di connessione standard per le Storage Area Networks nell'enterprise storage. Nonostante la connotazione comune del suo nome, il segnale Fibre Channel può andare sia su cavi di rame UTP che su cavi a fibra ottica. Il Fibre Channel Protocol (FCP) è il protocollo di interfaccia dello SCSI sul Fibre Channel.

**Fingerprinting**: insieme di dispositivi software attraverso i quali è possibile rilevare le impostazioni di computer remoti, al fine di consentire di catturare il maggior numero di informazioni sulla macchina.

**Firma digitale**: cifratura della chiave di hash atta a garantirne l'autenticità di un documento informatico. Sistema di autenticazione di documenti digitali tale da garantire non ripudio.

**Firmaware**: software registrato in una memoria particolare del computer (di solito di sola lettura), che comprende le istruzioni basilari per il corretto funzionamento del personal computer (ad esempio il BIOS), utilizzato all'avvio o per l'interazione con altri componenti, tramite l'implementazione di protocolli di comunicazione o interfacce di programmazione.

**Firewall**: Apparato di rete hardware o software che filtra tutti i pacchetti entranti ed uscenti, da e verso una rete o un computer, applicando regole che contribuiscono alla sicurezza della stessa. È uno degli strumenti principali della sicurezza informatica, progettato per impedire accessi non autorizzati a/da reti private. Il suo utilizzo tipico quindi è quello di impedire agli utenti provenienti da Internet l'accesso non autorizzato ad una Intranet. Un firewall si occupa di filtrare i dati che passano da un computer ad un altro sulla rete, quindi applica un modello di sicurezza di tipo "perimetrale".

**Floor distributor**: armadio di piano per la distribuzione della rete alle postazioni di lavoro degli utenti.

**FTP**: File Transfer Protocol, è un protocollo standard usato comunemente su internet per il trasferimento di file tra host diversi.

**GARR**: *Gruppo per l'Armonizzazione delle Reti della Ricerca*, il consorzio italiano che si occupa di gestione ed ampliamento della rete telematica nazionale ad altissima velocità delle università e della ricerca interconnessa con tutte le reti della ricerca europee e mondiali.

**Gateway**: dispositivo di rete che opera al livello di rete e superiori del modello ISO/OSI. Il suo scopo principale è quello di veicolare i pacchetti di rete all'esterno di una rete locale (LAN).

**Gigabit Ethernet**: è l'evoluzione a 1000 Mbit/s del protocollo Fast Ethernet operante a 100 Mbit/s.

tecnologia LAN che utilizza lo stesso metodo di trasmissione di Ethernet 10 Mbps, ovvero il protocollo CSMA/CS (Collision Detection), ma che opera con una velocità cento volte superiore, 1.000 Mbps, cioè 1 Gbps.

**GIIS**: Grid Index Information Service, parte del servizio MDS che riceve le informazioni dalle varie risorse (tramite i GRIS) e le organizza in un database LDAP.

**GIS**: Grid Information Service, è la parte del servizio MDS che si occupa di collezionare, organizzare e diffondere le informazioni relative allo stato delle risorse della grid.

**Globus**: Insieme di pacchetti software (tool kit) che costituiscono il nucleo dei servizi forniti da una griglia computazionale grid. Globus viene spesso considerato come un "collante" che tiene insieme e rende omogenee le parti che costituiscono la grid.

GRAM: Globus Resource Allocation Manager, è la parte di globus che gestisce le richieste di un utente o di un processo per quanto riguarda l'allocazione delle risorse.

**Grid**: infrastruttura di calcolo distribuito, utilizzati per l'elaborazione di grandi quantità di dati, mediante l'uso di una vasta quantità di risorse. In particolare, tali sistemi permettono la condivisione coordinata di risorse all'interno di un'organizzazione virtuale.

**Grid element**: elemento di una grid in grado di offrire un particolare servizio.

**GridFTP**: È un estensione del protocollo FTP standard che ha, come principale caratteristica, la possibilità di saturare il link su cui avviene la trasmissione dei dati, utilizzando più socket contemporaneamente.

**GRIS**: Grid Resource Information Service, parte del servizio MDS che si occupa di comunicare ai livelli superiori della struttura gerarchica dell'MDS lo stato di una risorsa.

**GSI**: Grid Security Infrastructure, è la parte del software che gestisce l'infrastruttura adibita alla sicurezza di una grid.

**Hacker**: esperto informatico che mira, attraverso attacchi a siti Internet o reti pubbliche o private, a fornire elementi per individuare i limiti di un programma o di un sistema di sicurezza. Nei casi peggiori possono diventare criminali che attaccano le istituzioni o le aziende private, modificano le pagine dei siti istituzionali o delle multinazionali, accedendo e modificando dati nei loro computer, organizzano truffe.

**Hash (Algoritmo di hash)**: la funzione hash è una funzione non iniettiva che mappa una stringa di lunghezza arbitraria in una stringa di lunghezza predefinita. Esistono numerosi algoritmi che realizzano funzioni hash con particolari proprietà che dipendono dall'applicazione. Spesso si intende una procedura attraverso la quale, dato un messaggio ed una chiave, è possibile determinare l'integrità del messaggio.

**Host**: sinonimo di computer, elaboratore connesso alla rete.

**HPC**: *High Performance Computing*, tecnologie utilizzate da computer cluster per creare dei sistemi di elaborazione in grado di fornire delle prestazioni molto elevate. Il termine è molto utilizzato essenzialmente per sistemi di elaborazioni utilizzati in campo scientifico.

**HTC**: *High Throughput Computing*, con questa espressione si identifica un ambiente di elaborazione in cui è richiesta una enorme potenza di calcolo per l'esecuzione di job che analizzano grandi quantità di dati, che vengono scambiati attraverso canali di comunicazione ad alte prestazioni .

**HTTP**: *HyperText Markup Language*, è il protocollo standard utilizzato su Internet per la trasmissione e lo scambio di ipertesti sul web.

**IEEE**: *Institute of Electrical and Electronics Engineers*, è l'organizzazione professionale mondiale degli ingegneri elettrici ed elettronici con gruppi di standardizzazione sulle reti di calcolatori.

**IEEE 802**: suite di standard per le reti locali (LAN) e per le reti metropolitane (MAN). Più precisamente, gli standard "802" sono dedicati alle reti che hanno pacchetti di lunghezza variabile.

**IEEE 802.1**: gruppo di lavoro del progetto IEEE 802 relativo al protocollo Ethernet. Il suo lavoro riguarda: architettura delle reti locali (LAN) e metropolitane (MAN), connessione (internetworking) fra reti 802 LAN, MAN e le reti geografiche (WAN), sicurezza del collegamento, gestione globale della rete, e protocolli di strato superiore agli strati MAC e LLC.

**IEEE 802.11**: standard per le reti wireless WLAN, con connessione a infrarossi o con onde radio. Si divide in:

- 802.11a (a 54 Mbps alla frequenza di 5.8 GHz) chiamata anche HyperLan2
- 802.11b (a 11 Mbps alla frequenza di 2.4 GHz), chiamata anche Wi-Fi
- 802.11g (che opera a 54 Mbps alla frequenza di 2.4 Mbps)
- 802.11a/g (che opera a 54 Mbps alla frequenza di 2.4 o 5 GHz)
- 802.11n (che deriva dalla 802.11g, l ma con doppia o tripla velocità)

**IEEE 802.11i**: conosciuto anche come WPA2, è uno standard sviluppato dalla IEEE specificamente per fornire uno strato di sicurezza alle comunicazioni basate sullo standard IEEE 802.11.

**IEEE 802.1x**: standard basato sul controllo delle porte di accesso alla rete LAN e MAN. Questo standard provvede a autenticare e autorizzare i dispositivi collegati alle porte della rete (switch e access point) stabilendo un collegamento punto a punto e prevenendo collegamenti non autorizzati alla rete locale. Viene utilizzato dalle reti locali wireless per gestire le connessioni agli access point e si basa sul protocollo EAP, Extensible Authentication Protocol.

**IEEE 802.2**: Protocollo appartenente al comitato IEEE 802, i cui diversi gruppi di lavoro si sono occupati per anni della standardizzazione delle reti LAN. L'802.2 in particolare fu il gruppo che si occupò della standardizzazione del livello *Logical link control*, sottolivello del livello *Data Link* della pila ISO/OSI delle reti di calcolatori. Sotto di esso (sempre nel livello 2) c'è il sottolivello MAC, che non è unico ma dipende dal mezzo di trasmissione scelto (Ethernet, Token ring, FDDI, 802.11, etc.).

**IEEE 802.3**: lo standard *CSMA/CD* descrive una tecnologia per reti locali (LAN) derivata nel 1985 dalla precedente tecnologia Ethernet. È probabilmente il più popolare di un'ampia famiglia di protocolli, IEEE 802. Nella pila di protocolli di rete del modello di riferimento ISO/OSI, 802.3 occupa il livello fisico e la parte inferiore del livello datalink. IEEE ha infatti ritenuto opportuno suddividere questo livello in due parti: LLC, *Logical Link Control* e MAC, *Media Access Control*. Il sottolivello LLC è comune a tutti gli standard della famiglia IEEE 802, mentre il sottolivello MAC è più strettamente legato al livello fisico, e le sue diverse implementazioni hanno il compito di offrire un'interfaccia comune al livello LLC. Fra queste implementazioni vanno ricordate in particolare 802.4, token bus e 802.5, token ring. Le caratteristiche di 802.3 sono ben riassunte nell'acronimo CSMA/CD:

- *Carrier Sense*: ogni stazione sulla rete locale ascolta continuamente il mezzo trasmissivo;
- *Multiple Access*: il mezzo trasmissivo è condiviso da ogni stazione;
- *Collision Detection*: le stazioni sono in grado di rilevare la presenza di collisioni dovute alla trasmissione simultanea, e reagire di conseguenza.

**IMAP**: *Internet Message Access Protocol*, permette l'accesso ad un server mail e di manipolare i messaggi come se si stesse lavorando in locale, in altre parole non è necessario scaricare i file. Questo tipo di accesso offre maggiore elasticità rispetto al POP (*Post Office Protocol*) che è stato pensato per l'uso offline, nel senso che i messaggi vengono scaricati e quindi cancellati dal server. Infatti è il metodo migliore se si ha un solo account e più computer che vi accedono (per esempio un notebook ed un PC), in altre parole permette di condividere le risorse del server.

**Infiniband**: protocollo di comunicazione ad alta velocità per l'interconnessione fra un processore e diversi nodi.

**Information Index (II)**: grid element che colleziona in un database le informazioni relative allo stato delle risorse di una Grid.

**Internetworking**: tecnologie di interconnessione tra LAN diverse.

**Intranet**: rete locale (LAN), o raggruppamento di reti locali, usata all'interno di una organizzazione per facilitare la comunicazione e l'accesso all'informazione, che può essere ad accesso ristretto. A volte il termine è riferito solo alla rete di servizi più visibile, il sistema di siti che formano uno spazio web interno. Ad essa si accede in genere tramite autenticazione.

**IP**: *Internet Protocol*, protocollo di rete su cui si basa la rete Internet. IP è un protocollo di rete a pacchetto; secondo la classificazione ISO/OSI è di livello rete (3). IP è un protocollo di interconnessione di reti (Inter-Networking Protocol), nato per interconnettere reti eterogenee per tecnologia, prestazioni, gestione.

**IPSec**: IPsec è l'abbreviazione di IP Security ed è uno standard per ottenere connessioni basate su reti IP sicure. La sicurezza viene raggiunta attraverso la cifratura e l'autenticazione dei pacchetti IP. La sicurezza viene fornita, quindi, a livello di rete. La capacità di fornire protezione a livello di rete rende questo protocollo trasparente al livello delle applicazioni che non devono essere modificate.

**iSCSI**: protocollo che permette di inviare comandi a dispositivi di memoria SCSI fisicamente collegati a server e/o altri dispositivi remoti. È un protocollo molto utilizzato in ambienti SAN poiché permette di consolidare l'archiviazione dei dati su dispositivi virtuali, collegati attraverso la rete, dando l'illusione di disporre localmente di un disco fisico che invece si trova in realtà su un dispositivo di storage remoto. A differenza del protocollo Fibre Channel consente l'impacchettamento su TCP/IP rendendo così possibile l'utilizzo dell'infrastruttura di rete esistente che rende di fatto possibile l'utilizzo su distanze maggiori.

**ISO**: *International Organization for Standardization*, nata nel 1947 a Londra, questa organizzazione a partecipazione volontaria presiede alla regolamentazione degli standard internazionali riguardanti molteplici settori. Nel 1978 l'ISO propose OSI, un modello di riferimento a 7 livelli per sistemi di rete diversi.

**JDL**: *Job Description Language*, è un linguaggio basato su espressioni, con cui è possibile specificare le caratteristiche di un job (parametri di input/output, requisiti di sistema, jobmanager da utilizzare ed altri attributi) in maniera da facilitare la ricerca, da parte del Resource Broker, delle corrispondenze tra requisiti e disponibilità di risorse.

**Jobmanager**: è un software che ha il compito di gestire le risorse di un singolo computer o di un cluster, scegliendo come e quando eseguire un job in base alle risorse disponibili e a una serie di regole definite dall'amministratore di sistema.

**Kernel**: la parte principale del sistema operativo. Viene caricato in memoria subito dopo il BIOS e si occupa della gestione del processore e del trasferimento dei dati fra le componenti del sistema fornendo ai processi in esecuzione sull'elaboratore un accesso sicuro e controllato all'hardware.

**Kerberos**: protocollo di rete per l'autenticazione tramite crittografia che permette a diversi terminali di comunicare su una rete informatica insicura, provando la propria identità tramite lo scambio di informazioni cifrate. Kerberos previene l'intercettazione e i replay attack, e assicura l'integrità dei dati. I suoi progettisti mirarono soprattutto ad un modello client-server, per fornire una mutua autenticazione: sia l'utente che il fornitore del servizio possono verificare l'identità dell'altro.

**LAN**: *Local Area Network*, Rete o gruppo di segmenti di rete confinati in un edificio o un campus, che collega computer e periferiche (es. stampanti, fax, scanner) installate nella stessa sede (es. stesso palazzo, anche a piani diversi) oppure in sedi vicine (es. due palazzi adiacenti).

**Larghezza di banda**: capacità di .trasporto dei dati. di un collegamento di rete utilizzata per indicare la velocità di trasmissione. Per esempio, un collegamento Ethernet è in grado di trasportare 10 Mbps (10 milioni di bit al secondo); un collegamento 100 Mbps (100 milioni di bit al secondo); 1.000 Mbps (1 miliardo di bit al secondo).

**LCFG**: *Local ConFiGuration System*, è un tool, con architettura client/server, per l'installazione automatica, configurazione e management centralizzato di macchine con sistema operativo Linux.

**LDAP**: *Lightweight Directory Application Protocol*, protocollo standard per la consultazione di grossi database distribuiti.

**Linux**: è un sistema operativo Unix che ha, tra le sue caratteristiche, quello di essere opensource, multipiattaforma, multiutente e supportato, a livello mondiale, da una vasta comunità di utilizzatori e sviluppatori.

**LLC**: *Logical Link Control*, protocollo di comunicazione per reti di computer, che fa parte della famiglia IEEE 802. È definito dal gruppo di lavoro IEEE 802.2.

**Logging & Bookkeeping**: grid element che colleziona in un database le informazioni relative allo stato dei job in esecuzione all'interno di una Grid.

**MAC address**: *Media Access Control Address*, Indirizzo di livello 2 di un nodo.

**Malware**: software creato con il solo scopo di causare danni più o meno gravi al computer su cui viene eseguito. Il termine deriva dalla contrazione delle parole inglesi malicious e software e ha dunque il significato letterale di "programma malvagio"; in italiano è detto anche codice maligno.

**MAN**: *Metropolitan Area Network*, Rete che si estende sull'area metropolitana. Ricopre quindi un'area geografica maggiore di un LAN ma minore di una WAN. Vedi anche LAN e WAN.

**MDS**: *Meta Directory Service*, sinonimo di GIS.

**MPI**: *Message Passing Interface*, è un set di librerie standard per lo sviluppo di applicazioni di calcolo parallelo.

MTBF: *mean time between failures* (tempo medio fra i guasti), è un parametro di qualità applicabile a dispositivi meccanici, elettrici ed elettronici e ad applicazioni software. Il MTBF è il valore atteso del tempo tra un guasto ed il successivo; la sua misura ha importanza, ad esempio, per la valutazione della vita media di un dispositivo elettronico, o di un componente meccanico.

**Multicast**: distribuzione simultanea di informazione verso un gruppo di destinatari.

**NAS**: *Network Attached Storage*, server dedicato esclusivamente a funzione di file server per una rete, attrezzato con dischi rigidi generalmente in tecnologia RAID. Viene progettato ed equipaggiato solo per questo scopo, evitando ogni altro hardware e software.

**NAT**: *network address translation* (traduzione degli indirizzi di rete), è una tecnica che consiste nel modificare gli indirizzi IP dei pacchetti in transito su un sistema che agisce da router.

**Netbeui** o **NetBIOS**: *Network Basic Input Output System*, software sviluppato da IBM e Sytek nel 1984, per consentire ai personal computer di accedere ad una rete.

**NFS**: *Network File System*, è uno dei protocolli standard, basato su un'architettura client/server, per la condivisione di file tra due o più sistemi Unix.

**OSI**: standard stabilito nel 1978 dall'International Organization for Standardization (ISO), il principale ente di standardizzazione internazionale, che stabilisce una pila di protocolli in 7 livelli per le reti di calcolatori.

**Overhead**: È la quantità di operazioni aggiunte all'esecuzione di un programma da una qualsiasi infrastruttura (per motivi di controllo o di gestione). Ha come effetto generale una perdita di performance del programma.

**Pacchetto**: sequenza di dati trasmessa su una rete, con un'intestazione (header) che può indicare il contenuto del pacchetto e il suo destinatario. Si può immaginare un pacchetto come una busta di dati con l'header che contiene l'indirizzo del mittente e del destinatario. I pacchetti vengono inviati ai destinatari attraverso le reti utilizzando la modalità "packet switching".

**PAP**: *Password authentication protocol*, è un protocollo di rete che richiede che l'utente si faccia riconoscere inviando, in chiaro, al server sia l'identificativo utente che la password.

**Patch**: nel contesto informatico il termine, che letteralmente significa "pezza", indica un programma il cui scopo è quelle di apportare delle modifiche ad un prodotto software esistente al fine di migliorarne l'affidabilità e l'efficienza. Solitamente le modifiche consistono nella risoluzione di bug e nell'aggiunta di funzionalità.

**PBS**: *Portable Batch System*, è un jobmanager che si occupa dello scheduling e dell'allocazione (statica) dei processi in un cluster locale di macchine fornendo degli strumenti per la gestione di code, accounting delle risorse di calcolo ed elaborazioni di tipo batch.

**PEAP**: Protected EAP.

**Ping**: *Packet InterNet Groper*, ossia il cercatore di pacchetto di Internet. È il metodo più semplice per testare e per avere i tempi di risposta delle connessioni IP. Il Ping manda una richiesta ad un host ed aspetta una risposta. Il ping ritorna i millisecondi necessari per il *round trip time*.

**PKI**: *Public-Key Infrastructure* (infrastruttura a chiave pubblica), è una serie di accordi che consentono a terze parti fidate di verificare e/o farsi garanti dell'identità di un utente, oltre che di associare una chiave pubblica a un utente, normalmente per mezzo di software distribuito in modo coordinato su diversi sistemi. Le chiavi pubbliche tipicamente assumono la forma di certificati digitali.

**Policy**: Una policy, è una serie di regole che stabiliscono il comportamento di un sistema di security, queste regole servono a chiarire in modo univoco chi, quando e come è autorizzato ad accedere ad un sistema informatico.

**PPP**: *Point-to-Point Protocol*, protocollo di rete di Livello datalink, comunemente usato nello stabilire connessioni dirette tra due nodi. PPP compone spesso il livello

datalink (il livello di collegamento dati) del modello OSI nelle connessioni su circuiti punto-punto sincronizzati e non sincronizzati, soprattutto in ambito WAN

**RADIUS**: *Remote Authentication Dial-In User Service*, è un protocollo AAA (*authentication, authorization, accounting*) utilizzato in applicazioni di accesso alle reti o di mobilità IP. RADIUS è attualmente lo standard de-facto per l'autenticazione remota, prevalendo sia nei sistemi nuovi che in quelli già esistenti.

**RAID**: *Redundant Array of Independent Disks*, è un sistema informatico che usa un insieme di dischi rigidi per condividere o replicare le informazioni. I benefici del RAID sono di aumentare l'integrità dei dati, le prestazioni e la tolleranza ai guasti, rispetto all'uso di un disco singolo.

**RAM**: *Random-Access Memory* (memoria ad accesso casuale), è una tipologia di memoria informatica caratterizzata dal permettere l'accesso diretto a qualunque indirizzo di memoria con lo stesso tempo di accesso. È una memoria di lettura e scrittura estremamente veloce, ma ha la caratteristica di essere volatile, ovvero di perdere le informazioni con il reset del computer.

**ROM**: *Read-Only Memory* (memoria a sola lettura), è una tipologia di memoria informatica, in particolare una tipologia di memoria non volatile (memoria informatica in grado di mantenere memorizzati i dati anche se non è alimentata elettricamente) in cui i dati sono memorizzati nella sua fase di costruzione e non possono essere più modificati per l'intera durata della sua vita.

**Router**: dispositivo di rete che si occupa di instradare pacchetti informativi. La caratteristica fondamentale dei router è che la funzione di instradamento è basata sugli indirizzi di livello 3 (rete) del modello OSI (corrispondente al livello IP dello stack TCP/IP), a differenza dello switch che instrada sulla base degli indirizzi di livello 2 (collegamento) "MAC". Gli elementi della tabella di instradamento (o Routing Table) sono intere reti, ovvero sottoinsiemi anche molto ampi dello spazio di indirizzamento. Questo è fondamentale per la scalabilità, in quanto permette di gestire reti anche molto grandi facendo crescere le tabelle di instradamento in modo meno che lineare rispetto al numero di host.

**RSA**: il più diffuso sistema di cifratura. Il suo nome deriva dalle iniziali dei suoi 3 inventori: Ronald River, Adi Shamir e Leonard Adleman. Fu inventato nel 1978. Algoritmo di crittografia asimmetrica, utilizzabile anche per cifrare o firmare informazioni.

**Replica Catalogue (RC)**: grid element che gestisce un database di informazioni relative ai file contenuti sugli Storage Element di una grid.

**Resource Broker (RB)**: è un grid element che ha il compito di trovare una corrispondenza tra i requisiti espressi dagli utenti per l'esecuzione dei propri job e le risorse disponibili sulla griglia, utilizzando opportuni algoritmi di scheduling.

**SAN**: *Storage Area Network*, rete o parte di una rete ad alta velocità (generalmente Gigabit/s) costituita esclusivamente da dispositivi di memorizzazione di massa, in

alcuni casi anche di tipologie e tecnologie differenti. Il suo scopo è quello di rendere tali risorse di immagazzinamento (storage) disponibili per qualsiasi computer connesso ad essa. I protocolli attualmente più diffusi, usati per la comunicazione all'interno di una SAN, sono FC (Fibre Channel) ed iSCSI (Internet SCSI). Una SAN può essere definita come: "una rete il cui scopo principale è il trasferimento di dati tra sistemi di computer ed elementi di storage e tra elementi di storage. Una rete SAN consiste in un'infrastruttura di comunicazione, che fornisce connessioni fisiche e in un livello di gestione, che organizza connessioni, elementi di storage e sistemi di computer in modo da garantire un trasferimento di dati sicuro e robusto".

**SATA**: *Serial Advanced Technology Attachment*, consiste in un'interfaccia standard per la connessione di dispositivi di memorizzazione quali hard disk e unità CD-ROM all'interno dei personal computer. Il Serial ATA è l'evoluzione dell'ATA (anche conosciuto come IDE), rinominato Parallel ATA (PATA) in seguito alla nascita del Serial ATA in modo da evitare fraintendimenti, rispetto al quale il Serial ATA presenta tre principali vantaggi: maggiore velocità, cavi meno ingombranti e possibilità di hot swap.

**SCSI**: *Small Computer System Interface*, è un'interfaccia standard progettata per realizzare il trasferimento di dati su bus, tra computer e periferiche di memorizzazione. Generalmente utilizzato su sistemi professionali di vecchia generazione. L'evoluzione di SCSI e' il Serial Attached SCSI (SAS).

**SAS**: *Serial Attached SCSI*, è una tecnologia di trasferimento dati studiata per lavorare sia con dispositivi ad accesso diretto, come i dischi fissi, sia per quelli ad accesso sequenziale, come i nastri magnetici. Il protocollo di comunicazione è seriale punto-punto diversamente dal bus SCSI parallelo introdotto nella metà degli anni '80. Nella tecnologia SCSI, la velocità del bus è condivisa fra tutti i dispositivi e il controller stesso, mentre nella tecnologia SAS è dedicata a ciascun dipositivo. Il set di comandi di SAS e mutuato esattamente da SCSI, così la completa compatibilità, a livello di comando, per tutti i software sviluppati per SCSI è garantita.

**Security policy**: insieme di regole che specificano quale rete o suo elemento è abilitata a comunicare con quali reti o elementi di rete.

**SDK**: *Software Development Kit*, è un termine che in italiano si può tradurre come "pacchetto di sviluppo per applicazioni", e sta a indicare un insieme di strumenti per lo sviluppo e la documentazione di software

**Server**: computer che fornisce servizi ad altre componenti (tipicamente chiamati client) attraverso una rete di trasmissione dati. Si noti che il termine server, così come pure il termine client, possono essere riferiti sia alla componente software che alla componente hardware che forniscono le funzionalità o servizi di cui sopra.

**Sistema Operativo**: insieme di programmi base che costituiscono l'interfaccia tra l'utente e l'hardware. Il sistema (OS) nasconde l'hardware non solo all'utente ma anche ai programmi. Permette una gestione ottimizzata tramite File System, dell'Hard Disk, incluse le seguenti risorse: CPU, memoria, memoria di massa, periferiche, etc.. Il sistema operativo organizza l'esecuzione dei programmi, fornendo i servizi

fondamentali su cui questi poggiano. Facilita fondamentalmente l'utente nel compito della programmazione.

**SLA**: *Service Level Agreement*, strumenti contrattuali attraverso i quali si definiscono le metriche di servizio che devono essere rispettate da un fornitore di servizi

**Smart card**: tipo di carta magnetica che utilizza un chip integrato nella carta stessa. Può registrare dati di qualsiasi tipo (nomi, indirizzi, numeri di telefono, importi, codici, password...) che, con l'inserimento in un lettore, può rendere disponibili.

**SMP**: architettura hardware dotata di più processori e dove questi possono accedere equamente a tutta la memoria RAM tramite un backplane basato su bus o meglio switched.

**SMTP**: *Simple Mail Transfer Protocol*, protocollo per lo scambio di messaggi di posta elettronica in rete TCP/IP (Internet, Intranet, Extranet, LAN). Progettato per messaggi in ASCII puro, attraverso le estensioni MIME può gestire messaggi formattati ed allegati grafici, audio e video, multimediali. È descritto nella RFC 821.

**SNMP**: *Simple Network Management Protocol*, Protocollo standard di Internet per la gestione di reti IP e dei dispositivi collegati, definito dalle specifiche STD 15 e RFC 1157.

**Socket**: canale logico su cui viene trasmesso un flusso di dati tra due computer in rete.

**Spam**: invio indiscriminato di grosse quantità di messaggi di posta elettronica, tramite Internet, a lunghi elenchi di indirizzi, spesso a scopo pubblicitario e commerciale. Molti server di posta elettronica hanno una funzione anti-spamming, che ne blocca l'inoltro a richiesta dell'utente.

**SSH**: *Secure Shell*, servizio per i login e l'amministrazione remota di server che utilizza le librerie SSL per la protezione in crittografia della connessione.

**SSID**: *Service Set Identifier*, nome con cui una rete Wi-Fi si identifica ai suoi utenti. Spesso gli access point sono configurati in modo da annunciare continuamente i loro SSID, cosicché i dispositivi Wi-Fi possano creare un elenco delle reti disponibili nella zona in cui si trovano. Tale elenco può poi essere mostrato all'utente perché scelga la rete a cui connettersi.

**SSL**: *Secure Socket Layer*, protocollo sviluppato da Netscape operante nel Transport Layer (Livello Trasporto), che consente, grazie a tecniche crittografiche, il trasferimento di dati tramite la rete Internet in modo sicuro, ovvero senza che possano essere "acquisiti" da enti non autorizzati.

**Storage Element (SE)**: è un nodo di una grid che fornisce i servizi necessari a immagazzinare, localizzare e replicare i dati. Un SE, inoltre, fornisce ad altri nodi della grid le informazioni relative alla disponibilità dei dati.

**Subnet**: sottoinsieme di una rete locale, facente parte della stessa rete ma distinta per limitazioni nell'uso di risorse e nell'accesso di utenti. I computer appartenenti ad una subnet vengono definiti dalla subnet mask.

**Subnet mask**: sequenza di bit che distingue quale porzione di un indirizzo IP identifica la (sotto)rete e quale l'host. È necessaria ad un dispositivo sorgente che vuole comunicare con un secondo dispositivo destinatario per decidere se spedire direttamente all'indirizzo IP del destinatario o se spedire i pacchetti tramite il router della propria rete locale.

**Swap**: estensione della capacità della memoria volatile complessiva del computer, oltre il limite imposto dalla quantità di RAM installata, attraverso l'utilizzo di uno spazio su un altro supporto fisico, ad esempio il disco fisso. A seconda del sistema operativo utilizzato è possibile avere file di swap (chiamato anche 'Memoria virtuale'), residenti nel normale file system del sistema, oppure partizioni di swap, cioè sezioni di disco integralmente dedicate allo swap ed inizializzate con un proprio specifico tipo di file system.

**Switch**: dispositivo hardware multiporta utilizzato per lo scambio di frame tra due o più calcolatori su una rete locale.

**TCP**: *Transmission Control Protocol*, è un protocollo di livello di trasporto della suite di protocolli Internet. È definito nella RFC 793, e su di esso si appoggiano gran parte delle applicazioni Internet. Il TCP può essere classificato al livello trasporto (OSI level 4). È utilizzato per la trasmissione di dati, in applicazioni che richiedono garanzia di ricezione dei pacchetti.

**TCP/IP**: *Transmission Control Protocol/Internet Protocol*, è un insieme di protocolli standard utilizzati su internet per lo scambio di dati tra più computer di una rete.

**Telnet**: protocollo di comunicazione che consente ad un personal computer di collegarsi ad un server remoto emulando un terminale client. Telnet è stato sviluppato per la rete ARPAnet con la RFC854, ma in seguito è stato incorporato nel protocollo TCP/IP e quindi è utilizzabile anche con le connessioni Internet. L'obiettivo del protocollo TELNET è fornire un supporto per le comunicazioni sufficientemente generalizzato, bidirezionale ed orientato ai byte (otto bit). È solitamente utilizzato per fornire all'utente sessioni di login remoto di tipo linea di comando tra host su internet.

**TFTP**: *Trivial FTP*, è una versione limitata del protocollo FTP standard che non offre, ad esempio, nessun meccanismo di autenticazione.

**TKIP**: *Temporal Key Integrity Protocol*, protocollo di sicurezza implementato nelle specifiche dello standard IEEE 802.11i. Prevede un sistema di ricombinazione della chiave di sicurezza per ciascun pacchetto di dati in transito, oltre ad un controllo di integrità del singolo pacchetto. Fa parte dello standard di sicurezza WPA

**TLS**: *Transport Layer Security*, TLS e il suo predecessore *Secure Sockets Layer* (SSL) sono dei protocolli crittografici che permettono una comunicazione sicura e una integrità dei dati su reti TCP/IP come, ad esempio, internet. TLS e SSL cifrano la comunicazione dalla sorgente alla destinazione (end-to-end) sul livello di trasporto.

**Tool kit**: è un insieme di pacchetti software, facilmente estendibili, che svolgono compiti specifici e possono essere utilizzati separatamente o in maniera congiunta.

**TSM**: *Tivoli Storage Manager*, sistema s/w di classe enterprise per la gestione centralizzata del backup e dell'archiviazione dei dati. Il software abilita l'utente ad inserire dati sia attraverso la funzionalità di *backup* sia attraverso la funzionalità di *archive* e analogamente consente il *restore* e il *retrieve*.

**TTLS**: *Tunneled Transport Layer Security*, estensione del TLS, sviluppato per superare la necessità di certificati lato client (sono invece richiesti certificati lato server).

**UDP**: User Datagram Protocol, protocollo standard della suite di protocolli IP, utilizzato per la trasmissione di dati, in applicazioni che non richiedono garanzia di ricezione dei pacchetti.

**Unicast**: un pacchetto destinato ad un solo computer (l'indirizzo usato per inviare tale pacchetto, è detto Unicast).

**UNINFO**: è il rappresentante italiano nell'ISO per le tematiche di reti di calcolatori.

**User Interface (UI)**: è un nodo di una grid a cui gli utenti si collegano per sottomettere i propri job. Una UI offre agli utenti un set di comandi e un ambiente testuale, grafico o di tipo web per la sottomissione e il management dei job.

**Virtual Organization (VO)**: gruppo di persone ed entità geograficamente distribuite, multi-istituzionali, in dinamica evoluzione e con interessi comuni, scientifici, economici o amministrativi. Un server VO è un grid element che colleziona in un database l'elenco e i membri delle Virtual Organization di una Grid.

**Virus**: Nell'ambito dell'informatica un virus è un software, appartenente alla categoria dei malware, che è in grado, una volta eseguito, di infettare dei file in modo da riprodursi facendo copie di sé stesso, generalmente senza farsi rilevare dall'utente. I virus possono essere o non essere direttamente dannosi per il sistema operativo che li ospita, ma anche nel caso migliore comportano un certo spreco di risorse in termini di RAM, CPU e spazio sul disco fisso.

**VLAN**: *Virtual Local Area Network*, insieme di tecnologie che permettono di segmentare il dominio di broadcast, che si crea in una rete locale (tipicamente IEEE 802.3) basata su switch, in più reti non comunicanti tra loro

**VPN**: *Virtual Private Network*, rete privata instaurata tra soggetti che utilizzano un sistema di trasmissione pubblico e condiviso come, per esempio, Internet. Lo scopo delle reti VPN è di dare alle aziende le stesse possibilità delle linee private in affitto ad un costo inferiore sfruttando le reti condivise pubbliche. Consente agli utenti di una rete di accedere alle sue risorse anche da postazioni remote, poste al di fuori della rete. I dati fra la postazione remota e server della VPN vengono inoltrati con la protezione di sistemi di cifratura del dei dati come ad esempio: SSL, IPSEC, etc..

**WAN**: *Wide Area Network*, tipologia di rete di computer che si contraddistingue per avere un'estensione territoriale pari a una o più regioni geografiche (quindi superiore sia a quella della rete locale che a quella della rete metropolitana). La WAN può connettere fra loro più reti locali e/o metropolitane. La più grande WAN mai realizzata, Internet, è una rete di computer ad accesso pubblico che copre l'intero pianeta.

**Web Server**: servizio che si occupa di fornire, su richiesta del browser, files di qualsiasi tipo oppure pagine web (spesso scritte in HTML). Le informazioni inviate dal server web viaggiano in rete trasportate dal protocollo HTTP. L'insieme di server web dà vita al World Wide Web, uno dei servizi più utilizzati di Internet.

**WDM**: *Wavelength Division Multiplexing*, un tipo di multiplazione utilizzato nei sistemi di comunicazione ottica. Per modulare diversi canali su una stessa fibra ottica si usano diverse portanti di differenti lunghezze d'onda, una per ogni canale, e per la singola portante si usa la modulazione di intensità. In questo modo è possibile sfruttare la grande banda ottica disponibile.

**WEP**: *Wireless Encryption Protocol* o *Wired Equivalent Privacy*, parte dello standard IEEE 802.11 (ratificato nel 1999) e in particolare è quella parte dello standard che specifica il protocollo utilizzato per rendere sicure le trasmissioni radio delle reti Wi-Fi.

**Wi-Fi**: *Wireless Fidelity*, è un termine che indica dispositivi che possono collegarsi a reti locali senza fili (WLAN) basate sulle specifiche IEEE 802.11.

**Wireless**: letteralmente “senza fili”, si intende una comunicazione, per lo scambio di dati, senza l’utilizzo di cavi attraverso la tecnologia radio IEEE 802.11 o Bluetooth.

**Workstation**: tipologia di computer contraddistinto dall'essere general purpose (cioè non destinato a specifici compiti), monoutente (cioè utilizzabile da un solo utente alla volta), destinato principalmente ad un utilizzo produttivo (da cui il suffisso work), e dall'avere alte prestazioni per poter assolvere compiti altamente professionali di vario genere.

**Worker Node (WN)**: è un nodo generico che offre potenza di calcolo. Un WN può essere considerato ad esempio come un elemento di una farm locale, o un computer che può offrire, come unico servizio, la sua cpu per l’esecuzione di job.

**World Wide Web (WWW)**: rete mondiale, sistema di strutturazione dell'informazioni e delle risorse in modalità ipertestuale con visualizzazione di immagini, filmati, suoni. Questo efficiente servizio creato al CERN di Ginevra da Tim Berners Lee è uno dei responsabili dell'esplosione di Internet.

**Worm**: letteralmente in inglese significa "verme". Particolare categoria di malware in grado di autoreplicarsi. È simile ad un virus, ma a differenza di questo non necessita di legarsi ad altri eseguibili per diffondersi.

**WPA**: *Wi-Fi Protected Access*, specifiche di sicurezza per la connessione wireless appartenenti allo standard IEEE 802.11i.

**X.509**: Si tratta di uno standard ISO che definisce un sistema di certificazione usato per rendere sicure le comunicazioni elettroniche. È uno standard per le infrastrutture a chiave pubblica (PKI). X.509 definisce, fra le altre cose, formati standard per i certificati a chiave pubblica.

**X.25**: Protocollo di rete standard adottato dal CCITT. Opera su rete a "commutazione di circuito", è utilizzato soprattutto su reti WAN, reti pubbliche telefoniche e Frame Relay.

**XML**: *Extensible Markup Language*, è un linguaggio di codifica estensibile, e rappresenta un formato universale per dati e documenti strutturati sul Web.

# Bibliografia


**Capitolo 1:**

1. http://www.lnf.infn.it/
2. http://www.infn.it/
3. http://www.infn.it/lhcitalia/
4. http://www.asimmetrie.it/
5. http://www.cern.ch/
6. http://www.lnf.infn.it/computing/

**Capitolo 2:**

7. S. Gai, P. Nicoletti, P. L. Montessori
   Reti locali. Dal cablaggio all'internetworking – Telecom Italia
8. Andrew S. Tanenbaum
   Reti di Computer – Prentice Hall International
9. Lydia Parziale, David T. Britt, Chuck Davis, Jason Forrester, Wei Liu, Carolyn Matthews, Nicolas Rosselot
   TCP/IP Tutorial and Technical Overview – ibm.com/redbooks, Dicembre 2006
   http://www.redbooks.ibm.com/redbooks/pdfs/gg243376.pdf
10. http://www.ieee.org/
11. http://www.iso.org/
12. http://www.iana.org/
13. http://www.ietf.org/rfc.html
14. http://www.cisco.com/

**Capitolo 3:**

15. SAN Security: A Best Practices Guide, Revision 2 – Brocade Communications Systems, Inc.
16. Secure SAN Zoning Best Practices – Brocade Communications Systems, Inc.

17. Jon Tate, Brian Cartwright, John Cronin, Christian Dapprich
IBM SAN Survival Guide – ibm.com/redbooks, Agosto 2003
http://www.redbooks.ibm.com/redbooks/pdfs/sg246143.pdf

18. Jon Tate, Jim Kelly, Pauli Ramo, Leos Stehlik
IBM TotalStorage: Introduction to SAN Routing – ibm.com/redbooks, Maggio 2006

19. A. Blank, P. Kiefer, C. Sallave Jr., G. Valencia, J. Wain, A. M. Warda
Advanced POWER Virtualization on IBM System p5 – ibm.com/redbooks, Novembre 2005

**Capitolo 4:**

20. Daniele Giacomini
Appunti di Informatica Libera, 8 Aprile 2010
http://appuntilinux.mirror.garr.it/mirrors/appuntilinux/a2/a2.pdf

21. http://www.isc.org/software/dhcp

22. http://www.isc.org/software/bind

23. http://web.mit.edu/kerberos/

24. http://www.freeradius.org/

25. http://www.openafs.org/

26. Charlotte Brooks, Peter McFarlane, Norbert Pott, Martin Trcka, Eduardo Tomaz
Tivoli Storage Management Concepts – ibm.com/redbooks, Maggio 2006
http://www.redbooks.ibm.com/redbooks/pdfs/sg244877.pdf

27. http://www.apache.org/

28. http://www.sendmail.org/

29. http://www.dovecot.org/

30. http://www.sophos.com/products/enterprise/email/

31. http://technet.microsoft.com/it-it/library/dd349801(WS.10).aspx

32. http://technet.microsoft.com/it-it/library/cc732488(WS.10).aspx

33. http://technet.microsoft.com/it-it/library/cc731636(WS.10).aspx

**Capitolo 5:**


34. Ian Foster
    What is the Grid? A Three Point Checklist, GRIDToday – July 20, 2002
    http://www-fp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf

35. I. Foster, C. Kesselman, S. Tuecke
    The Anatomy of the Grid: Enabling Scalable Virtual Organizations – International J. Supercomputer Applications, 2001
    http://www.globus.org/alliance/publications/papers/anatomy.pdf

36. I. Foster
    The Grid: A New Infrastructure for 21st Century Science – Physics Today, 2002

37. I. Foster, C. Kesselman
    The Grid 2: Blueprint for a New Computing Infrastructure – Morgan Kaufmann; 2nd edition (November 18, 2003)


**Capitolo 6:**


38. S. Burnett, S. Paine
    RSA Security's Official Guide to Criptography – McGraw-Hill, 2001

39. Maximum Apache Security – Sams, 2002

40. Michael E. Whitman, Herbert J. Mattord, Richard Austin, and Greg Holden
    Guide to Firewalls and Network Security – Paperback, Giugno 2008

41. Richard A. Deal
    Cisco Router Firewall Security – Paperback, Agosto 2004

42. Greg Holden
    Guide to Firewalls and Network Security: Intrusion Detection and VPNs – Paperback, Aprile 2003

43. http://www.syslog.org/

44. http://www.nagios.org/