RACCOLTA ED ANALISI DI LOGS

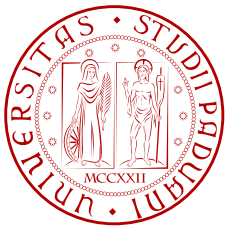
LEGGE ANTI TERRORISMO, AUDITING AMMINISTRATORI

RELATORE: **Ch.mo Prof. Filira Federico**

LAUREANDO: **Szabo Karoly Albert**

Corso di laurea triennale in Ingegneria Informatica

ANNO ACCADEMICO 2009-2010



UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
CORSO DI LAUREA IN INGEGNERIA INFORMATICA
TESI DI LAUREA

RACCOLTA ED ANALISI DI LOGS LEGGE ANTI TERRORISMO ED AUDITING AMMINISTRATORI

RELATORE: Ch.mo Prof. Filira Federico

LAUREANDO: *Szabo Karoly Albert*

Padova, 26 Marzo 2010

Indice

Sommario

La presente tesi descrive la mia esperienza di tirocinio svoltasi presso *ne-t by Telerete Nordest*, azienda di ICT che opera nel Padovano, che recentemente si sta espandendo oltre i confini della provincia.

La tesi si svilupperà seguendo l'andamento temporale dello stage, durante il quale mi sono occupato di varie mansioni, in particolare ho impiegato gran parte del tempo per adempiere ad obblighi legali quali: la legge anti terrorismo ed il provvedimento del Garante della Privacy per quanto riguarda l'Auditing degli amministratori.

Per poter gestire progetti così ampi e complessi ho deciso di utilizzare le nozioni di Project Management apprese, ma, nonostante questo approccio, ho spesso rivisto e ripianificato il mio operato, in quanto, alcune conoscenze teoriche che avevo, non erano sufficienti ad affrontare problematiche reali; al contrario, nella parte finale del mio stage a Telerete, grazie all'esperienza fatta ed alla familiarità acquisita con gli strumenti presenti in azienda, sono stato in grado di portare a termine progetti complessi restando entro i tempi previsti.

Nella prima parte verrà definita l'azienda ed il contesto organizzativo nel quale opererò, successivamente verranno descritte le problematiche che il primo progetto si proponeva di risolvere, e l'attuazione del progetto stesso.

La seconda parte tratta il rispetto delle cogenze di legge imposte dal provvedimento dal Garante della Privacy relativo agli amministratori di sistema, ovvero l'auditing degli amministratori. Verranno mostrate le soluzioni proposte e la soluzione effettivamente implementata.

Questa esperienza è stata, a mio parere, molto utile, sia come crescita professionale che personale, in quanto ho dovuto approfondire e collegare tra loro molti argomenti affrontati in questi anni di Università, che, nel mondo reale, si sono rivelati spesso strettamente correlati tra loro.

Capitolo 1

Introduzione

La mia attività di testi in azienda ha avuto principalmente due obbiettivi:

- Aggiustamento ed 'ingegnerizzazione' del sistema integrato per la gestione delle reti wireless sotto il controllo e la responsabilità di Telerete NordEst;
- La creazione di un sistema di supporto per l'evasione delle richieste provenienti dalla Polizia Postale;
- Rispetto delle cogenze di legge dettate dal Garante della Privacy, configurando quindi un sistema auditing interno.
- Far sì che i lavori proposti non risolvano unicamente i requisiti di legge, ma che ci sia un approccio migliorativo in tutto ciò che viene fatto:
  - Rilevamento Problematiche di rete
  - Centralizzazione degli accessi

A tal proposito, considerando che l'ambito operativo è quello di una realtà aziendale in fase di sviluppo ed espansione, si rende necessario il raggiungimento di una situazione in cui tutte queste risorse ed informazioni siano organizzate in maniera ottimale.

I dati rilevanti che l'azienda non può permettersi di perdere per una corretta gestione di rete, al pari dei dati che pur non essendo strettamente necessari potrebbero tornare utili all'azienda per collaborazioni e progetti futuri, dovranno essere immagazzinati in un Data Warehouse, una sorta di Database che ammette la presenza di informazioni ridondanti; tale ridondanza può essere infatti utilizzata per l'analisi incrociata dei dati, che permette di ottenere svariate informazioni

utili per l'azienda. Si cercherà di progettare il magazzino di dati con lo scopo di consentire di produrre facilmente relazioni ed analisi.

Il tutto dovrà permettere, da una parte una consultazione completa ed efficiente dei dati contenuti nel sistema, dall'altra parte l'inserimento automatizzato o semi-automatizzato dei dati di interesse nel sistema stesso. Tramite appositi script, la procedura di inserimento delle informazioni nel Data Warehouse potrà essere impostata secondo le esigenze dettate dalla specifica applicazione o dalla natura del servizio: in maniera tale da svolgersi automaticamente ogni periodo temporale di durata predeterminata; oppure in modo tale da poter essere azionata al volo ogniqualvolta l'operatore lo ritenga opportuno.

Durante la prima parte dello stage (circa una settimana) ho effettuato un periodo di ambientamento all'interno dell'azienda, in particolare nel settore tecnico, allo scopo di prendermi il tempo di carpire le filosofie e le dinamiche aziendali, e inserendomi così in maniera graduale all'interno del mondo lavorativo.

In questo periodo, mi sono preso il tempo necessario per documentarmi su alcuni argomenti informatici e di telecomunicazioni, colmando alcune lacune in materie che, con ogni probabilità, avrebbero avuto a che vedere con la mia attività di tesi in azienda. Oltre a ciò, ho studiato a fondo il problema che ero in procinto di affrontare, creando un documento tecnico che fornisca indicazioni complete sul progetto e del suo contesto.

Nel periodo di ambientamento mi sono documentato in particolare sulla configurazione e le funzionalità di Squid quale transparent proxy, sul funzionamento della distrubuzione Gentoo di Linux e sulla struttura della rete interna e delle reti riguardanti Padova Wifi e Monselice Wifi.

Come già detto il modulo 'Log Mining' è stato oggetto del mio stage nella fase iniziale. In questa parte del lavoro ho identificato problematiche di rete e le ho associate ai logs adatti, preseguendo il lavoro iniziato da altri stagisti prima di me, ho successivamente studiato e configurato il sistema di raccolta logs che contenessero i dati richiesti dalla legge anti terrorismo, fornendo, infine, un sistema di analisi dei log.

Il sistema relativo al log mining, fungerà anche da base per l'estrazione di informazioni utili per il secondo progetto sviluppato, la raccolta e la consultazione dei log relativi agli accessi degli amministratori di sistema, come definito dal Provvedimento del 27 novembre 2008 del Garante della Privacy.

Questo sistema è stato progettato per integrarsi al meglio nella struttura aziendali già presente, con l'obiettivo di avere un effetto migliorativo sulla routine degli aministratori oltre che ad ottemperare alle cogenze di legge, le quali impongono un tracciamento degli accessi ed un'identificazione di chi li ha compiuti.

Il progetto iniziale non è stato seguito completamente a causa di numerosi problemi riscontrati con macchine non aggiornabili perchè vitali, di necessità di sicurezza su alcune macchine e di vincoli temporali troppo stretti, si è dovuti ricorrere alla gestione del rischio, seguendo le linee guida dello studio di fattibilità, fatto prima della pianificazione del lavoro per identificare la soluzione ottima nel caso specifico ed eventuali vie di fuga in caso di problemi.

Durante i periodi 'morti' della mia esperienza in azienda, ovvero quando terminavo dei lavori prima della scadenza, mi sono state assegnate mansioni di routine, quali:

- il testing di apparati di telefonia VOIP e di rete, in particolare ho configurato un IPPBX e testato vari Codec per esso,
- la ricerca di informazioni sui servizi informatici/informativi presenti nel territorio,
- brevi ricerche di mercato riguardanti gli RFID e loro appilcazioni.

Capitolo 2

Presentazione Azienda ospitante

La tesi di laurea verrà svolta in azienda, presso Telerete NordEst. La sede centrale è situata a Padova, in zona industriale. L'azienda opera principalmente all'interno della provincia padovana, ma, recentemente, sta allargando la propria area lavorativa anche al di fuori dei confini della stessa, all'interno di un progetto di espansione aziendale.

Telerete NordEst mette la propria competenza e le proprie risorse a disposizione, tra gli altri, del Comune di Padova, della Provincia di Padova, di APS Holding, dell'Azienda Ospedaliera di Padova e dell'Università degli Studi di Padova. L'azienda dispone di numerose risorse, sia a livello di competenze umane e professionali, sia a livello tecnologico con un'ingente quantità di dispositivi hardware e componenti software.

Ne-t fornisce inoltre servizi di consulenza per ottimizzare i processi organizzativi e servizi a supporto dell'intero ciclo di vita della soluzione. Con una struttura di oltre 60 collaboratori, è in grado di assicurare assistenza post-vendita e servizi di manutenzione ai clienti in modalità 24h x 365. La missione è:

- Fornire una gamma di servizi completa che, attraverso l'utilizzo delle tecnologie più avanzate, rispondano alle necessità del tessuto urbano e delle aziende, sia pubbliche che private che in esso operano.

- Porsi come interlocutore privilegiato per chiunque abbia la necessità di ottenere in tempi rapidi risposte concrete a esigenze di natura tecnologica,

fornendo inoltre servizi accessori con il valore aggiunto della professionalità espressa dai singoli collaboratori.

- Affiancare gli Enti con il proprio know-how nello sforzo di offrire la maggiore accessibilità possibile ai servizi, sia attraverso l'uso di media diversi sia attraverso la connettività diffusa, con un'attenzione particolare alla tutela della sicurezza dei cittadini.

2.1 Soci

L'azienda è sotto la proprietà di altre 4 socità, tra queste, la più impostante, è APS Holding, la quale possiede, tra l'altro, anche delle quote di Infracom

- APS HOLDING - 50,123%
- PRONET - 38,120%
- INFRACOM - 8,044%
- CAMERA DI COMMERCIO - 3,713%

2.2 Servizi

L'azienda offre svariati servizi ai propri clienti, alle azienda partner e in buona parte dedicati al territorio, è opportuno classificare i diversi servizi forniti dall'azienda Telerete.

2.2.1 Tassonomia dei servizi offerti

- Progetti di integrazione tecnologica in ambito urbano: progettazione ed implementazione di tecnologie urbane e progetti integrati, tra cui la rete del metrobus di Padova (nell'ambito del progetto SAE - Sistema di Ausilio all'Esercizio)
- Videosorveglianza, per conto di Carabinieri, Polizia Municipale e Questura:
  - Progetto "Padova città sicura" - Via Anelli
  - Progetto "Padova città sicura" - Palazzo Moroni
  - Consorzio Padova Ovest
  - Pensiline alle fermate del metrobus padovano
  - Stadio Euganeo
  - ZTL (Zone a Traffico Limitato)
- Installazione infrastrutture e gestione di rete per servizi di connettività wireless:
  - Padova WiFi
  - Unipd WiFi
  - Monselice WiFi
- Fornitura di connettività: MAN cittadina (protocollo HiperLAN), apparecchiature tra cui hotspot e antenne WiFi, gestione di postazioni co-siting per antenne di telefonia mobile, gestione della rete in fibra ottica e disaster recovery
- Fornitura di servizi ISP: connessioni ISDN, DSL, wireless, fibra ottica, Hosting, Housing, servizi di sicurezza, registrazione domini, servizi DNS, server farm e disaster recovery
- Servizi informatici:
  - Sviluppo ed installazione software
  - Sviluppo website, servizi ed applicazioni web
  - Servizi avanzati di gestione anagrafica animale
- Call center: servizi informativi per clienti, servizi di CRM, assistenza e Help Desk, booking di eventi culturali

- Campagne pubblicitarie di eventi nel padovano
- Anagrafe Canina (e di altri anumali) di varie città
  - Padova
  - Trento
  - Roma
- Portali e applicazioni informatiche di supporto all'E-Government e per le aziende
- Fornitura hardware di supporto

2.3 System Integration

L'integrazione ha lo scopo di creare collegamenti e mettere in comunicazione tra loro vari sistemi, servizi, applicazioni che risiedono su sistemi operativi differenti, si appoggiano a database e DBMS differenti, utilizzano linguaggi differenti. L'obiettivo finale dell'integrazione è la crezione di una piattaforma in cui i vari sotto-sistemi funzionino sia presi singolarmente sia presi globalmente assieme all'intero contesto: all'utente, il tutto dovrebbe apparire in maniera trasparente come un unico sistema.

I metodi di integrazione sono essenzialmente 3:

1. **Integrazione verticale**: è il processo di integrazione tra i vari sottosistemi in accordo con le loro funzionalità, creando entità funzionali spesso chiamate con il termine *silos*. Il beneficio di questo metodo consiste nel fatto che l'integrazione viene raggiunta in maniera rapida e coinvolge solo un certo numero di sottosistemi necessari; il costo nel breve periodo è quindi più contenuto. Dall'altra parte però il costo di mantenimento è più elevato, dal momento che in caso vengano richieste funzionalità aggiuntive, l'unico modo possibile consiste nell'implementare un altro silo apposito.

2. **Integrazione a stella**: è anche nota come *Star Integration* o *Spaghetti Integration*. E' il processo di integrazione in cui ciascun sottosistema è interconnesso con ognuno dei rimanenti sottosistemi. Il costo è variabile e dipende fortemente dalle tipologie di interfacce che mettono a disposizione le varie componenti di sistema. I tempi ed i costi necessari per integrare il sistema con nuove funzionalità e nuove componenti crescono in maniera

esponenziale nel numero di sottosistemi. Nel caso siano necessari un numero limitato di componenti od un numero limitato di interconnessioni tra essi, il metodo è efficace e si rivela estremamente flessibile e riutilizzabile nell'insieme di funzionalità.

3. **Integrazione orizzontale**: un nome alternativo è Enterprise Service Bus (ESB). E' un metodo di integrazione in cui un sottosistema specializzato viene dedicato esclusivamente a realizzare la comunicazione tra gli altri sottosistemi. Ciò permette un notevole taglio nel numero di connessioni (interfacce) da realizzare: è infatti sufficiente prevederne una per ciascun sottosistema, la quale lo connette direttamente all'ESB. Quest'ultimo è in grado di tradurre un'interfaccia in un'altra. Il costo di integrazione viene sensibilmente ridotto ed il grado di flessibilità del sistema è elevato. Anche i costi di mantenimento e aggiornamento restano limitati: per esempio la sostituzione di una componente con un'altra affine richiede al più di definire una nuova interfaccia (tra il nuovo modulo e l'ESB), in maniera del tutto trasparente al resto del sistema.

Ulteriori informazioni sono reperibili all'URL:
http://en.wikipedia.org/wiki/System_integration.
Attualmente in azienda le varie piattaforme di integrazione presenti sono utili ad ottimizzare compiti relativi a:

- Networking, Connettività
- servizi di Web-call, call-center, ticketing
- Info-mobilità
- E-service, E-Government, E-Service
- Integrazione di applicativi e scambio di dati fra di essi

2.4 Piattaforma di integrazione ESB

Mule è un sistema ESB (Enterprise Service Bus) di integrazione orizzontale, soluzione utilizzata come dorsale per servizi software e componenti applicativi. Si occupa di interconnessione tra servizi, brokering (intermediazione), orchestration tra i servizi stessi, routing, messaging, data transformation, sicurezza, ...

Attualmente, la piattaforma Mule presente in Telerete integra:

- un sistema di pagamento, compreso un meccanismo di gestione degli account (correntemente utilizzato da 2 siti web)
- uno strumento di reportistica, utilizzato per la generazione automatica di report periodici relativi alla gestione della rete del Metrobus padovano, contenente anche dei programmi (oggetti) per la generazione di file pdf, grafici (formato jpg), allegati (come fogli Excel)
- 2 database contenenti informazioni sulla rete del Metrobus
- un sistema per la generazione e l'invio automatici di fax, per conto dell'Ufficio Diritti Animali di Roma, servizio fornito tramite web service (con stile architetturale di tipo Rest: REpresentational State Transfer)

Mule permette di integrare potenzialmente tutte le applicazioni si desideri; il limite deriva solo ed esclusivamente dalla disponibilità di connettori, oggetti che hanno il ruolo di mettere in comunicazione l'applicazione con la piattaforma di integrazione. Come detto, alla base dell'integrazione orizzontale risiede il fatto per cui i vari applicativi non sono mai in comunicazione diretta tra di loro; al contrario, ciascun applicativo sa come comunicare con l'Enterprise Service Bus (ESB), il quale agisce poi da intermediario (broker) tra i vari moduli integrati. Ogni applicazione fa uso di uno specifico linguaggio per parlare con l'esterno; Mule si occupa di imparare tutti i linguaggi utilizzati dalle varie applicazioni che integra, e per farlo si serve di appositi connettori implementati al suo interno. Per i protocolli di comunicazione standard, quelli più utilizzati, esiste una moltitudine di connettori liberamente scaricabili da Internet.

Ulteriori informazioni sono reperibili all'URL http://www.mulesource.org.

2.5 Gestionale ERP

Altra piattaforma presente è l'ERP (Enterprise Resource Planning), sistema informativo aziendale utilizzato per la gestione delle risorse, degli acquisti e delle vendite, della qualità, del rapporto con il cliente, e di altri aspetti di business di natura prevalentemente commerciale e amministrativa. Per conoscere ed approfondire gli aspetti di interesse sull'ERP, si sono svolti alcuni incontri con i quadri

del settore commerciale e con il referente della compagnia che ha realizzato e venduto il software gestionale a Telerete.

Il sistema ERP Freeway Skyline by Eurosystem, attualmente utilizzato in azienda, mette numerose funzionalità gestionali a disposizione del settore commerciale ed amministrativo. Al suo interno gestisce un database di magazzino, il cui DBMS è Oracle. Tutti i dati gestiti dall'ERP sono contenuti in un unico database; noi siamo interessati solamente alla porzione di database che contiene le informazioni relative al magazzino, in particolar modo ai prodotti tecnologici con cui ha a che fare l'area tecnica. Le funzionalità dell'ERP sono svariate e coprono molteplici ambiti del business aziendale, la gestione del magazzino è soltanto uno di questi aspetti.

Cespiti Se in ambito tecnico si parla di prodotto tecnologico, nel settore commerciale si utilizza il termine cespite (in inglese *asset*). I cespiti sono tutti quei valori strumentali, materiali e immateriali, che sono di proprietà dell'azienda; il termine ingloba anche tutte le attività aziendali che sono fonte di profitto per la compagnia. I cespiti aziendali di nostro interesse possono essere:

- merci già in servizio/utilizzo (per esempio apparati di rete attivi e funzionanti);
- merci stoccate in magazzino, già allocate e in attesa di essere messe in opera;
- merci stoccate in magazzino, con funzioni di scorta, per eventualità future.

2.6 OTRS

La gestione di manutenzione ed interventi porta il nostro sistema a doversi interfacciare con il sistema OTRS (Open-source Ticket Request System), che si occupa di gestire il flusso relativo alle richieste di assistenza da parte degli utenti sui prodotti e sui servizi acquistati. OTRS è installato in azienda in un'apposito server di produzione, ed è accessibile tramite interfaccia grafica a pagine web.

Ticketing Un ticket è un numero identifiactivo che viene associato ad ogni richiesta di supporto derivante dall'apertura di ticket. Attualmente i Ticket OTRS vengono creati ed aperti manualmente, da parte delle dipendenti del call center (help-desk):

- in seguito a telefonate di richiesta assistenza da parte di clienti

- in seguito alla ricezione di email o di fax provenienti da clienti (es: utente del personale APS)

Successivamente l'OTRS invia tramite mail la segnalazione al corretto gruppo di addetti delll'area tecnica e mantiene aperto il ticket fino a risposta confermata, l'OTRS tiene costantemente traccia di ogni ticket e del suo stato : 'new', 'open', 'closed unsuccessful', 'closed successful'.

Profili gestiti OTRS implementa le nozioni di 'utente', 'gruppo' e 'ruolo', relativi ai dipendenti aziendali (membri dell'area tecnica nel nostro caso). Per ciascun gruppo o utente è definito un insieme di 'code' cui quel gruppo o utente è addetto. Per quanto riguarda i clienti dell'azienda, sono definiti 'utente cliente' e 'gruppo cliente'. Oltre alle informazioni di contatto (persona di riferimento, indirizzo email, ...) ad ogni cliente, privato o ente che sia, è associata una coda, a sua volta legata ad un utente o gruppo di assistenza tecnica.

OTRS::ITSM OTRS è uno strumento molto importante e molto utile per il supporto agli utenti e per la gestione degli interventi; permette inoltre una collaborazione stretta e veloce tra help desk (call center), che riceve le richieste di assistenza, e l'area tecnica, che fornisce effettivamente il supporto al cliente. Per venire incontro anche alle esigenze del settore commerciale/amministrativo, in azienda si utilizza OTRS in combinazione con l'estensione ITSM (Information Technology Service Management), che aggiunge a OTRS i concetti di 'servizio' e di 'service level agreement' (SLA).

Capitolo 3

Progetto Log Mining

In questo capitolo si esploreranno gli obiettivi che il progetto relativo al log mining si propone di risolvere. Il punto fermo del progetto è la messa in regola rispetto alla legge anti-terrorismo, le richieste dell'area tecnica riguardanto invece la risoluzione o, se non altro, facilitare l'identificazione di problemi di rete e le cause di essi. Una volta elencati i problemi sarà possibile definire i requisiti da rispettare in fase di progettazione ed implementazione.

Una volta delineati gli obiettivi del lavoro, andranno identificati i logs e i servizi in grado di fornire i dati utili alla risoluzione di entrambe le categorie di problemi: di rete e di navigazione degli utenti (anti-terrorismo), le informazioni ricavabili da questi logs saranno spesso ridondanti, condizione utile per effettuare incrociamento di dati, ma scomoda per quanto riguarda lo spazio occupato e la gestione dei log stessi, andranno quindi effettuate scelte tra alcuni logs per evitare di memorizzare dati che si sovrappongano eccessivamente.

3.1 Contesto del Progetto

Il Progetto ha come scopo la realizzazione di un sistema in grado innanzitutto di raccogliere ed organizzare grossi moli di dati, e successivamente dare in pasto questi dati ad un algoritmo di Data Mining, il risultato finale saranno pochi dati o grafici che riassumono l'andamento o semplicemente una selezione dei dati iniziali.

L'obiettivo del progetto che si è prefissata Telerete è la realizzazione di un

sistema che possa essere di ausilio al troubleshooting ed alla risoluzione delle problematiche di rete, basandosi sulla raccolta e sull'analisi delle informazioni contenute nei vari log. La piattaforma finale deve essere in grado di monitorare in tempo reale i log di sistema, evidenziando l'eventuale verificarsi di anomalie all'interno degli stessi o trend di interesse; si cercherà, quindi, di sviluppare una piattaforma con una certa capacità di predizione, in grado di capire quando si stanno verificando quelle condizioni che in passato hanno già portato ad anomalie.

Il progetto, pur essendo molto interessante nelle logiche di business aziendale (in quanto in grado di produrre una piattaforma di ampio utilizzo in ambito tecnico, con una potenziale riduzione notevole dei costi), si è dimostrato sin da subito particolarmente complesso ed ambizioso. La tecnologia attuale è ancora distante dal fornire la possibilità di rispondere a questo tipo di esigenze: per esempio, il software ProM, che alle prime battute del progetto era subito apparso una buona base su cui poggiare le fasi successive del progetto, si è poi rivelato essere non così utile nè adatto a raggiungere gli obiettivi prefissi. Per tale motivo il progetto si concentrerà su alcuni lati basilari che vengono a monte dell'inferenza, come la raccolta e l'analisi di log di rete contenenti informazioni relative alla legge anti-terrorismo.

Il sistema quindi si pone come obiettivo iniziale la raccolta e l'analisi dei dati riguardanti i servizi di connettività forniti da Telerete nordest: Padova Wifi, Monselice Wifi, Unipd Wifi, e la successiva analisi per poter identificare tutti i dati riguardanti il traffico di un certo ip in una fascia temporale ed evadere così velocemente possibili richieste della polizia postale ottemperando peraltro alla legge anti-terrorismo (per maggiori informazioni vedere la sezione sulla legge Pisanu 3.2).

Il passo successivo è l'utilizzo dei log raccolti per identificare e risolvere le problematiche che possono verificarsi nella rete e nei servizi forniti dall'Azienda, facendo quindi da supporto al Troubleshooting attuale.

3.2 Legge Pisanu

Decreto-legge 27 luglio 2005, n. 144 : Legge 31 luglio 2005, n. 155
Ogni azienda che fornisca, tra i propri servizi, la connessione ad Internet ha

l'obbligo, dettato dalle norme di legge vigenti (155/2005, anche noto come 'pacchetto Pisanu'), di raccogliere i dati che permettano di identificare chi accede ai servizi telefonici e telematici offerti, acquisendo i dati anagrafici riportati su un documento di identità.

Oltre a questo, l'azienda che funge da Internet Service Provider (ISP) deve memorizzare e mantenere i dati relativi alla data ed ora della comunicazione ed alla tipologia del servizio utilizzato, esclusi comunque i contenuti delle comunicazioni,e deve essere in grado di estrapolare informazioni (in un range temporale) quali: chi è entrato in determinato server, quale server èè stato visitato da un certo utente o da un certo IP (da noi fornito).

Tra gli obiettivi del lavoro, oltre a raccogliere questo tipo di dati, si vuole anche predisporre un metodo di consultazione ed analisi delle informazioni raccolte. Ciò può essere utile: per uso interno, per effettuare uno studio statistico ed estrapolare trend di utilizzo del servizio, oppure su richiesta delle autorità competenti in materia di crimini informatici.

Per quanto riguarda i tempi di mantenimento delle informazioni contenute nei log, i requisiti di data retention dettati dalle norme di legge sono di 6 mesi per il traffico telematico, periodo eventualmente estendibile a 12 mesi totali su richiesta delle autorità competenti (relativamente ai casi di reati particolarmente gravi).

3.3 Analisi problemi di rete

Le problematiche che la piattaforma si propone di identificare sono quelle che seguiranno. L'elenco seguente non ha alcuna pretesa di completezza nell'effettiva casistica nè di assoluta precisione nella descrizione delle problematiche. Presenta invece, in un'ottica di soddisfazione del cliente, le principali motivazioni che lo spingono a rivolgersi all'area tecnica e all'help-desk di Telerete.

Server o access point *down* Per verificare che sia effettivamente così, si effettua un'operazione di pinging: nel caso l'host a cui si vuole accedere sia effettivamente *down*, esso non risponderà al ping. In molti casi, il problema rientra e si risolve autonomamente nel giro di alcuni secondi o minuti: questo perchè l'host può aver subito un riavvio inatteso oppure l'ultimo segmento di rete può essersi interrotto temporaneamente. Qualora dopo un breve periodo di tempo permanga lo stesso problema, si rende necessario recarsi fisicamente a spegnere e

riaccendere l'apparato. Se questo tentativo non dovesse sortire gli effetti sperati, il dispositivo probabilmente è guasto e va pertanto riparato oppure sostituito.

Server o access point non raggiungibile (*unreachable*) Molto probabilmente non si tratta di un problema dell'host al quale si vuole accedere, bensì di un problema intermedio, relativo ad un host frapposto (es. un gateway) oppure ad un segmento di rete. Un nodo interno della rete è *down* o comunque non inoltra correttamente i pacchetti verso la destinazione esatta. Una serie di operazioni di pinging, eseguite considerando la topologia di rete e le gerarchie tra i nodi, permettono di individuare la collocazione del problema, in modo da poter poi agire per ripristinare la corretta operatività dell'host malfunzionante.

In questo caso, la serie di operazioni effettuate dal tecnico per risolvere il problema e ripristinare il corretto funzionamento può variare a seconda della tipologia di connessione. I collegamenti cablati (*wired*) sono nettamente più affidabili rispetto alle connessioni wireless: può capitare che un segmento di fibra ottica o un cavo Ethernet si guasti, ma la situazione è decisamente rara; il wireless, d'altro canto, è un canale intrinsecamente variabile e determinato fortemente da fattori esterni (temperatura, condizioni meteo, oggetti frapposti, ...), motivo per cui è più frequente che un canale wireless si interrompa ("down" o "lost carrier").

Degradazione della connessione wireless A volte accade che il livello della connessione wireless subisca un sensibile degrado, evidenziato in termini di riduzione del traffico in transito (*throughput*), di alta latenza (*Round Trip Average*) e di elevata perdita di pacchetti (*Packet Loss*). Solitamente questa situazione di connettività limitata è dovuta a motivi di trasmissione fisica del segnale radio. Pertanto si agisce manualmente cambiando la frequenza portante del segnale o aumentando la potenza in trasmissione. In altri casi, il problema è dovuto alla presenza di un'altra sorgente di segnale wireless interferente; anche in questo caso si effettua un cambiamento della frequenza, portandola in un range privo di disturbi rilevanti.

Esaurimento delle capacità computazionali di un server A volte si ricevono segnalazioni di malfunzionamenti presso server, che possono essere dovuti alla mancanza di memoria RAM, di spazio libero su hard-disk, o a limiti della CPU che non riesce a mantenere costante il livello di servizio richiesto. In questo caso, per risolvere il problema bisogna agire manualmente, aumentando le risorse fisiche del server. Si preferisce però prevenire situazioni di questo tipo; è possibile

impostare dei check (anche con Nagios) con lo scopo di avvertire quando si siano raggiunti livelli di guardia in relazione alle risorse computazionali (per esempio: utilizzo RAM superiore al 80% ininterrottamente per 3 minuti).

Mancata autenticazione tra le antenne Spesso l'utente non riesce ad accedere ad un servizio di rete wireless perchè non dispone delle chiavi di sicurezza, oppure le ha settate erroneamente: l'effetto risultante è che l'antenna del terminale wireless utente non è in grado di associarsi con l'antenna dell'access point. In questo caso è sufficiente fornire all'utente la chiave WPA di autenticazione e spiegare la procedura per il corretto inserimento della chiave stessa.

Un nodo critico perde la configurazione I motivi possono essere molti e in alcuni casi la causa è difficilmente individuabile. Qualora un nodo critico (es: gateway, server) perda i file di configurazione, è necessario che questi vengano ricaricati manualmente sull'apparato; per questo motivo, è opportuno disporre di copie di backup dei file di configurazione, in modo tale da non essere costretti a ri-effettuare in toto l'operazione di configurazione.

Mancata autenticazione dell'utente con server RADIUS Problemi di autenticazione possono nascere anche a livello di server Radius (protocollo AAA). In questi casi, le possibilità sono due: l'utente ha inserito dei dati d'accesso (login e password) errati, oppure l'utente non è ancora stato registrato. Nel secondo caso, un tecnico deve provvedere ad aggiungere un'apposita riga all'interno del database del server Radius, per rendere disponibile l'accesso al servizio da parte dell'utente.

Mancata assegnazione dell'indirizzo di rete Se l'utente è riuscito ad autenticarsi alla rete, può verificarsi una situazione in cui siano terminati gli IP-address disponibili presso l'host che si occupa di assegnare gli indirizzi di rete: tale host può essere una macchina Windows impostata per fornire il servizio DHCP, oppure un server DHCP collocato presso un *captive portal*. Questa tecnica, implementata a livello degli access point wi-fi (negli *hotspot*, aree di copertura wireless), raccoglie o letteralmente 'cattura' le richieste di accesso da parte dei client HTTP che si vogliano connettere in rete, forzandoli a visitare una speciale pagina web, prima di poter accedere alla navigazione effettiva: questa pagina web permette l'autenticazione (login e password) oppure la registrazione (inserimento dati personali e accettazione delle condizioni e dei termini d'uso del servizio) da parte

dell'utente. Quando si verifica questa anomalia a livello di servizio DHCP, è necessario liberare qualche slot (ossia, indirizzo IP), se possibile; in caso contrario, si devono rivedere i meccanismi di assegnazione, eventualmente estendendo la rete e quindi il range di indirizzi IP a disposizione. Per prevenire questo inconveniente, è possibile impostare un check sul numero di indirizzi IP rimanenti, che invii un segnale di notifica qualora questi siano sotto una certa soglia critica.

Filtraggio sul traffico HTTP operato da Squid Alcuni siti, servizi o contenuti di rete sono stati resi volutamente non accessibili, questo filtraggio è attivo unicamente per la rete aziendale interna.

Filtraggio dei pacchetti ad opera di un firewall Alcuni siti, servizi o contenuti di rete sono stati resi volutamente non accessibili, e sono quindi stati bloccati per mezzo di firewall tra cui Linux's IPTables, Clavister, Pix.

Problemi nella risoluzione dei nomi La sede del problema può essere presso il terminale utente oppure presso Telerete. Nel primo caso, l'interfaccia di rete presso il computer del cliente non conosce l'indirizzo IP del server DNS. Tale informazione, solitamente, viene rilasciata dal server DHCP, e raggiunge il terminale utente assieme all'indirizzo di rete assegnato; in caso contrario, banalmente, bisogna provvedere a fornire al cliente l'indirizzo IP del server DNS. Nel secondo caso, invece, un tecnico deve procedere a riconfigurare o ad aggiornare le informazioni cui ricorre il server DNS per effettuare la traduzione dei nomi in indirizzi IP.

Problemi di e-mail A volte si riscontrano problemi relativi alla posta elettronica, sia presso clienti che presso la sede aziendale. Può capitare che, per un motivo o per l'altro, il server di posta situato presso Telerete sia andato in *down*. Ad ogni modo, considerata la crucialità di questo nodo, il server viene costantemente monitorato; perciò, eventuali problemi di questo tipo vengono rilevati e risolti in tempi brevi. Altri problemi di posta possono essere dovuti al filtraggio di spam da parte di Barracuda, un firewall che esegue controlli su tutte le email che entrano nella rete aziendale (e quindi anche le email indirizzate ai terminali presso clienti). L'azienda produttrice di Barracuda mantiene ed aggiorna una blacklist di indirizzi IP, etichettati come fonti di spam; il firewall Barracuda accede a questo database e ne confronta le entry con gli IP-address sorgente delle mail in ingresso. Statisticamente viene bloccato più del 90% delle email entranti. Oltre al filtraggio a livello IP, viene eseguito anche uno scan sui contenuti in

allegato, in cerca della presenza di eventuali virus, spyware, ... Tale sistema è attivo in azienda solamente da inizio 2009, e finora non ha presentato particolari problematiche nè comportamenti inattesi.

Impossibile accedere ad una certa sottorete Ciò è dovuto alla gestione di rete tramite utilizzo della tecnologia delle VLAN. Eventuali comportamenti anomali o inattesi possono essere dovuti solamente ad un'errata configurazione degli switch che implementano le reti virtuali: per far fronte a tali problematiche, deve essere eseguita una modifica manuale da parte di un tecnico sui file di configurazione di questi apparati.

3.3.1 Assistenza ai clienti

In azienda sono presenti due sistemi di monitoraggio: Nagios ed MRTG B.7, ma non esistono veri e propri meccanismi di ausilio al personale tecnico che si trova di fronte a richieste di supporto.

Solitamente, le richieste di supporto provenienti dai clienti di Padova Wifi, o di altri servizi di Telerete, non sono esaustive, nella descrizione del problema raramente si trovano informazioni utili all'identificazione del problema e tantomeno alla sua risoluzione; con queste informazioni il tecnico non ha la possibilità di capire rapidamente la causa del disservizio. Il tecnico dunque prova, molto spesso per tentativi, ad intuire la natura del problema effettuando una serie di controlli (*check*) che gli permettano quanto meno di circoscriverne l'area di interesse. Questa serie di check spazia dalla consultazione dei file di log, alla visualizzazione di file di configurazione degli apparati di rete, all'interrogazione delle informazioni contenute in alcuni database, fino all'esecuzione di script o comandi di tipo CLI (*Command Line Interface*) con valutazione dei rispettivi output.

3.4 Stato del Servizio WiFi

Prima di procedere col lavoro ho studiato la struttura della rete, essendo molto complessa ho deciso di riassumerla nello schema in figura 3.1, nel quale vengono mostrati unicamente i nodi e le strutture fondamentali per il problema in esame. La struttura della rete è stata semplificata essendo, altrimenti, molto complessa e non pertinente col nostro obiettivo.

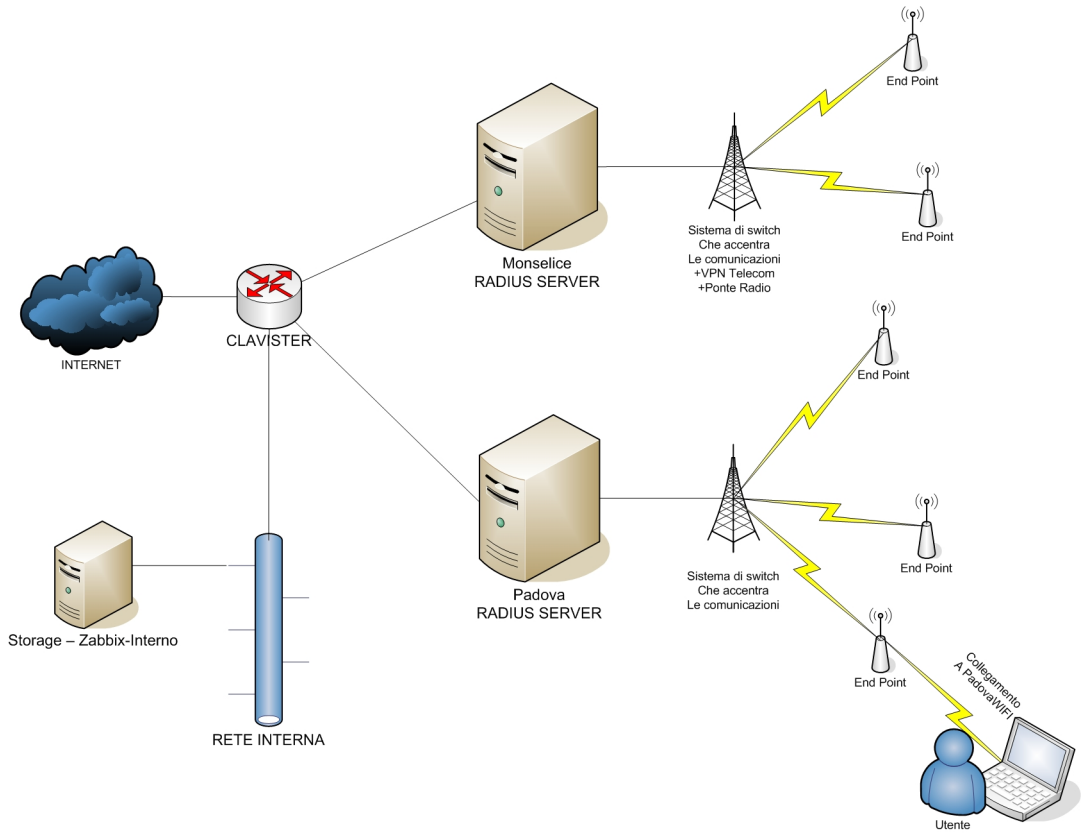


Figura 3.1: Schema fisico riassuntivo dei servizi Wifi offerti da Ne-t

3.4.1 Sistema di autenticazione

Il sistema di autenticazione si basa sul protocollo Radius, questo sistema si interfaccia con un database PostGres, e riceve informazioni dal Captive Portal, Coovachilli B.3, Squid B.4 è stato estromesso dall'autenticazione ma merita di essere nominato in questa sezione in quanto, se non fosse stato tenuto sotto controllo, sarebbe stato la causa di una falla nella sicurezza del sistema, falla che è stata corretta grazie ad una corretta impostazione di Iptables.

3.4.2 Radius

Il sistema Radius presente nella rete Padova WiFi è, in particolare, l'implementazione FreeRadius B.2 del protocollo. Questa istanza di FreeRadius si frappone tra il Database, che contiene le credenziali e i dati personali degli utenti iscritti a Padova Wifi, ed il Captive Portal, che ha il compito di validare, tramite Radius, i dati inseriti in fase di login.

Il compito di questo servizio è di verificare i dati che Coovachilli inoltra e di segnalare a Coovachilli se le credenziali inserite risulano valide o meno. Radius inoltre può fornire, se richiesti dalla polizia postale, i dati personali degli utenti.

3.4.3 Coovachilli

Coovachilli, ovvero il captive Portal, è la prima interfaccia che l'utente si trova davanti al momento dell'accesso alla rete di Padova WiFi (o Monselice WiFi), questa richiede l'inserimento di credenziali prima di poter navigare a tutti gli effetti. Inizialmente fornisce un IP per permettere all'utente di inviare le proprie credenziali. A questo indirizzo, Coovachilli, blocca tutto il traffico uscente, fino al momento dell'avvenuta autenticazione. Una volta che Radius conferma a Coovachilli l'effettiva autenticità delle credenziali, Coovachilli sblocca l'IP dell'utente e ne permette la navigazione libera. Sarà compito di Iptables e di Squid effettuare un filtraggio dei Server o degli URL che gli utenti andranno a visitare.

Coovachilli ha, quindi, il compito di assegnare un IP (che in realtà assegna il DHCP) all'utente che si collega, bloccarlo e consentire di immettere le credenziali che verranno inoltrate al Radius presente, se questo conferma le credenziali è possibile navigare, quindi Coovachilli sbloccherà l'IP, altrimenti lo manterrà bloccato. Coovachilli mantiene nel proprio database e nei propri logs, un'associazione

tra IP e username corrispodente, con associata data ed ora di inizio e di fine sessione.

3.4.4 Squid

Inizialmente era stata pensata una versione di Padova Wifi contentente Squid come Proxy, sul quale effettuare le autenticazioni. L'ipotesi è stata scartata a causa della, seppur lieve, difficoltà che introduceva, ovvero la necessità da parte dell'utente di impostare sul proprio PC il proxy.

Il Web-Proxy è stato comunque mantenuto per poter usufruire dei servizi di Web-Cache, utili per ottimizzare l'utilizzo della rete. Dopo un periodo di utilizzo però si è scoperto che Squid forniva un sistema per bypassare l'autenticazione e tutti i controlli successivi. Il metodo che è stato scoperto si appoggia all'IP temporaneo, che Coovachilli fornisce agli utenti per permettere di effettuare l'autenticazione, una volta che l'IP era assegnato era sufficiente impostare Squid come Proxy pur senza effettuare il login, a causa di queste problematiche il servizio è stato momentaneamente sospeso.

3.5 Requisiti

Rispettando le cogenze della legge antiterrorismo (Legge Pisanu) e cercando di creare un sistema che sia di supporto all'identificazione dei problemi di rete sopra descritti, si sono potuti definire i requisiti formali, che saranno da rispettare nella progettazione e nell'implementazione della piattaforma nel suo insieme.

I requisiti saranno suddivisi in :

1. Requisiti Obbligatori, una serie di funzionalità ed obiettivi che non sono omissibili durante la progettazione

2. Requisiti Facoltativi, se possibile da rispettare ma se diventano un problema o allungano troppo le tempistiche del lavoro si possono tralasciare

3. Requisiti Interessanti, sono le funzionalità che si ipotizza facili da implementare e soddisfacneti per il cliente anche se non le aveva richieste

4. Requisiti Desiderabili, sono quelle funzionalità aumenterebbero la soddisfazione dell'utilizzatore finale del prodotto che si sta progettando

3.5.1 Requisiti Obbligatori

Sono i requisiti emersi direttamente dalle cogenze di legge e dalle necessità riscontratein area tecnica. Molti sono stati definiti esplicitamente ma una parte è stata estrapolata in quanto spesso ai committenti delle funzionlità o considerazioni sembrano implicite.

1. Gestione della rotazione dei log per dividere temporalmente in maniera adeguata i log.
2. Gestione da interfaccia web della configurazione per la raccolta dei log.
3. Interrogabilità dei log e opportuna base di query predefinite.
4. Sistema di tracciabilità dell'attività internet delle utenze internet di Telerete.
5. Modularità degli applicativi di interrogazione dei log sviluppati.
6. Schermate informative sullo stato della rete interna.
7. Caricamento dei log di iptables, kern tables e coovachilli.
8. Portabilità del sistema.
9. Il sistema farà parte di un progetto preesistente, ed in quanto tale utilizzerà parzialmente codice già scritto.
10. Gestione dei cron jobs esistenti nel sistema.
11. Modifica configurazione dei syslog client e server esistenti.
12. Modularità degli applicativi sviluppati
13. Trasmissione dei log macchina-macchina e raccolta su supporto ridondante esterno.
14. Informazioni e filtri sugli URL visitati
15. Sistema di audit trail degli accessi al sistema.
16. Allineamento temporale tra i vari dati raccolti

3.5.2 Requisiti Facoltativi

I requisiti facoltativi non sono 'core' per il progetto, sono comunque utili, ma non necessari, forniscono indicazioni su future espansioni del progetto o sulle funzionalità che devono essere aggiunte nelle varie versioni prodotte e su eventuali migliorie da apportare:

1. Sistema di gestione degli accessi alle interfacce di query e configurazione.
2. Caricamento di log generici nella base dati.
3. Possibilità di specificare colonne indice nel caricamento di log generici.
4. Caricamento dei log di posta Postfix.
5. Creazione manuale passo passo della configurazione dell'intero sistema in uso.
6. Riscrittura ed ottimizzazione codice non portabile o errato.
7. Rappresentazione grafica del traffico di rete.

3.5.3 Requisiti Interessanti

I requisiti interessanti non sono stati espressamente richisti ma risultano relativamente semplici da soddisfare ed aumenterebbero la qualità del prodotto:

1. Sistema di Web-cache per ottimizzare le prestazioni della rete
2. Costruzione automatica delle query durante le interrogazioni
3. Ottimizzazione regular expression
4. Grafici che rappresentino l'utilizzo dei servizi

3.5.4 Requisiti Desiderabili

I requisiti desiderabili sono stati suggeriti dai committenti come lavori da eseguire terminate le mansioni principali:

1. Sistema user-friendly
2. Integrazione delle varie funzionalità

3. Esportabilità del risultato delle query sui log in formati csv (comma separated values), excel, sql.

4. Creazione guidata della parte grafica degli applicativi.

5. Eliminazione delle query al database di Freeradius ed utilizzo esclusivo dei file di log per la determinazione delle utenze internet.

3.6 Identificazione Problematiche di Rete

Analizzate le problematiche e la casistica con cui si possono presentare, risulta evidente la necessità di analizzare delle possibili variabili che possano essere ricondotte alla classificazione del problema o alla risoluzione di questo.

Come anticipato, la possibilità di inferire nuove informazioni da dati in possesso in forma di log testuali é legata alla identificazione di variabili che siano il più possibile estranee al singolo nodo della rete o al singolo log e che siano invece correlate a fattori che condizionano l'intera rete e ne descrivano caratteristiche sensibili.

3.6.1 Numero utenti connessi

Il numero di utenti connessi é evidentemente una variabile sensibile della rete che condiziona lo spazio degli indirizzi assegnabili ed il traffico transitante.
I problemi a cui potrebbe quindi essere correlata sono:

1. Mancata assegnazione dell'indirizzo di rete

2. Degradazione della connettività di rete

3. Problemi di e-mail

Il numero di utenti lo si può dedurre in varie maniere, più o meno precise e fedeli alla realtà, come anticipato prima, però, ci troviamo nella posizione di doverlo dedurre da informazioni di cui già siamo in possesso, ed in particolare di log.
Il numero di utenti connessi può essere dedotto in vari modi che analizzeremo sotto.

Log Radius : é il server di autenticazione degli utenti della rete. Dai log possiamo venire a conoscenza del numero di autenticazioni eseguite con successo in un arco temporale. I log di Radius memorizzano ogni tentativo fallito di login e il momento in cui essi avvengono. L'inconveniente potrebbe essere che radius non fornisce log di disconnessione ed inoltre non ci fornisce indicazione degli utenti autenticati al di fuori dell'arco temporale preso in considerazione.

Log dns : i log dns ci danno un'idea del numero di utenti connessi in termini di MAC addresses associati ad indirizzi IP.

3.6.2 Numero autenticazioni fallite

Anche se non direttamente associata ad una problematica di rete e generalmente associata ad un errore dell'utente, questa variabile potrebbe evidenziare importanti correlazioni che verranno esposte più avanti.
Le problematiche correlate potrebbero essere:

1. nodo server o access point down
2. mancata autenticazione dell'utente con server RADIUS
3. Un nodo critico perde la configurazione

Questa variabile può essere utilizzata come percentuale di autenticazioni fallite, media di autenticazioni fallite in un arco di tempo etc.

Log Radius : i log di radius registrano ogni autenticazione fallita e l'istante in cui vengono effettuate.

3.6.3 Traffico attraversante un nodo

Questa variabile é evidentemente molto cruciale nel descrivere una rete e può preludere ad eventi successivi frequenti come disconnessioni, lentezza nei servizi etc.
Le problematiche in particolare a cui é legata sono:

1. Problemi di e-mail
2. Degradazione della connettività di rete

I log da cui si può dedurre il traffico attraversante un nodo possono essere molteplici ma si prenderà in considerazione solo quello che lo fornisce come dato esplicito e senza nessun onere aggiuntivo.

MRTG log : i log di mrtg ci forniscono direttamente il traffico in Kbyte attraversante un nodo senza nessuna attività sui singoli log di traffico.

3.6.4 RTA medio

In forma estesa Round Trip Average rappresenta il tempo medio richiesto ad un pacchetto per raggiungere destinazione: questa variabile potrebbe essere molto indicativa sull'andamento generale della rete e potrebbe allertare gli amministatori del verificarsi di eventi o di condizioni che, in passato, hanno creato problemi.
I problemi che potrebbe identificare sono quindi:

1. degradazione della connettività di rete
2. nodo server o access point down
3. server unreachable

L'RTA può essere dedotto da più log e può essere utile sia considerato come media degli RTA di un singolo nodo sia di più nodi raggruppati.

Nagios log : i log di nagios contengono RTA di ping ad host preimpostati

3.6.5 Packet Loss

Il packet loss come l'RTA é una variabile che descrive la connettività di un nodo e la degradazione di quest'ultima.
I problemi a cui potrebbe essere correlata, come per l'RTA, sono:

1. degradazione della connettività di rete
2. nodo server o access point down
3. server unreachable

Generalmente il packet loss é legato all'RTA, quindi si trova negli stessi log in cui si trova l'RTA:

Nagios log : i log di nagios contengono packet loss di ping ad host preimpostati

3.6.6 Altri parametri

Altri parametri utili ad identificare problematiche di rete, o che aggiungerebbero informazioni interessanti durante la fase di analisi saranno ora brevemente descritti.

Rapporto traffico uscente / traffico entrante

Il rapporto fra traffico entrante ed uscente, con il traffico totale in un dato intervallo, dovrebbe, di massima, restare costante, oppure seguire dei pattern ben definiti e ripetitivi nell'arco della giornata o della settimana.

Topologia della rete

Pur non essendo una variabile normalmente deducibile dai log in possesso, dovrebbe essere possibile specificare la topologia della rete come grafo di nodi connessi.

Fascia oraria

Potrebbe essere utile inserire anche una variabile che rappresenta la fascia oraria, ma questa é deducibile durante l'analisi dei logs, risulta quindi superfluo aggiungere dati facilmente ottenibili.

Numero messaggi anomali

Si potrebbe prevedere di fare un parsing dei file di log, e, tramite un pattern matching, contare il numero di messaggi previsti e generalmente presenti nei log, ed il numero di messaggi di log anomali (in quanto generalmente non presenti, oppure presenti ma in numero inferiore).

Process Mining

Il process mining é utile nel momento in cui evidenzia trend o devianze altrimenti non osservabili dai singoli eventi. E' ipotizzabile un certo numero di correlazioni che un process mining efficace potrebbe trovare sui dati e che potrebbe indirizzare all'aggiunta di nuove variabili o di nuovi log e quindi di dati in ingresso al processo di data mining.

L'aumento di utenti, la diminuzione della banda, l'aumento del Round Trip Average e la degradazione risultante dei servizi é una correlazione immediata, che l'algoritmo di process mining dovrebbe evidenziare.

3.7 Integrazione con altri sistemi

Il nostro sistema dovrá interagire con gli altri sistemi presenti in azienda, la struttura aziendale risultate, a livello logico, avrá una forma a stella, uno schema rappresentativo del sistema che dovrà risultare al termine del progetto é rappresentato in figura 3.2. E' utile avere un'idea di come procedere prima di iniziare il progetto, in modo da non trovarsi spiazzati nelle fasi seguenti.

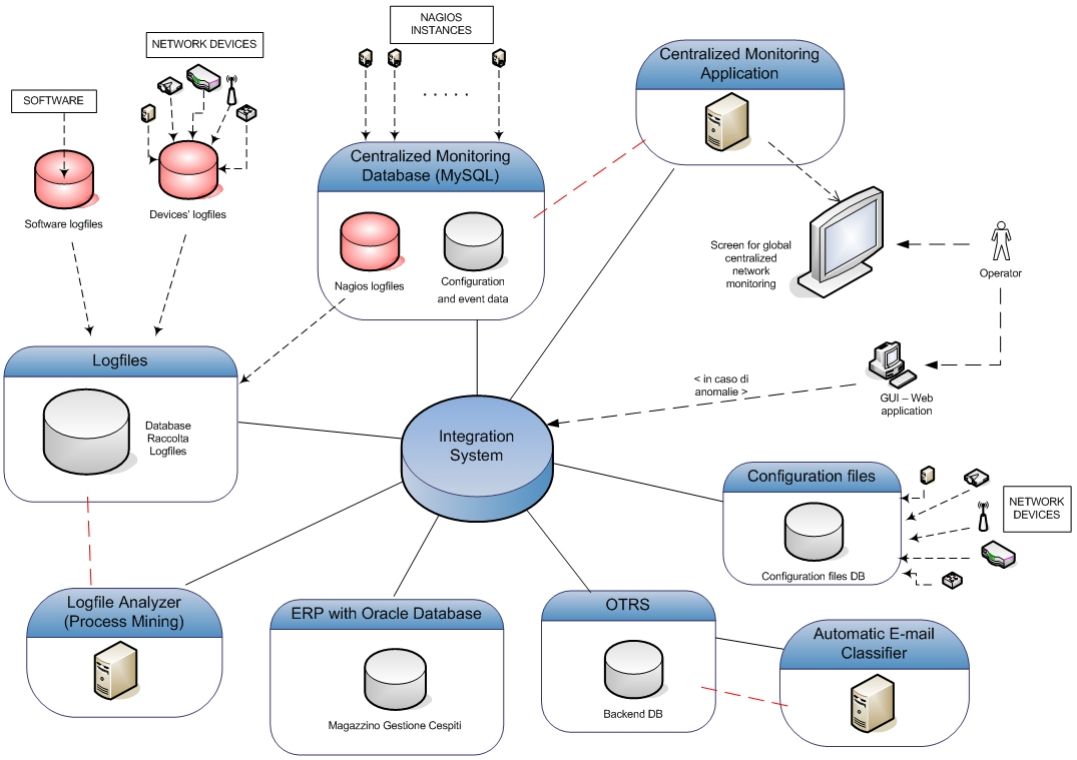


Figura 3.2: Schema dell'integrazione trai vari sistemi presenti in azienda, il log mining dovrà inserirsi in questi sistemi

Capitolo 4

Implementazione Raccolta

In questo capitolo verranno descritte le fasi che hanno contraddistinto la configurazione del sistema di logging per adempiere alla legge anti-terrorismo e per tentare di fornire una base per l'identificazione dei problemi di rete, in questa fase sono stati già selezionati i logs che verranno utilizzati nel progetto, e sono stati, quindi, specificati anche i dati che da essi verranno estratti. La priorità, come è prevedibile, è stata assegnata alla messa in regola rispetto alle leggi 144,155 (definito pacchetto Pisanu 3.2).

Una volta definiti i dati e le priorità, è stato necessario effettuare valutazioni sulla sicurezza del sistema, andando a testare l'effettiva inviolabilità del server e sulla dimensione dei dati che si andranno a memorizzare, cosa che ha imposto la creazione di sistemi per organizzare al meglio, mediante automatismi, il server centrale di raccolta e i dati in esso contenuti.

4.1 Riepilogo obiettivi prefissati

Il progetto ha lo scopo di realizzare un sistema integrato per la gestione delle reti Wireless sollo il controllo e la responsabilità di Telerete NordEst, il controllo degli accessi a questa rete verrà affidato ad un applicativo che si basa sull'analisi dei log, un sistema non invasivo dal quale si possono ricavare molte informazioni circa l'andamento e l'utilizzo della rete. Al contrario di altri sistema di monitoring di rete, questo non crea traffico al momento delle richieste, le informazioni sono già presenti nei logs, i quali saranno stati, precedentemente, centralizzati. Il nostro obbiettivo è di utilizzare questi log come risorsa, non solo per adempiere ad obblighi legislativi, ma anche per studiare l'andamento della rete, rilevare eventuali

malfunzionamenti e saper rispondere in fretta, e possibilmente in maniera automatizzata, a determinate problematiche.

L'obiettivo generale dell'analisi dei log è quello di trarre informazioni utili ed immediatamente fruibili, partendo da dati di cui si é già in possesso, senza cioè alcuna attività aggiuntiva di ricerca delle informazioni.

Tali log non sono idonei ad essere analizzati da un operatore umano, in quanto soggetto a limiti di tempo, attenzione e possibilità di errore.
La possibilità di ottenere informazioni nuove è legata, quindi, alla capacità di automatizzare la raccolta, l'analisi e l'interpretazione di grossi moli di log provenienti da macchine ed applicativi eterogenei. Trarre informazioni utili ed immediatamente fruibili da dati in possesso è il punto di forza del process mining.

Nel caso di una rete complessa, comprendente variabili quali host, servizi ed utenti, per ottenere informazioni generali, è necessario disporre di una buona base di dati, da incrociare ed interpretare, i dati in causa vengono forniti dai log generati dai vari nodi componenti la rete e dai vari servizi associati. Questi log presi singolarmente non sarebbero in grado di descrivere la rete in cui si trovano, nè di fornire nuove informazioni che non siano direttamente osservabili dai dati grezzi ricevuti in input. La possibilità di inferire informazioni nuove e legata alla identificazione di variabili che siano il più possibile slegate dal singolo nodo della rete o dal singolo log e che siano invece legate a fattori che condizionano l'intera rete e ne descrivano caratteristiche sensibili.

Il fine della raccolta sarà quindi quello di fornire dati ed identificare variabili da utilizzare per uno o più algoritmi di process mining, nel caso in cui si riuscisse a calare un processo alla rete.

Bisogna tenere presente che queste variabili devono essere deducibili dai log, mediante lettura diretta o lettura incrociata dei dati contenuti nei vari log e non devono coinvolgere altre attività specifiche, devono cioè essere informazioni che arrivano senza costi aggiuntivi ed unicamente dall'osservazione di questi. Le variabili devono essere inoltre deducibili con un costo preventivato e sostenibile, spesso infatti i log sono dell'ordine di Gigabyte e prodotti nell'arco di millisecondi, e non è ipotizzabile una loro analisi in tempo reale.

Il prodotto ultimo sarà un sistema distribuito su più macchine, con funzionalità che andranno dalla raccolta dei log, alla loro reperibilità, interrogabilità e disponibilità, al fine di creare un sistema di integrazione delle informazioni a beneficio dei sistemisti di rete.
Più praticamente il sistema dovrebbe guidare un operatore nella ricerca di eventuali malfunzionamenti dell'infrastruttura di rete, suggerendo dei test effettuabili, e, per alcuni test, fornire già delle risposte deducendole dai logs.

L'idea di fondo è che, molte informazioni, possono essere estratte dai log senza bisogno di eseguire dei test specifici, che utilizzino nuovamente la rete per ritornare un esito positivo o negativo. Nel caso più semplice, un host 'down', può essere recepito da una assenza di log di traffico, piuttosto che da un ping attivo sulla macchina (anche in assenza di attività un host ha comunque un certo traffico ARP o di altro genere per mantenere la connessione alla rete, quindi per mancanza di traffico nei log si intendono sessioni di utilizzo attivo della rete).

4.2 Studio dei log da raccogliere

Definite le variabili di rete cui si vuole venire a conoscenza, è ora opportuno definire delle fonti per recuperare queste informazioni. Oltre ai problemi di rete, non è da dimenticare la mission principale, che in questo frangente è la creazione di un sistema atto ad evadere rapidamente richieste provenienti dalla Polizia Postale.

La prima fase del lavoro è consistita nello studio approfondit delle modalità con le quali viene effettivamente erogato il servizio di Connettività da parte di Telerete. Questo studio ha richiesto di capire nel dettaglio quali sono le applicazioni ed i dispositivi coinvolti nella fornitura del servizio, operazione che ha permesso di identificare chiaramente i sistemi di logging in gioco, ossia i sistemi che sono in grado di fornire informazioni utili, da mantenere nel server centrale per:

- ottemperare alle norme di legge,
- fornire indicazioni utili in merito ai problemi di rete ed alla loro risoluzione.

La prima analisi sul servizio Connettività è stata fatta su PadovaWiFi, la rete wireless cittadina che poggia su un'infrastruttura di rete sotto la gestione di

Telerete. Da questa fase di analisi sono usciti i primi sistemi di logging su cui si è ritenuto di cominciare il lavoro, nello specifico:

- FreeRadius
- Coova-Chilli
- IPTables
- Squid

4.2.1 FreeRadius

FreeRadius è un'implementazione open-source del protocollo Radius (vedi appendice B.2). Nell'infrastruttura di rete, FreeRadius svolge il ruolo di server di autenticazione, dotato di un database PostgreSQL che mantiene traccia:

- dei dati anagrafici (forniti durante la fase di registrazione) degli utenti, che permettono di identificare la persona fisica che accede al servizio,
- delle sessioni di utilizzo da parte di ogni singolo utente, ma non delle disconnessioni,
- dell'accounting relativo all'utilizzo del sistema da parte di ogni singolo utente,
- altre informazioni, tra cui quelle relative ai NAS (*Network Access Server*), ai profili di accesso al servizio e relativo billing, ecc...

Il log di FreeRadius contiene informazioni relative ai tentativi di autenticazione da parte degli utenti, sia nel caso di *Login OK* che in quello di *Login incorrect*: tra queste informazioni compaiono lo username ed il MAC address della postazione da cui l'utente effettua la connessione al servizio, mentre la password (CHAP) non viene riportata nei log per ovvi motivi di privacy. Vengono poi registrati eventuali messaggi d'errore, per esempio la mancata connessione con il database PostgreSQL o il fallimento dell'esecuzione di alcuni script di supporto.

Questa implementazione di Radius non fornisce, come già enunciato nel precedente capitolo 3.6.1, i dati riguardanti disconnessioni e utenti attivi al di fuori del'arco temporale selezionato. Una soluzione sicuramente attuabile sarebbe quella di accedere al database di Radius per verificare il numero di utenti autenticati ed attivi ma ci troveremmo a percorrere una via che esula dalla sola analisi dei log a disposizione.

4.2.2 Coova-Chilli

Coova-Chilli è il software che implementa il *captive portal*; fornisce all'utente l'interfaccia di autenticazione in cui inserire login e password, e dialoga col server Radius per effettuare il controllo degli accessi al servizio (vedi appendice B.3). Funge inoltre da DHCP server, assegnando un'indirizzo IP agli utenti che si autenticano con successo e mantenendo l'associazione IP address - MAC address - Data e ora.

Il log di Coova-Chilli quindi registra:

- i login, salvandone username e IP address;
- i logoff, salvandone ancora username e IP address;
- le DHCP request, memorizzando il MAC richiedente;
- le assegnazioni di IP address ad un MAC, da parte del DHCP server;
- i rilasci, da parte di un MAC address, del corrispondente indirizzo IP, che torna quindi a disposizione del DHCP server per nuove allocazioni;
- notifiche di eventuali anomalie (es. nella comunicazione con il server Radius) o dell'esecuzione di operazioni di routine (es. reload dei configuration files).

Altra informazione utile, che Coovachilli fornisce, è il numero di utenti che si sono collegati con successo durante la giornata corrente e nell'ultima ora.

4.2.3 IPTables

IPTables è il noto firewall utilizzato in ambiente Linux, esegue un *packet filtering* a livello di rete (livello 3 dello stack ISO-OSI, *network layer*).

Iptables è un modulo integrato nei sistemi Linux, il suo ruolo è, non solo quello di firewall, ma anche gestore del traffico di rete all'interno di ogni macchina, è possibile specificare regole molto precise a livello di rete, ma, essendo per l'appunto un protocollo di basso livello, non è possibile effettuare controlli elaborati.

Ogni pacchetto in entrata, in uscita, o anche solo di passaggio in una macchina, viene analizzato da Iptables, dipendentemente dalle regole impostate, una o più volte, esistono quattro *table* sulle quali impostare regole : filter, nat, mangle, raw; nel nostro caso useremo quasi unicamente la tebella FILTER (Nat sarà utile solo

per Squid).

Il log di IPTables memorizza numerose informazioni relative ai pacchetti IP in transito. Possono essere impostate regole di logging diverse, a seconda della specifica tabella e della *chain* utilizzata: INPUT (regole di ricezione pacchetti), FORWARD (regole di inoltro pacchetti) e OUTPUT (regole di invio pacchetti). In particolare, vengono salvati i seguenti dati:

- interfacce di rete impiegate (in entrata e in uscita);
- l'IP address sorgente;
- l'IP address destinazione;
- la porta usata presso la sorgente;
- la porta usata presso la destinazione;
- la lunghezza del pacchetto;
- l'identificativo (ID) del pacchetto;
- il protocollo di trasporto utilizzato;
- il valore della window size di TCP;
- altri parametri specifici, tra cui il TTL (Time-To-Live), il ToS (Type-of-Service), o i flag di TCP.

4.2.4 Squid Log

Squid é un Proxy con capacitá di web-cache, il suo ruolo in questa struttura, fino ad oggi, era ristretto a web-cache, utile per evitare di effettuare richieste ridondanti verso l'esterno, spesso capita che piú utenti richiedano le stesse informazioni in brevi lassi temporali, grazie a Squid é possibile mantenere nella cache i dati di ogni richiesta per un breve periodo, rendendo cosí il server capace di restituire la pagina o il contenuto già in cache, senza dover inoltrare richieste uguali verso Internet, con questo sistema si riesce ad alleggerire il 'collo di bottiglia' che normalmente si crea verso l'esterno.

Squid analizza ogni pacchetto che inoltra, ed è in grado di estrarre molte informazioni, di queste solo alcune sono di interesse per i nostro obiettivi:

- l'IP address sorgente;

- l'IP address destinazione;
- la porta usata presso la sorgente;
- la porta usata presso la destinazione;
- la lunghezza del pacchetto;
- l'URL visitato;
- l'oggetto MIME della comunicazione;
- il metodo utilizzato (POST/GET/CONNECT/...)
- l'esito della richiesta HTTP (codice);
- altri parametri specifici.

I log di squid contengono informazioni utili riguardanti gli oggetti in transito, ottenibili unicamente da servizi che agiscono a livello applicativo, come fa, per l'appunto, Squid.

La conoscenza dei contenuti dei server che vengono visitati fornisce la possibilità di effettuare filtraggio ad alto livello e limitare, o se non altro tracciare, la visione di contenuti illegali (es: pedo-ponografia, terrorismo ecc..)

4.2.5 Log utili per problematiche di rete

Verranno ora elencati e descritti i logs legati a problematiche di rete e che, se bene utilizzati, potrebbero guidare l'area tecnica nell'identificazione e nella risoluzione di questi problemi.

Log DNS

I log dns ci danno un'idea del numero di utenti connessi in termini di MAC addresses associati ad indirizzi IP. Ancora una volta non possiamo ottenere un dato certo sul numero effettivo di utenti, in quanto questo sarà altamente dipendente dall'ampiezza dell'intervallo preso in considerazione e dal lease time (potrebbe essere dell'ordine di minuti come un mese) in cui una associazione MAC address IP viene considerata valida.

Delle opzioni definite nel capitolo precedente (3.6.1) quella dei log dns sembrerebbe la soluzione più attuabile, bisogna tenere conto che la verosimiglianza

della variabile confrontata al numero effettivo di utenti, non è strettamente indispensabile ai fini del data mining, che dovrebbe essere, comunque, in grado di individuare trend e correlazioni a partire da variabili 'sfalsate'.

Resta salvo il fatto che la variabile del modello deve comunque descrivere, al suo variare, un effettivo cambiamento nel mondo reale.

MRTG log

Multi Router Traffic Grapher (MRTG) B.7 è un software per il monitoraggio delle apparecchiature di rete tramite il protocollo SNMP, questo monitoraggio ha lo scopo di creare automaticamente dei grafici sul traffico che riguardano i nodi analizzati. I grafici in questione vengono accompagnati da una pagina HTML che guida all'interpretazione dei valori, facilitandone così la pubblicazione.

I log di MRTG ci forniscono direttamente il traffico in Kbyte attraversante un nodo senza nessuna attività sui singoli log di traffico, solitemente vengono utilizzati per creare grafici rappresentanti l'andamento di ogni apparato o della rete in generale su diverse scale temporali, i dati vengono regolarmente storicizzati, per fornire grafici che rappresentino l'andamento secondo scale temporali differenti. Questa variabile ci verrebbe fornita senza nessun intervento, andando a reindirizzare i logs forniti dell'applicativo MRTG, da tempo presente ed operativo in azienda.

Nagios log

I log di nagios contengono RTA e packet loss di ping ad host preimpostati, sono utili per effettuare analisi 'al volo' sullo stato generale delle macchine e degli apparati di rete. Di base Nagios fornisce un'interfaccia Web, utile per effettuare questi controlli periodicamente, nei log vengono memorizzate informazioni maggiori, ma queste informazioni sono ottenibili unicamente installando su ogni apparato dei servizi che comunichino col Nagios centrale, che ha il compito di effettuare la raccolta e l'analisi dei dati contenuti nei logs.

Anche Nagios è già presente ed attivo in azienda nella sua versione originaria, grazie alla quale è possibile effettuare un monitoring degli apparati di rete, pur senza venire a conoscenza di cosa accade al loro interno.

4.3 Valutazioni prima di procedere

La tecnologia attuale e ancora distante dal fornire la possibilità di rispondere a questo tipo di esigenze: per esempio, il software ProM, che alle prime battute era apparso come una buona base su cui poggiare le fasi successive del progetto, si è poi rivelato essere non così adatto a raggiungere gli obiettivi prefissati senza una manipolazione iniziale efficiente dei log files.

ProM non è il primo tool del suo genere, esistono sia soluzioni accademiche (e.g., EMiT, Little Thumb, InWoLvE, Process Miner e MinSoN) che soluzioni commerciali (e.g., ARIS PPM, HP BPI e ILOG JViews) per effettuare process mining. Il problema con le soluzioni citate però è che utilizzano formati diversi per leggere e scrivere i log files e presentano i risultati in maniera differente tra loro rendendo la portabilità un problema cruciale nello sviluppo.

Ciò rende problematico ed inefficiente l'uso di piú tool su un medesimo dataset per riuscire ad estrarne informazioni utili. ProM invece è stato sviluppato ponendo la flessibilità come primo requisito e permette quindi grande personalizzazione dei log sia in input che in output, la creazione di nuovi algoritmi di process mining è, quindi, enormemente semplificata per l'utente finale.

Nonostante la presenza di questo tool la scelta è stata comunque quella di concentrarsi su un aspetto particolare del progetto: la raccolta e l'analisi di log di rete e l'estrazione di informazione da questi, lasciando le inferenze più complesse ad uno sviluppo successivo che utilizzi ProM. La scelta è dettata dalla necessità di disporre di un sistema efficiente di raccolta e di manipolazione iniziale dei log, che possa astrarre le problematiche di reperimento, collezionamento, estrazione e formattazione delle informazioni.

4.3.1 Valutazioni dimensionali

Prima di procedere si è stimata la reale dimensione che i log prodotti potrebbero raggiungere, un altro parametro necessario da valutare per una progettazione di buon livello è la velocità di estrazione dei dati, o meglio della lettura dei dati contenuti all'interno dei logfiles compressi. Sono stati eseguiti test empirici per determinare quale delle soluzioni seguenti sia la più indicata, e si è giunti alla conslusione che, nel nostro caso, Gzip offre il miglior compremesso in termini di tempo e spazio occupato, allontanandosi solo minimamente dalle due soluzioni ideali, ovvero ricercare all'interno di un file non compresso (RAW) o la massima compressione offerta dall'algoritmo TARBALL (con uno 0.5 % di spazio in più

occupato), BZip2 offre un miglior rate di compressione ma la lettura del file e la compressione stessa richiedono uno sforzo eccessivo per la macchina, quindi è stato scartato assieme a RAR, che, da alcuni benchmark trovati online, offre prestazioni molto simili a BZip2, anche RZip è stato escluso sul nascere in quanto richiede enormi quantità di RAM e CPU, che non ci si può permettere di occupare su di un Server attivo.

Nella tabella 4.1 sono stati effettuati i test su di un file testuale di 413.25 MB con i due algoritmi più promettenti :

Sistema di archiviazione	*Dimensione risultante(MB)*	*Tempo per il Grep(sec.)*
RAW	413.25	2
GZIP	27.66	3
TARBALL	27.52	19

Tabella 4.1: Tabella sulla valutazione dimensionale

4.3.2 Valutazioni sulla sicurezza del sistema

Per l'autenticazione la scelta è ricaduta su Coovachilli, o Captive portal, questa scelta è stata fatta per semplificare l'utilizzo del servizio da parte degli utenti dei servizi WiFi di Ne-t meno esperti, l'applicazione presa singolarmente non crea falle di sicurezza, ma, accoppiata a Squid, a causa della sua particolarità di essere un Proxy, crea una situazione grazie alla quale, utenti non autenticati, sono in grado di usufruire del servizio WiFi, minando quindi la sicurezza generale del sistema.

Per poter accertare le cause di questo bypass è stato effettuato uno studio sull'effettivo schema di comunicazione tra processi che avviene all'interno dei server Radius, lo studio ha portato sul tavolo varie possibilità, alcune di queste sono poi state confermate grazie a vari test, i quali consistevano in tentativi di emulazione di 'malintenzionati'.

Essendo necessario, per terminare con successo il progetto, disporre delle informazioni fornite da Squid, si è deciso di attivarlo non appena fosse stata disponibile una soluzione valida per eliminare il buco nella sicurezza. Inoltre, la capacità di

effettuare Web-Cache localizzata nei server Radius, alleggerisce il traffico verso Internet e sul Firewall aziendale: Clavister.

Dopo circa una settimana dedicata allo studio approfondito della problematica ed al funzionamento di Squid, ho tentato di risolvere il problema tentando svariate configurazioni di Squid, al fine di impedirgli di essere efruttato come 'ponte' per scavalcare Coovachilli.

Dopo questo periodo di testing e di avvicinamento alla distribuzione Gentoo, la quale, data la sua complessità, mi ha creato non pochi problemi, ho deciso di modificare strategia, non puntando più sull'applicativo stesso per risolvere il problema, bensì tentando di risolverlo a livello di Rete. Prima di proseguire è stato schematizzato lo scambio di informazioni che avviene all'interno dei server Radius, come mostrato in figura 4.1, si nota che è stata riportata sia la situazione problematica che una rappresentazione della soluzione.

Avevo notato che il Bypass non avveniva di per sè per un buco nella sicurezza del Firewal, di Squid o del Captive Portal, ma, appunto, per l'impiego di più strumenti sullo stesso sistema, i quali fornivano servizi simili (login) ma in maniera differente, questo creava più 'agganci' verso la macchina, la configurazione trasparente di Squid era risultato essere il punto debole.

Il bypass avveniva a causa della presenza del Proxy, il quale, essendo inferiore a livello TCP/IP rispetto al Captive Portal, permetteva un collegamento da parte degli utenti verso la macchina, anche se si imponeva che il Proxy si comportasse in modalità trasparente. Era possibile simulare un'autenticazione impostando sul proprio browser, o sul proprio PC, il Proxy. Inserendo come indirizzo IP quello del Server Radius e la porta asseganta a Squid, si evitava la necessità di autenticarsi, scavalcando così anche ogni tipo di filtraggio.

Soluzione IPTables

La soluzione ritenuta ottimale si raggiungeva andando ad impostare correttamente il Firewall, a livello Rete è possibile bloccare ogni pacchetto rivolto all'istanza di Squid nel server Radius.

Praticamente ogni pacchetto che avesse come porta di destinazione quella di Squid, o al suo utilizzo come proxy, andava eliminato. Una volta appresa la corretta sequenza di passaggi che avvengono, dentro Iptables, al passaggio di ogni pacchetto, è stato sufficiente ragionare su un artificio semplice ed efficace

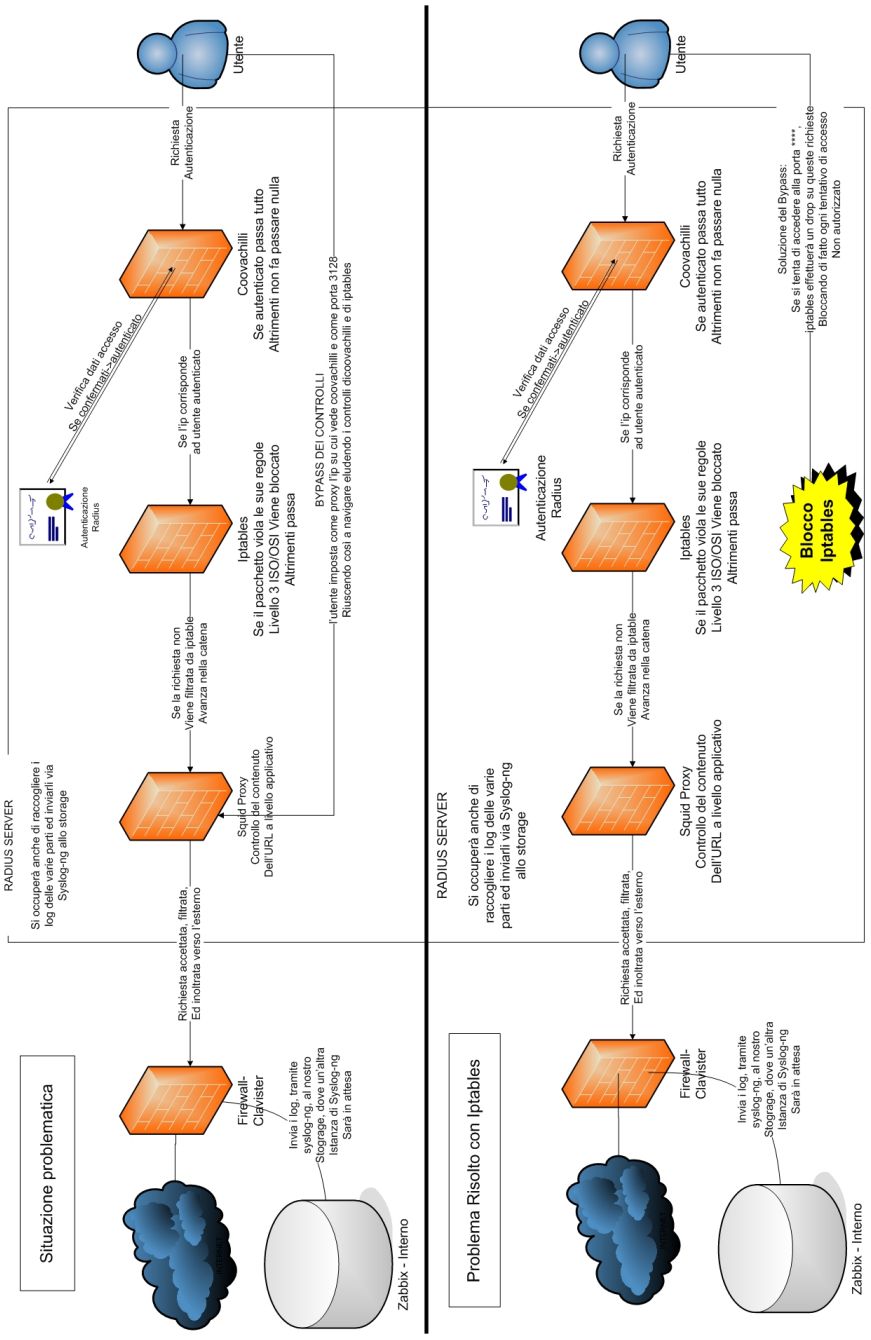


Figura 4.1: Visione a livello logico della comunicazione tra processi nei server Radius

che eliminasse questa problematica, lo schema su cui mi sono basato è riferito al comportamento di Iptables all'interno di una macchina, come mostrato in figura 4.2.

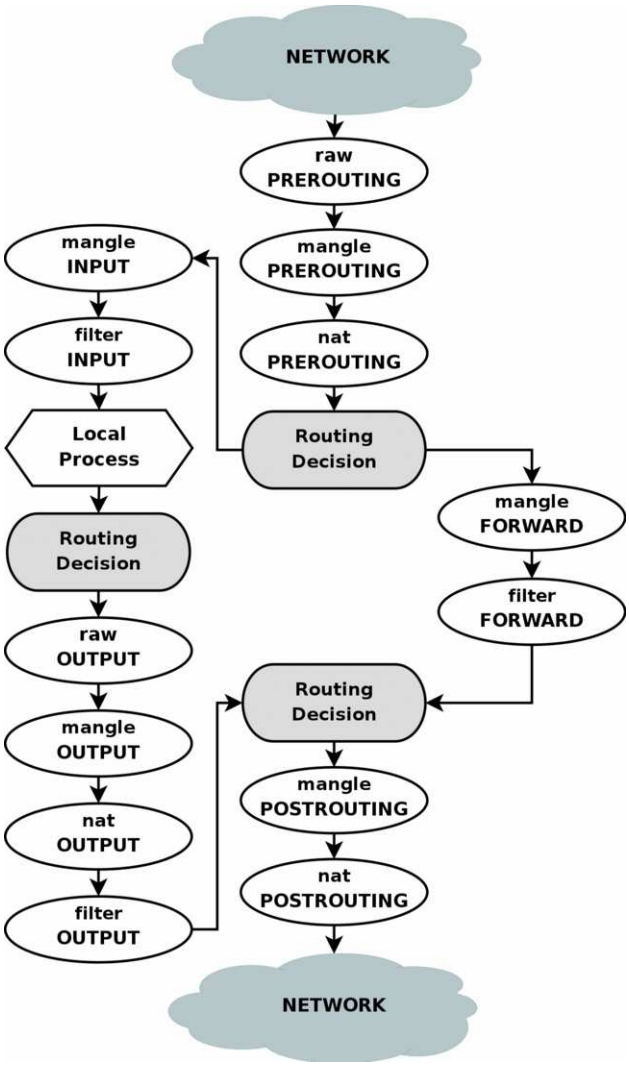


Figura 4.2: Diagramma flusso IPTables dentro una macchina

Le regole utilizzate per la risoluzione di questo problema sono le seguenti

Blocco di codice 4.1

```
#Accetto traffico diretto verso squid:
iptables -A INPUT -p tcp -m tcp --dport ---- --syn -j ACCEPT

# Droppo porta inutilizzata su cui redireziono chi tenta l'accesso
# bypassando Coovachilli
```

```
iptables -A INPUT -p tcp -m tcp --dport **** --syn -j DROP

# Traffico che prova ad accedere direttamente al proxy --> viene
# inoltrato su una porta inutilizzata
iptables -t nat -I PREROUTING -i eth0 -p tcp -m tcp --dport ----
--syn -j REDIRECT --to-ports ****

# Traffico autorizzato da Coovachilli resta sulla porta 80 --> lo
# invio su squid :
iptables -t nat -A PREROUTING -i eth0 -p tcp -m tcp --dport 80
-j REDIRECT --to-ports ----
```

Dove con la prima istruzione accetto tutto il traffico che è indirizzato verso squid, mentre nella seconda istruzione butto via ogni richiesta che è indirizzata verso una porta (con alto numero per evitare errori frequenti alle richieste degli utenti), le due operazioni di pre-routing servono rispettivamente, la prima reindirizza sulla porta chiusa in precedenza tutte le richieste indirizzate dagli utenti direttamente a squid, evitando così il bypass, la seconda invece indirizza le richieste che hanno superato il firewall di iptables e l'autenticazione di coovachilli verso squid.

Fondamentale, per questo set di istruzioni, è l'ordine in cui vengono inserite, infatti, se si invertono due qualunque tra le righe sopra definite, viene perso ogni pacchetto, o vengono accettati tutti.

Da notare che in un primo periodo questa soluzione causava un sottile problema, difficle da notare, che è emerso in fase di testing dell'applicativo, analizzando i logs si è notato che, questa sequenza di comandi sul netfilter, non permettava il logging di tutti i pacchetti. Il log nella tabella del firewall avviene in fase di FILTER, mentre i comandi sopra riportati vengono eseguiti in fase di PREROUTING, ovvero prima degli altri, è stata quindi aggiunto il logging anche in fase di PREROUTING per i pacchetti destinati alla porta 80.

4.4 Processi

Prima di procedere con l'implementazione, è stato necessario uno studio delle comunicazioni tra processi che avvengono all'interno dei server Radius, i servizi Coovachilli e Radius lavorano in tandem nel nostro caso.

Coovachilli è in costante attesa di utenti, Radius invece ha il ruolo di verificare ed eventualmente accettare, le credenziali che gli vengono fornite, effettuando tale

verifica all'interno del Database PostGres presente nel server stesso.

Nel momento in cui un utente si collega a Padova Wifi viene reindirizzato verso il Captive Portal, ovvero Coovachilli, il quale fornisce un IP non autenticato, finchè resta in questo stato, Coovachilli, blocca ogni tentativo di accesso dell'utente, ed accetta da questo IP unicamente l'inserimento di credenziali. Una volta inserite le credenziali da parte dell'utente, Coovachilli le inoltra a Radius, il quale potrà restiture una risposta negativa o affermativa; a questo punto Coovachilli salva sul database (PostGresql) il tentativo e l'esito dell'accesso, l'orario, e l'IP assegnato, e, in caso di risposta affermativa da parte di Radius, non blocca più nulla all'IP.

Iptables ha un ruolo 'passivo' in questo schema, svolgendo il suo ruolo di firewall ed inoltrando i logs a syslog-ng, Squid invece svolge il suo ruolo di Transparent Proxy e Web-Cache, pur inoltrando i logs all'istanza di Syslog presente sul Server. Per meglio chiarire la comunicazione che avviene tra i processi nell'autenticazione e nella raccolta dei logs, è riportato in figura 4.3 uno schema che rappresenta le varie richieste che può ricevere e l'invio dei logs verso il server centrale di raccolta. La ricezione dei log, invece, è semplicemente un syslog che

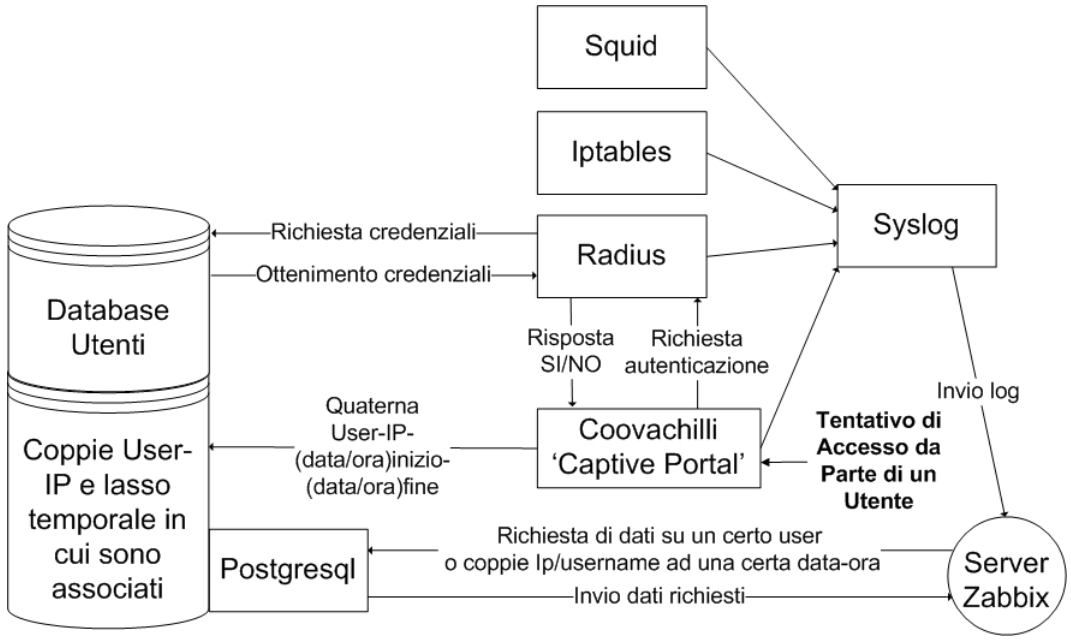


Figura 4.3: Schema comunicazioni tra processi reltivi al progetto all'interno di un server Radius

effettua il salvataggio organizzato di ciò che riceve, le regole in questo caso vanno

mantenute più semplici possibili, per evitare sovraccarichi di dati in ingresso nella macchina.

4.5 Configurazione

La configurazione utilizzata nei server oggetto del logging è suddivisa in varie sezioni, innanzitutto andavano allineati gli orari, per evitare logs sfalsati, cosa che avrebbe complicato l'incrocio di logs e delle informazioni in essi contenuti, mentre, mantenendo le ore allineate, le informazioni restano coerenti tra loro; lo stoccaggio e la configurazione per la centralizzazione andavano effettuate considerando la grande mole di dati da gestire e da memorizzare, senza quindi dimenticare le considerazioni dimensionali fatte in precedenza.

Analizzeremo ora le varie fasi che si sono susseguite per configurare correttamente le macchine per la sola raccolta e gestione dei logs.

4.5.1 Accuratezza temporale

Come definito nei requisiti, la precisione nella sincronizzazione tra le macchine è un aspetto cruciale, sia per avere informazioni coerenti (e non relative ad istanti precedenti o futuri nel sistema di riferimento dell'host considerato), sia per poter successivamente effettuare operazioni di analisi incrociata tra i dati contenuti nei log raccolti da diversi nodi di rete. Il Syslog Server e gli host attualmente impostati per inviargli i propri log sono stati dotati di supporto NTP, sincronizzandoli innanzitutto alla rete interna, per avere sincronizzazione locale, dove ciò non fosse possibile, si propenderà a server NTP di strati ad indice basso (ossia più vicini ai nodi radice, dove si trovano gli orologi atomici) nella gerarchia dei Time Servers. Per esempio, le 3 righe seguenti, aggiunte al file */etc/ntp.conf*, istruiscono il Syslog Server riguardo ai pool server NTP (di strato 3) cui connettersi periodicamente (dal più prioritario al meno prioritario) per effettuare l'aggiornamento del clock di sistema, il primo sarà il clock definito dalla macchina che funge da LDAP aziendale, così da avere una sincronizzazione locale tra le macchina in azienda:

Blocco di codice 4.2

```
server PDC.dominio.net
server 1.it.pool.ntp.org
server 1.europe.pool.ntp.org
```

Durante questa operazione, si è notato che alcuni host erano sfasati anche di diversi minuti rispetto all'ora corrente. Per esempio, la macchina di Padova WiFi, il cui pool server preimpostato era stato rimosso, è stato nuovamente impostato in maniera tale da sincronizzarsi periodicamente.

In futuro, quando si aggiungeranno nuove macchine all'insieme di host che inviano i propri log, sarà necessario predisporre le configurazioni NTP. Alcuni programmi utili sono *ntpd*, demone NTP in grado di funzionare sia in ambiente Linux che in Windows, e SNTP (Simple NTP), impiegato soprattutto in sistemi embedded ma non altrettanto preciso.

4.5.2 Stoccaggio

I logs vengono immagazzinati all'interno di un NAS (Windows) di rete, sul quale è stata crata una cartella, dedicata alla piattaforma di Log Mining, la cartella è montata all'interno della macchina, cosicchè è possibile utilizzarla come se fosse fisicamente presente all'interno del server di stoccaggio; che quindi, in realtà, è un server di stoccaggio solo a livello logico, fisicamente i dati risiedono su una macchina Windows che condivide un proprio disco in rete.

La gestione dei log, una volta immagazzinati, è gestita da una serie di script Bash e Perl, questi script hanno la funzione di effettuare il logrotate e la compressione dei dati salvati. A causa di un bug presente nella macchina Linux, ed in particolare un bug di Samba, è necessario smontare e rimontare la cartella ad ogni esecuzione dello script, mentre l'eliminazione dei log datati è affidata ad un'altro script : clean.pl, che verrà descritto nei prossimi paragrafi.

Lo Storage Server Il Data Storage Server è una semplice macchina che dev'essere dotata di notevoli capacità di memorizzazione, dovendo essere utilizzata come archivio (raccoglitore) di tutti i file di log provenienti dai vari sistemi di logging. A Telerete, la macchina adottata è un NAS da 2 TB con ridondanza RAID5 (circa 1.3 TB effettivi), che monta il sistema operativo Windows NT 4.9 Server.

Lo Storage Server viene montato sul Web Server grazie al file system Samba (*smbfs*, informazioni al sito http://www.samba.org), e viene utilizzato per archiviare le informazioni richieste dalla legge anti-terrorismo; queste informazioni

sono disponibili in forma di file di log zippati. Il NAS mette a disposizione le caratteristiche di affidabilità e persistenza, derivanti dalla ridondanza apportata dal RAID (Redundant Array of Inexpensive Disks). Le funzionalità di questa macchina consistono nel fornire la garanzia di integrità dei dati e la *data retention* nei tempi dovuti (6 + 6 mesi).

4.5.3 Syslog

Syslog (abbreviazione di System Log) è uno standard per l'invio di messaggi di log in una rete IP. Il termine 'syslog' viene utilizzato per indicare sia l'effettivo protocollo Syslog, sia per l'applicazione o la libreria che si occupa della spedizione e della ricezione dei messaggi di log.

Syslog e un protocollo di tipo client/server: il syslog sender invia un piccolo messaggio testuale (al massimo 1 KB o 1024 caratteri) al syslog receiver. Questo syslog receiver viene comunemente chiamato 'syslogd', 'syslog daemon' o 'syslog server'.

I messaggi Syslog possono essere inviati sia via UDP sia via TCP e vengono generalmente spediti in chiaro: sebbene non faccia parte delle specifiche del protocollo originario, è possibile utilizzare un wrapper in grado di fornire cifratura alla connessione tramite SSL/TLS. Per fare un esempio, un'applicazione Syslog viene spesso impiegata in simbiosi con stunnel (http://www.stunnel.org/).

Syslog può quindi essere sfruttato per integrare informazioni di log provenienti da differenti sistemi, convogliandole in un'unica repository centralizzata. La porta assegnata dalla IANA (Internet Assigned Numbers Authority) al protocollo Syslog è la 514: bisogna però prestare attenzione in quanto la porta registrata è relativa al solo protocollo UDP, mentre la 514/TCP è allocata al protocollo shell (cmd), assicurandosi comunque che la porta in questione non venga già impiegata dalla shell, nulla vieta all'istanza Syslog di utilizzare la porta 514 in TCP. La porta 601, assegnata a syslog-conn (descritto nel RFC 3195), prevede l'utilizzo di entrambi i protocolli di trasporto. Infine, la porta 6514 di TCP è associata all'estensione Syslog over TLS (standardizzato in RFC 5425).

La scelta è quindi stata quella di utilizzare Syslog-NG: un'implementazione open source del protocollo Syslog, che estende le funzionalità dell'originale syslogd, con alcune aggiunte che lo hanno reso piu versatile ed adattabile.

I punti di forza nell'adozione di Syslog-NG sono:

- nuova implementazione di un protocollo standard,
- compatibilità col suo predecessore, il daemon 'syslogd',
- compatibilità con apparati Linux, Unix-based e molti dispositivi di rete (Syslog è presente nativamente nei devices prodotti da Cisco), certi applicativi Windows comunicano con questo protocollo,
- il software è open source offre una buona versatilità a livello di trasporto, permettendo di stabilire se utilizzare TCP oppure UDP (unica opzione nel caso di syslogd),
- il numero della porta è impostabile (514 è la porta di default assegnata dalla IANA per il servizio di Syslog),
- effettua un remote logging di tipo incrementale. Nel nostro caso, questa è una caratteristica molto importante: infatti, considerate le dimensioni che possono raggiungere alcuni dei nostri log (fino ad alcuni GB/giorno), costruire il file di log remoto mano a mano che giungono le log entries (alcuni KB/s) è decisamente meno oneroso per la rete rispetto al trasferimento dell'intero file, operazione che rischia di sovraccaricare la rete,
- fornisce la possibilità di effettuare secure logging, adottando meccanismi di crittografia basati sul tunneling SSH oppure su SSL / TLS. Questa proprietà può essere ritenuta trascurabile per log che vengono trasferiti all'interno della intranet aziendale, considerata sicura, ma rappresenta invece una caratteristica essenziale qualora le informazioni loggate debbano transitare attraverso la rete internet o altre reti non date,
- fornisce la possibilità di riscrivere, o di aggiungere informazioni, allo stesso header del pacchetto syslog.

Altra soluzione, che potrebbe rivelarsi molto efficiente sotto l'aspetto del throughput, potrebbe essere l'instaurazione di una VPN ad-hoc sicura (per esempio tramite IPsec oppure OpenVPN) a supporto di un semplice trasferimento FTP, ma per non complicarsi la vita è stato scelto Syslog. La tabella 4.2 elenca le varie fonti di log e le rispettive logging facilities:

Syslog sui server Radius

Segue, a titolo di esempio di configurazione di un'istanza di Syslog-NG, un estratto dal file *syslog-ng.conf* del server Radius di Padova WiFi, installato su una

Tabella 4.2: Log sources

Log source	*Logging facility*	*Descrizione*
Authentication	auth	accessi SSH
Authentication	authpriv	apertura sessioni SSH
Cron	cron	cron job eseguiti
IPTables	kern	pacchetti a livello 3 (IP)
dhcpd	local1	warning errori del demone DHCP
Coova-Chilli	local6	login e messaggi DHCP
dhcpd	local7	info del demone DHCP
sSMTP	mail	errori del demone di posta sSMTP
Squid	user	informazioni provenienti dal web-proxy
Syslog-NG	syslog	info del demone syslog

macchina Gentoo Linux. I due *statement* sono quelli relativi alla destinazione cui far pervenire i log, e alla effettiva regola di logging, la fonte, essendo quella di default, non è stata riportata. Interessa far notare che i log vengono trasportati sulla porta 514 (default) di TCP.

Blocco di codice 4.3

```
destination remote { tcp("**.**.**.**"); };
log { source(src); destination(remote); };
```

Al Syslog Server pervengono, attualmente, tutte le log entries gestite dal client Syslog-NG montato sul server di Padova WiFi. Le logging facilities definite localmente presso il syslog client sono i valori che danno poi il nome ai file generati dall'istanza di syslog presente presso il Web Server.

Rsyslog di Monselice WiFi Anche il server Radius di Monselice WiFi è stato configurato in modo tale da inviare i propri file di log al Web Server. La macchina server di Monselice WiFi presenta alcune differenze rispetto a quella di Padova WiFi. Innanzitutto monta un sistema operativo diverso, ossia una Linux Fedora Core. Ad ogni modo, la differenza più importante per quanto riguarda i nostri

obiettivi, è che questo host non utilizza Syslog-NG, bensì Rsyslog. I comandi fondamentali per perseguire il nostro scopo sono stati i seguenti:

Blocco di codice 4.4

```
local6.*                /var/log/coova.log
&                       @@**.**.**.***
kern.warn;kern.info     /var/log/iptables.log
&                       @@**.**.**.***
```

Il trasferimento avviene via TCP (avendo definito '@@', mentre '@' avrebbe utilizzato UDP), ancora una volta sulla porta 514 di default. Vengono spedite tutte le entry relative al Coova-Chilli (associato alla logging facility *local6*) e a IPTables (associato alla logging facility *kern*), il comando '&' indica di inviare le stesse informazioni in più destinazioni, nel nostro caso sia in remoto che su file locale.

4.5.4 Configurazione Server Zabbix

La configurazione di Syslog-NG nel Web Server è stata modificata in maniera tale da eseguire automaticamente la log rotation giornaliera. In particolare, i file di log vengono organizzati secondo precise regole di formattazione, determinate dai seguenti parametri: host di provenienza, anno, mese, giorno, logging facility (che rappresenta, in breve, il componente di sistema che genera quella log entry).

Seguono le righe di codice che istruiscono il Syslog-NG server a creare i diversi file di log, a seconda dei valori assunti da *$HOST*, *$YEAR*, *$MONTH*, *$FACILITY* e *$DAY*, tutti parametri definiti dal protocollo Syslog. Sono presenti anche i comandi relativi ai permessi che attribuiscono a Syslog-NG la facoltà di creare i file e le tabelle di cui necessita per rispettare lo schema descritto.

Blocco di codice 4.5

```
source logsudp { udp(ip(0.0.0.0)); };
source logstcp { tcp(ip(0.0.0.0)); };
destination logfile {
    file("/root/local/NAS/$HOST/$YEAR/$MONTH/
         $FACILITY/$FACILITY.$YEAR$MONTH$DAY"
    owner(apache) group(apache) perm(0600)
```

```
        create_dirs(yes) dir_owner(apache)
        dir_group(apache) dir_perm(0700));
};
```

Vediamo come tale configurazione di Syslog-NG imponga una struttura ben definita e prestabilita per l'insieme delle directory e dei file utilizzati, insieme che va a comporre un albero gerarchico (con radice */root/local/NAS/*) di formato noto all'interno del file system di Linux.

4.5.5 Gestione Log

Il Web Server è stato poi impostato per eseguire un *cron job* notturno, il quale ha il compito di comprimere i log file chiusi, diciamo tutti i file relativi al giorno solare precedente al giorno corrente. Il cron job viene attualmente eseguito alle 2 di notte (2:00 AM), in una fascia in cui, molto probabilmente, la macchina Linux è impegnata soltanto in minima parte. Il *cron job* è stato impostato aggiungendo in coda al file */etc/crontab* la seguente riga:

Blocco di codice 4.6

```
0  2  *  *  *     root    /opt/log_find.sh
```

Notiamo come lo script venga eseguito dall'utente `root`, in maniera tale da attribuire al programma tutti i permessi di cui necessita. Lo script *log_find.sh* è stato scritto in Bash Shell, ed è composto da due soli comandi effettivi. Il primo di questi, ricerca e seleziona tutti i file contenuti nella base directory */root/local/NAS/* in maniera ricorsiva (ossia anche nelle sottocartelle di quella cartella), che non siano già stati compressi (non aventi quindi già l'estensione .gz) e che non siano relativi al giorno corrente (in quanto è possibile Syslog ci stia ancora scrivendo all'interno). Il fullpath di questi file viene fornito come parametro in input allo script *log_ zip.sh*.

Blocco di codice 4.7

```
/usr/bin/find /root/local/NAS/
    ! -name "*.gz" -type f
    ! -path "*‘/bin/date +%Y%m%d‘"
    -exec /opt/log_zip.sh '{}' \;
```

Anche */opt/log_zip.sh* è stato scritto in Bash Shell. Le operazioni eseguite da questo script sono principalmente le seguenti:

1. prende in input il fullpath di un file (passatogli da */opt/log_find.sh*),
2. copia il file in una directory temporanea,
3. esegue la compressione del file (NOTA: il nome del file rimane il medesimo, con l'aggiunta dell'estensione '.gz'),
4. ricopia il file compresso nella directory originaria,
5. copia il file compresso in una directory nel disco fisso dello Storage Server (NAS01). Qualora non fosse presente la directory opportuna, lo script gode dei permessi per crearla (anche qui per mezzo dei parametri $HOST, $YEAR, $MONTH, $FACILITY),
6. rimuove il file compresso dalla directory temporanea (qualora questa operazione non fosse già stata eseguita dal software di compressione) e rimuove il file originario dalla directory di provenienza.

Rimozione dei vecchi log file

Avendo a che fare con file di log di grandi dimensioni (fino ad alcuni GB/giorno, circa 100 MB/giorno per il file compresso con GZip), si rende necessario definire delle *policy* di cancellazione dei log file di vecchia data; l'assenza di questo genere di regole comporterebbe infatti, nel lungo periodo, un esaurimento delle capacità di memorizzazione su disco rigido. Considerati i requisiti temporali relativi alla *data retention* da soddisfare per ottemperare alle norme di legge, è stato predisposto uno script Perl, richiamato alla fine del *cron job* descritto in precedenza, con il compito di controllare il timestamp di creazione di tutti i log salvati, rimuovendo i file più vecchi di un anno. Quest'ultimo valore potrà essere variato in maniera molto semplice, a seconda degli interessi di analisi dei dati da parte dell'area tecnica di Telerete, ma tenendo sempre ben presenti i requisiti richiesti dalla legge in vigore.

Il secondo comando dello script */opt/log_find.sh* richiama semplicemente il programma Perl responsabile della rimozione dei vecchi logfile.

Blocco di codice 4.8

```
perl /opt/clean.pl;
```

Lo script analizza i timestamp relativi a tutti i file presenti nella base directory impostata come parametro all'inizio del codice: attualmente il parametro `$base_dir` punta a */root/local/NAS/*. Anche in questo caso la ricerca avviene ricorsivamente su tutte le sottocartelle della base directory: il comando impiegato è ancora `/usr/bin/find`. Il timestamp che viene controllato è il decimo parametro (`$stat[9]`) dell'array restituito dal comando `/usr/bin/stat($file);` . Il valore, che rappresenta l'*epoch time* trascorso dall'ultima modifica, viene attualmente confrontato con 31536000 secondi (corrispondenti a 365 giorni) per determinare se il file è da rimuovere oppure va lasciato dov'è; ovviamente tale valore potrà essere portato a 15811200 (6 mesi, secondo la *data retention* richiesta per legge) o ad un'altra cifra in qualsiasi momento. Lo script produce un file di log, */opt/clean_file.log*, che registra con modalità appending tutte le invocazioni dello script e tutte le rimozioni da esso operate.

Sistema di alerting

Durante la fase terminale del lavoro si é notata la necessitá di introdurre un sistema che avverta gli amministratori in caso di problemi. Questo sistema sfrutta un'utility integrata di Linux, `ssntp`, un sistema in grado di inviare mail senza la necessitá di un operatore, caratteristica fondamentale per essere sfruttata all'interno di uno script.

Queste mail vengono inviate dopo la compressione ad opera dello script precedentemente descritto, viene compiuta una rapida verifica delle dimensioni del logfile compresso, nel caso in sui il log sia di dimensione minore alla media, verrá inviata la mail di alert agli amministratori, che avranno il compito di verificare le motivazioni di questo sotto-dimensionamento.

Verrá ora mostrato un estratto delo script che svolge questo compito:

Blocco di codice 4.9

```
info=‘du -lah | grep kern | grep ${ieri} |
 grep -e "^[[1-9][0-9]\|^[3-9]][0-9]M.*$"‘
if [ "${info[0]}" ]
then
echo "Ok,dimensioni file sono nella norma"
else
cat /opt/mailheader /var/log/temp.log |
 ssmtp ++@telerete.net
fi
```

4.6 Testing e Struttura finale

La struttura generale del sistema di logging può essere semplificata come in figura 4.4, sapendo che in realtà i logs provengono da altre istanze di Syslog presenti su altri Server, i quali li inoltrano in remoto al Server Centrale.

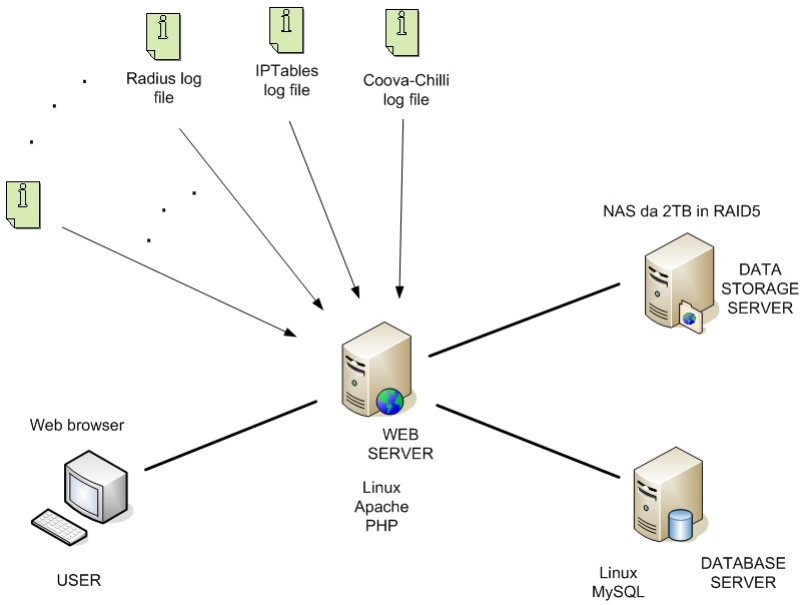


Figura 4.4: Struttura del raccogilmento dei logs

Ad ogni passaggio che veniva portato a termine, era necessaria una fase di testing, durante la quale si verificava che effettivamente tutto funzionasse, per prima cosa è necessario verificare che i logs siano corretti, che i log vengano inviati, arrivino a destinazione e che vengano gestiti correttamente, infine va verificata la completezza ed eventuali dati mancanti.

Il metodo più semplice per verificare che syslog funzionasse era di salvare i logs da inviare su di un file temporaneo sulla macchina in analisi ed analizzare il logs, successivamente, per controllare che l'invio avvenga regolarmente e non crei problemi si utilizza il comando :

Blocco di codice 4.10

```
tcpdump host ***.***.*.**
```

Indicando l'IP dell'host al quale vanno inviati dati o dal quale vengono ricevuti, infatti lo stesso test andava fatto anche alla destinazione dei logs, per le macchine

windows invece conviene utilizzare wireshark se possibile, per ottenere maggiori informazioni sulle tempistiche e sui percorsi di rete compiuti dai pacchetti.

Infine, se tutto funziona correttamente, si effettua una veloce verifica della struttura delle cartelleche Syslog va a creare e dei logfiles che vengono salvati. In caso di problemi la causa può essere, per lo più, un firewall posto tra le macchine, che impedisce la comunicazione, una volta aperte le porte necessarie raramente accadono in altri intoppi.

Documentando con regolarità ho notato che, col passare del tempo, compiti che inizialmente parevano complessi e per i quali perdevo parecchio tempo, sono ora diventati quasi routinari, in caso di problemi le difficoltà sono spesso già state affrontate, o se non altro valutate, e, disponendo di una documentazione organizzata, si trovano soluzioni rapidamente.

Capitolo 5

Implementazione Sistema di analisi

Una volta effettuata la raccolta ci si è trovati di fronte alla necessità di effettuare un'analisi veloce ed efficente dei dati raccolti, per soddisfare questa necessitá si é scelta una soluzione classica, ovvero un'interfaccia web, nel nostro caso scritta in PHP e gestita da Apache (un web-server), inoltre verranno sfruttate le potenzialitá di MySQL (un DBMS).

5.1 Analisi del Problema

In un primo momento ci siamo concentrati sull'estrazione delle informazioni richieste dalla legge Pisanu 3.2. In particolare si vuole sapere, in ciascun momento o periodo temporale, chi era connesso ad Internet e quali pagine web stava visitando. Per ottenere queste informazioni è sufficiente analizzare le informazioni di registrazione degli utenti, contenute in un apposito database Postgresql, e le informazioni presenti nei log e nel database interno del server Radius. In azienda viene utilizzato FreeRadius, un'implementazione open-source del protocollo di autenticazione, autorizzazione e accounting (brevemente AAA), realizzata per funzionare in ambienti Linux e Unix-like in generale (v. appendice B.2). Scenario tipico:

1. gli organi di competenza (es. Magistratura, Polizia Postale e delle Comunicazioni) richiede chi ha visitato un determinato sito Internet in una determinata fascia oraria. Nota: tutti gli utenti accedono alla intranet di Telerete con un IP-address unico all'interno di quella rete, ma sono visibili all'esterno (in Internet) tutti con il medesimo indirizzo IP, l'IP-address associato a Telerete. Questo è un aspetto tipico della Network Address Translation (brevemente NAT), operazione effettuata dai firewall aziendali

(per esempio il Clavister). Di conseguenza, se viene commesso qualche illecito in Rete, l'indirizzo tracciato è quello di Telerete, da qui la necessitá di poter rilevare l'effetiva identitá della persona che ha visitato ogni sito, la cui richeista é associabile a Telerete.

2. dal sito Internet si ricava l'indirizzo IP del Server che ospitante, tramite una semplice richiesta DNS.

3. dal file di log *messages* (collocato di default nella cartella */var/log/*) di `IPTables`, si selezionano le entry relative a quel periodo temporale che contengono quell'indirizzo IP.

4. si ottiene così l'IP-address interno dell'utente richiedente (eventualmente anche più di uno), ossia colui che ha visitato quel sito in quella fascia oraria.

5. dalla tabella *radacct* del database di Radius si ottiene l'identificativo cliente associato a quell'IP.

6. con l'ID del cliente si accede alla tabella *cliente*, contenente i dati degli utenti (dati forniti dagli stessi durante la fase obbligatoria di registrazione al servizio) e si ottengono tutte le informazioni di identità, contatto e recapito relative a quell'utente.

7. in caso di necessitá é possibile verificare i dettagli della richiesta (se HTTP) grazie ai log forniti da Squid.

5.1.1 Casi d'uso del sistema - Legge Antiterrorismo

L'interazione tra utente e sistema avviene unicamente via interfaccia web. Come detto, *start* e *end* della fascia temporale sono campi *required*, per discriminare la porzione di log entry da estrarre ed inserire nel database. I campi *ipaddr* e *login* sono anch'essi obbligatori da inserire, ma possono valere '*'. Le possibili combinazioni, derivate dagli input utente, definiscono i seguenti 4 casi d'uso dell'applicazione:

1. *ipaddr* e *login* valgono entrambi *

2. *ipaddr* è specificato mentre *login* vale *

3. *ipaddr* vale * mentre *login* è specificato

4. *ipaddr* e *login* vengono entrambi specificati

5. *request* specificata da iptables e si desidera sapere l'URL

6. *fascia temporale* se sidesiderano tutte le richieste HTML è sufficiente specificare il lasso temporale in Squid

Case 1 A seconda della fascia temporale richiesta vengono inserite nel database le informazioni dai log relative a quel range. Non essendo state fornite esplicitamente le altre due informazioni, si suppone che l'utente voglia semplicemente avere accesso alla consultazione di tutte le informazioni contenute nel database e relative a quell'arco di tempo. In seguito alla popolazione del database viene quindi fornita all'utente la possibilità di effettuare query in lettura (SELECT) al database MySQL appena creato. Per fare ciò, una volta che la tabella è stata popolata viene visualizzata una pagina web con un form HTML che prevede l'inserimento di un input testuale, in cui il tecnico digita la SELECT SQL; i risultati vengono visualizzati sulla stessa webpage.

Se si desiderano dettagli ulteriori su una determinata richiesta è sufficiente cliccare su Dettagli, si verrà riportati sulla pagina di squid con i dettagli sulla richiesta; Questo vale anche per il case 2.

Case 2 In questo caso si vuole sapere chi, all'interno della rete wireless, ha effettuato una richiesta (possibilmente anche più di una persona, o anche nessuno) verso un certo sito. Pertanto, dopo la fase di popolazione della tabella, l'applicazione PHP effettua una ricerca dell'IP address specificato in input. La ricerca avviene sul campo "*destination*" della tabella MySQL, e risale agli indirizzi IP "*source*" dell'utente (eventualmente degli utenti) che ha effettuato una connessione verso quell'indirizzo. Una volta in possesso di questo IP address, l'applicazione è in grado di risalire al login dell'utente legato a quell'IP address: questa operazione è possibile consultando la tabella *public.radacct* del database di FreeRadius, vedendo la corrispondenza tra "*framedipaddress*" e "*username*" nel preciso istante temporale in cui è avvenuta la richiesta. Infine, l'applicazione ricava tutti i dati del cliente dalla tabella *public.cliente* del database di FreeRadius, reperendo le informazioni anagrafiche associate a quel "*idutentebase*".

Case 3 Quì si suppone che l'utente tecnico voglia sapere, per conto della Magistratura, quali sono state le richieste che ha effettuato un dato utente cliente all'interno di una precisa fascia temporale. Dal nome utente (il login specificato), l'applicazione ricerca l'indirizzo IP assegnatogli dal server DHCP in quella fascia

temporale: questa informazione è presente nella tabella *public.radacct* del database di FreeRadius, seguendo l'associazione tra "*username*" e "*framedipaddress*". L'indirizzo IP ricavato diventa poi la "*source*" all'interno del database MySQL, valore che permette di risalire a tutte le informazioni registrate dai log (in particolare alle "*destination*") relative alle richieste effettuate da quell'utente cliente in quel lasso di tempo.

Case 4 Essendo stati forniti in input sia l'IP address richiesto sia lo username del cliente richiedente, l'output sarà un valore booleano che è vero se e solo se quel cliente ha effettivamente effettuato uno scambio di pacchetti con il sito legato a quell'indirizzo IP. Questo caso segue il flusso descritto per il caso precedente, con un'aggiunta finale: viene effettuato un controllo che verifica se l'IP address in input fa parte dell'insieme composto da tutte le "*destination*" ricavate al termine dell'esecuzione del caso 3.

Case 5 Nel caso in cui un tecnico necessiti di informazioni aggiuntive in merito ad una determinata richiesta, già trovata, è sufficiente cliccare sull link dettagli, il quale passerà alla pagina Squid il compito di fornire i dati richiesti.

Case 6 Qui si suppone che l'utente tecnico sia in cerca di informazioni in merito a contenuti e URL visitati in un certo lasso temporale, a richiesta in questo caso è da effettuare direttamente sulla pagina di Squid, verranno caricate tutte le righe che rientrano nel lasso temporale specificato e sarà possibile effetuare richerche semplificate grazie ai form presenti, per effettuare la costruzione delle query per i parametri voluti basterà cliccare su 'Costruire query'.

5.1.2 Requisiti e regole

Sia durante la progettazione dell'applicativo, che nella fase di sviluppo, sono stati tenuti presenti i requisiti di progetto e si è lavorato in maniera tale da rispettare alcune regole prefissate:

- realizzare un sistema utile ai dipendenti dell'area tecnica. Per questo sono state effettuate, durante tutto l'arco temporale di sviluppo, delle interviste sulle funzionalità richieste, sulle informazioni di interesse e su altre questioni pratiche per ottimizzarne l'utilità;

- realizzare un sistema centralizzato estendibile. L'implementazione PHP è stata portata avanti in un'ottica di modularità ed estendibilità. Le funzionalità principali dell'applicativo non sono applicabili solamente ai particolari sistemi di logging analizzati, ma possono essere utilizzate per qualsiasi sistema di logging interessi in futuro, qualunque sia il formato adottato, semplicemente configurandone i parametri nei file di configurazione. Alcuni moduli e funzioni, per quanto non attualmente utilizzati, sono stati aggiunti ritenendo possano con ogni probabilità tornare utili in un futuro prossimo.

- realizzare un sistema efficiente. Lavorando con file di grandi dimensioni (diversi gigabyte), l'efficienza temporale non è un obiettivo facile da perseguire. Numerose scelte di implementazione sono state prese allo scopo di minimizzare i tempi di esecuzione dell'applicativo, dovendo a volte sacrificare la comprensibilità e l'eleganza nella stesura del codice.

- realizzare un sistema a costo zero per l'azienda. L'utilizzo di applicazioni e tecnologie open source, facilmente reperibili in Rete, ha permesso di ottenere un sistema a costo zero sia per quanto riguarda i costi di realizzazione iniziali, sia per quanto riguarda i costi di gestione, configurazione ed estensione futuri.

- realizzare un sistema semplice. La semplicità non riguarda solamente l'immediatezza nell'utilizzo dell'applicativo, ma anche la possibilità di capire il funzionamento sottostante del programma e delle funzionalità che mette a disposizione, in particolare in ottica estendibilità: per questo motivo, si è cercato di mantenere il codice il più pulito e leggibile possibile, rispettoso delle buone norme e comuni pratiche di programmazione, oltre che documentato quantomeno nei passaggi non banali.

5.1.3 Fonti utilizzate

Seguirá una descrizione dei log e delle fonti di dati che sono state identificate come piú appropriate, o contenenti informazioni utili alla risoluzione del problema sopra descritto.

Subito è saltata all'occhio la necessitá di prelevare i log forniti da Coovachilli, essendo gli unici che forniscono la vera associazione Utente-IP-fascia oraria, successivamente per associare il nome utente ad un'identitá sono stati selezionati anche i log di Radius, ora trovsti di fronte alla richeista bisogna identificare quale IP l'ha effettuata e verso dove era diretta, i log del DNS potrebbero sembrare adatti, ma un utente accorto potrebbe utilizzare indirizzi IP, senza far quindi passare la propria richiesta attraverso il DNS, i log forniti da Iptables sono parsi i piú adatti.

Infine andavano registrate informazioni piú precise in merito al contenuto del server che é stato visitato, questo perché esistono server che contentogono sia informazioni utili che illegali; va quindi tenuta traccia dei contenuti visitati. I dati che i vari log forniscono sono stati descritti nel capitolo precedente, ora analizzeremo il contenuto effetivo e la struttura dei logs.

Coovachilli

Da coovachilli nel nostro caso erano interessanti più dati:

- Il modulo coovachilli richiede al volo dati sul numero di utenti connessi con successo nell'ultima ora
- il modulo IpTrace utilizza le informazioni di coovachilli per collegare gli IP rispetto al nome utente, per ogni lasso temporale
- Il modulo IpTrace utilizza non soloi log ma anche i dati memorizzati sul Database Postgresql presente sui server Radius
- Vengono utilizzati i log di Coovachilli per venire a conoscenza di tentativi di accesso falliti

Un esempio di log fornito da coovachilli è il seguente :

Blocco di codice 5.1

```
Feb 18 10:37:48 192.168.12.9 coova-chilli[29677]: chilli.c: 2694:
New DHCP request from MAC=00-12-F0-58-E7-B8
```

```
Feb 18 10:41:09 192.168.12.9 coova-chilli[29677]: chilli.c: 2785:
DHCP addr released by MAC=00-24-7D-59-C1-02 IP=10.192.0.43
Feb 18 10:47:49 192.168.12.9 coova-chilli[29677]: chilli.c: 2785:
DHCP addr released by MAC=00-12-F0-58-E7-B8 IP=0.0.0.0
Feb 18 11:28:21 192.168.12.9 coova-chilli[29677]: options.c: 809:
Rereading configuration file and doing DNS lookup
Feb 18 11:28:21 192.168.12.9 coova-chilli[29677]: chilli.c: 1010:
Unknown downlink protocol
Feb 18 12:03:54 192.168.12.9 coova-chilli[29677]: chilli.c: 2694:
New DHCP request from MAC=64-B9-E8-6E-17-07
Feb 18 12:03:54 192.168.12.9 coova-chilli[29677]: chilli.c: 2661:
Client MAC=64-B9-E8-6E-17-07 assigned IP 10.192.0.44
```

In cui si nota che Coovachilli assegna ad un MAC un determinato IP, nei dati di Coovachilli invece è presente la coppia di dati MAC:Username, invece in Radius sono presenti i dati personali dell'utente, da verificare in caso di richieste della Polizia Postale.

FreeRadius

Dai database utilizzati da FreeRadius sono recuperabili informazioni riguardo le credenziali, che sono utilizzate da Coovachilli al momento dell'autenticazione, mentre al nostro scopo verranno utilizzate queste informazioni direttamente, non essendo, come è giusto che sia, memorizzate all'interno di log.
I dati personali veranno ottenuti direttamente dal Database PostGres, mediante connessione ed interrogazioni remote.

Dai logs di FreeRadius verranno estratte informazioni per le future espansioni della piattaforma, quindi attualmente non verranno analizzati.

Iptables

I log forniti dal firewall sono utili ad identificare sorgente e destinazione di ogni pacchetto, sono i log più voluminosi essendo quelli più di basso livello.

Seguiranno alcune righe del protocollo NetFilter, o IPtables, in particolare è da sottolineare che i dati provenienti da tun0 sono i tentativi di autenitcazione, e tun0 è l'interfaccia virtuale fornita da Coovachilli, mentre frw sono i pacchetti che transito attraverso la macchina, e che provengono, a livello logico, da Coovachilli, e quindi se passano vuol dire che effettivamente l'autenticazione è avvenuta.

Blocco di codice 5.2

```
Feb  5 20:16:43 192.168.12.9 kernel: frw: IN=tun0 OUT=eth0
      SRC=10.192.0.109 DST=69.63.176.165 LEN=300 TOS=0x00
      PREC=0x00 TTL=63 ID=26633 DF PROTO=TCP SPT=52031 DPT=80
      WINDOW=65535 RES=0x00 ACK PSH URGP=0
Feb  5 20:34:40 192.168.12.9 kernel: frw: IN=tun0 OUT=eth0
      SRC=10.192.0.112 DST=217.148.122.39 LEN=77 TOS=0x00
      PREC=0x00 TTL=68 ID=16072 PROTO=UDP SPT=57509 DPT=53 LEN=57
Feb  5 20:34:43 192.168.12.9 kernel: tun0: IN=tun0 OUT= MAC=
      SRC=10.192.0.112 DST=10.192.0.2 LEN=64 TOS=0x00
      PREC=0x00 TTL=69 ID=16078 DF PROTO=TCP SPT=37287
      DPT=3990 WINDOW=64240 RES=0x00 SYN URGP=0
```

I dati che interessano maggiormente sono Data, Ora, Interfaccie utilizzate, IP di prevenienza e di destinazione, l'Ip di destinazione in particolare è quello del server di destinazione del pacchetto, questa informazione, assieme all'IP di provenienza, è utile per collegare i dati forniti da Squid con quelli di Iptables.

Altri dati rilevanti sono in fondo al log, se presenti gli ACK, i SYN fanno capire il motivo del transito del pacchetto; i dati rimanenti sono utili per capire l'andamento della rete, eventuali rallentamenti, saturazioni ecc.., saranno dati fondamentali per la piattaforma di analisi dei problemi di rete. Al momento, per analizzare questi dati, si utilizzano Regular Expression limitate dai dati costanti, ovvero i TAG presenti prima di ogni dato.

Squid

I log di squid erano necessari per avere informazioni sulle richieste individuate da IpTrace, ma è stato creato un modulo apposito per verificare tutti gli URL visitati e gli IP reali dei server visitati, leinformazioni fondamentali per capire cosa, realmente, gli utenti vedno all'interno di un server, si possono ricavare unicamente da qui.

Seguono alcune righe particolari di logs prelevati dal file del 06-02-2010 :

Blocco di codice 5.3

```
Feb  6 09:45:20 192.168.1.9 10.192.9.246 1567 - GET text/html
404 http://api.simile-widgets.org/exhibit/locales/locale.js
Feb  6 09:55:10 192.168.1.9 10.192.10.31 56852 - GET text/plain
200 http://www.msftncsi.com/ncsi.txt
```

```
Feb  6 20:23:41 192.168.1.9 10.192.15.46 1459 81.29.194.214 GET -
304 http://www.weblula.it/fengard/pulsanti/minipost.gif
Feb  6 20:36:43 192.168.1.9 10.192.15.120 5385 209.85.135.100 GET
  image/gif 200 http://www.google-analytics.com/__utm.gif?
Feb  6 20:36:43 192.168.1.9 10.192.12.130 542 207.46.124.202 POST
application/x-msn-messenger 200 http://207.46.124.202/
```

Si nota che non sempre si dispone di tutte le informazioni, il server ad esempio non compare nelle richieste che non hanno avuto successo (codice 404 = Not Found) o in quelle di semplice conferma (codice 200 = OK), talvolta invece non è specificato il tipo MIME della comunicazione, ma è un'informazione facilmente ricavabile dall'URL.

Le informazioni vitali invece sono sempre presenti, l'URL e l'IP dell'host che ha prodotto la richiesta.

5.2 Struttura Apache: Web server

Per poter affrontare bene la problematica dello sviluppo di una piattaforma è bene avere idea delle comunicazioni tra processi che avverranno all'interno della macchina che la ospiterà, la figura 5.1 è uno schema dei processi in esecuzione e della loro interazione, all'interno della macchina Zabbix.

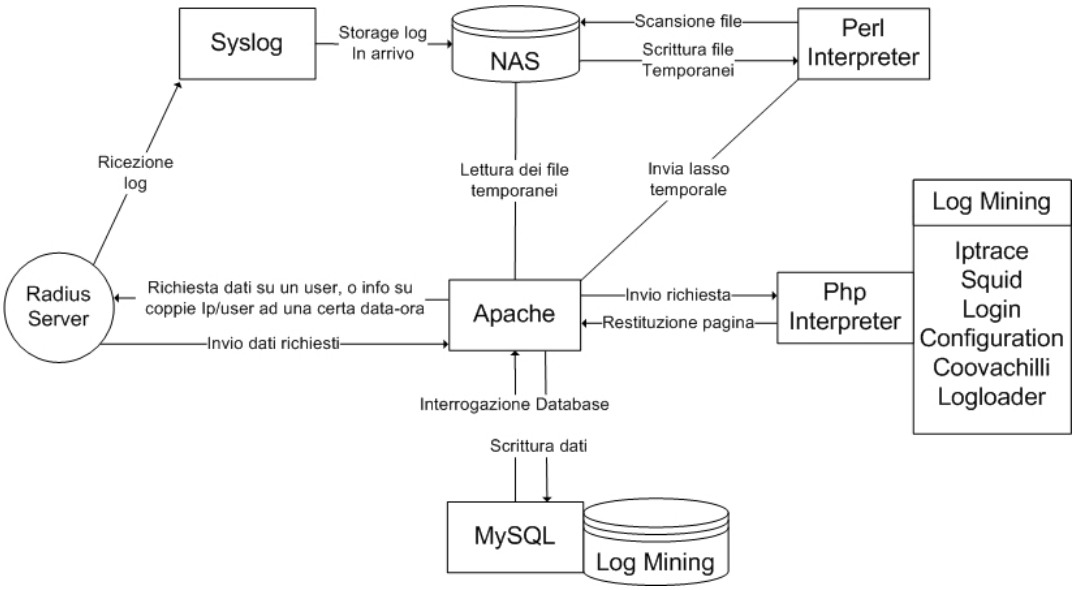


Figura 5.1: Schema della comunicazione tra processi, riguardanti il log mining, nel server Zabbix

Sono stati indicati anche i processi riguardanti lo storage ed il mantenimento ad opera rispettivamente di syslog-ng e di crontab. La struttura centrale in questo caso è Apache, il web-server che effettivamente smista le varie richieste. Il Web Server fornisce l'interfaccia di accesso al servizio che permette di analizzare le informazioni contentute nei file di log raccolti. Il sistema è attivo e raggiungibile come Web Service, messo a disposizione dal server Apache installato. L'utente finale (l'operatore dell'area tecnica) vi accede in HTTP dalla rete interna per mezzo di un comune browser, e nella pagina visualizzata inserisce gli input in appositi form.

5.2.1 Database Server: MySQL

Il Database Server è una macchina Linux, con un database installato: il DBMS (*DataBase Management System*) potrebbe essere di qualsiasi tipo, ma i test ed il deployment sono stati eseguiti utilizzando MySQL. Attualmente il database viene creato on-demand su richiesta dell'operatore tecnico: in sostanza, quando l'utente inserisce nel form gli input richiesti, l'applicativo PHP inserisce automaticamente i dati relativi alla richiesta effettuata.

La creazione della tabella e l'inserimento dei dati in essa sono operazioni che richiedono *full access* sul database MySQL. Per questo motivo, l'accesso allo script PHP viene garantito con l'account di "root". Dall'altra parte invece, l'accesso sul database già creato dovrebbe essere in sola lettura, si è quindi creato un secondo account chiamato "readonly" (sono ammessi soltanto i SELECT SQL).

Le tabelle principali sono quelle relative ai vari moduli :

- coovachilli
- iptables
- squid

le tabelle relative alle sole sessioni degli accessi alla piattaforma stessa sono :

- logged_users
- user
- active_users

Non essendo la descrizione delle tabelle degli accessi, strettamente inerente al problema che la piattaforma si propone di risolvere, verranno escluse dalla trattazione, è sufficente essere a conoscenza della loro esistenza e della loro utilità.

Coovachilli : La tabella squid serve a memorizzare i dati relativi alle connessioni giornaliere, utile per identificare i trend di utilizzo dei servizi WiFi che Ne-t offre, viene aggiornata ogni ora mediante degli script che Cron richiama, mentre, per non affollare la tabella di dati inutili, vengono eliminate tutte le entry più vecchie di 365 giorni, in modo tale da fornire sia un trend giornaliero che annuale. La struttura della tabella è

1. Data
2. Data e Ora
3. 24 colonne, una per ogni ora della giornata
4. Log totali della giornata

Questo sistema è facilmente espandibile, per visualizzare trend mensili o settimanali.

Iptables : Fornisce il supporto per il cariacmento dei dati livello 3 dello stack TCP/IP, ovvero quelli provenienti da Netfilter:

1. Data e ora,
2. IP dell'host,
3. IP del Server oggetto della comunicazione,
4. le porte (sorgente e destinazione),
5. il protocollo,
6. indica se il pacchetto è un tentativo di login oppure una normale comunicazione.

Squid : questa tabella, come per Iptrace, fornisce il supporto per il caricamento dei dati, mediante l'utilizzo di queste due tabelle è possibile ottenere informazioni che, utilizzando un solo strumento alla volta, sarebbe impossibile ottenere. Le colonne utilizzate per Squid sono :

1. Data e ora,

2. ip server e host,

3. URL di destinazione,

4. oggetto e tipo di comunicazione

5. il codice HTTP risultante dalla richiesta.

5.3 Documentazione codice

Seguirà una descrizione del codice sviluppato, per brevità verranno affrontati unicamente i punti cruciali, e le funzioni principali di ogni modulo,

5.3.1 Struttura del programma *log_mining*

Riportiamo in seguito la struttura gerarchica ad albero, che comprende file e cartelle costituenti l'applicativo. Il programma è composto da diversi file contenenti procedure e definizioni di funzioni.

```
/var/www/log_mining/
     |
     |------------- adodb5/
     |                  |
     |                  |-------- LIBRERIA ADO DB
     |
     |
     |------------- include/
     |                   |
     |                   |-------- connect.php
     |                   |
     |                   |-------- constants.php
     |                   |
     |                   |-------- database.php
     |                   |
     |                   |-------- form.php
     |                   |
     |                   |-------- mailer.php
     |                   |
     |                   |-------- session.php
```

```
|
|
|------------- page_files
|                    |
|                    |-------- style.css
|                    |
|                    |-------- *.GIF/*.JPG - GRAFICA
|
|
|------------- phpgraphlib_v2.02/
|                     |
|                     |-------- phpgraphlib.php
|                     |
|                     |-------- phpgraphlib_pie.php
|
|
|
|------ anomalous_lines.pl
|
|------ clean.pl
|
|------ configuration.php
|
|------ coovachilli.php
|
|------ coovachilli_logins.php
|
|------ database_functions.php
|
|------ date_convert.php
|
|------ dbquery.php
|
|------ get_interval.php
|
|------ index.php
|
```

```
|------ ini_write.php
|
|------ input_validator.php
|
|------ interval.sh
|
|------ iptables.php
|
|------ iptables_logins.png
|
|------ iptables_parse.pl
|
|------ iptrace.php
|
|------ log2db.php
|
|------ log_find.sh
|
|------ log_zip.sh
|
|------ logged_users.pl
|
|------ login.php
|
|------ logloader.php
|
|------ mask1.php
|
|------ menu.php
|
|------ playtime.php
|
|------ process.php
|
|------ settings.ini
|
|------ squid.php
```

Gli script .pl e .sh servono ad eseguire operazioni in maniera ottimizzata mentre il file settings.ini contiene tutte le configurazioni riguardanti ogni modulo che compone la piattaforma di analyzing. I file .php invece sono le pagine vere e proprie o parti di esse, la libreria grafica è utilizzata unicamente per il modulo coovachilli.php, per disegnare il grafico indicante il numero di login per ora.

5.3.2 Login

La pagina relativa al login è semplicemente un modulo che permette l'inserimento delle credenziali per utilizzare la piattaforma. Utilizza le sessioni, che sono parte delle funzionalità PHP, inoltre, parte di questo modulo (/include/session.php) è utilizzata in tutti gli altri, per verificare l'effettiva validità della sessione corrente; è stato implementato anche il salvataggio di ogni accesso eseguito, per soddisfare un'altra cogenza di legge, ovvero l'auditing degli amministratori (argomento che verrà trattato a fondo nella seconda parte del presente documento).

5.3.3 Logloader

Il codice di questo modulo è la base di partenza per gli altri simili, in quanto effettua l'operazione cruciale della piattaforma, ovvero il caricamento di log su database a partire da dati in formato testuale.

Questo modulo si appoggia, come gli altri, su mysql, la struttura della pagina consente di inserire il path del log che si desidera analizzare, il numero di colonne, i loro nomi e quali di esse sono da utilizzare come indici, tutti questi dati vengono utilizzati per creare la tabella ad-hoc all'interno del Database Log_Mining, infine è presente il campo per inserire la regular expression con la quale analizzare il log ed estrarne i dati da inserire nella tabella appena creata.

Il modulo 'portante' della piattaforma risulta essere *log2db.php*, al quale è affidato l'effetivo compito di estrarre dati utili dai logs e di inseririli nella tabella opportuna, è in grado di gestire un singolo file oppure array di files, inoltre qualunque sia il numero di files in ingresso, gestisce sia file in formato RAW (non compresso), che file compressi con gzip, ottenendo prestazioni molto simili in entrambi i casi.

Il file contiene due funzioni, la prima riceve un file unico, ed è utilizzata nel modulo Log Loader, la seconda riceve in ingresso la tabella utilizzata, il pattern

che le righe devono rispettare, le colonne da riempire e l'array dei path dei files di interesse (la prima funzione differisce unicamente da questo ultimo parametro, che indica un unico file). Non appena il log verrà caricato, sarà possibile eseguire Query sui dati ed effettuare l'analisi vera e propria del log.

Nota Bene : il numero di campi inseriti nella regular expression deve coincidere con il numero di campi utiilzzati nella tabella, altrimenti verrà restituito un errore, la INSERT non potrà avere luogo.

5.3.4 Iptrace

Questo modulo è necessario per evadere rapidamente richieste provenienti dalla Polizia Postale in merito alla legge anti-terrorismo, i Casi d'uso sono stati descritti all'inizio del capitolo, le richieste forniscono un elenco di logfiles che vengono passati al sopra citato 'log2db.php', l'elenco viene creato da una funzione Perl che, dati in ingresso data/ora iniziale e data/ora finale, preleva dai logfiles unicamente le righe comprese in questo intervallo e le scrive si di un files temporaneo creato appositamente, restituendo un array conenente i path dei files prodotti.

La visualizzazione avviene analizzando i risultati delle select effettuate, una volta caricati i dati sul DB è possibile analizzarli rapidamente, conviene utilizzare questa pagina minimizzando il numero di caricamenti su DB, essendo l'operazione più dispendiosa. Una volta che l'elenco viene visualizzato compare un bottone per ogni linea, questo invia i dati necessari al modulo Squid, che li utilizzerà per estrarre i dettagli della richiesta corrispondenti ala linea presente su Iptables.

5.3.5 Squid

La pagina relativa ai log di Squid segue la struttura di Iptrace, un campo iniziale permette di specificare il range temporale, vengono poi selezionati i log corrispondenti e lo script perl ne estrae le linee utili.

I dati possono provenire anche dal modulo Iptrace, indicando l'istante e l'IP di provenienza del pacchetto del quale si desidera conoscere ulteriori informazioni, il modulo Squid in questo caso preleva non solo la richiesta specificata ma tutte le righe corrisopondenti a quell'utente nei 5 minuti successivi (lasso temporale modificabile dal modulo Configuration).

Una volta caricati i dati sulla tabella relativa a squid, è possibile specificare parti di URL da ricercare, Indirizzi IP di Host e di server, e costruire automaticamente l'interrogazione che verrà inviata al DBMS, questo semplifica l'utilizzo del sistema, che altrimenti risulta essere poco 'user friendly',ma è possibile unicamente ove la struttura della tabella e dei log sia nota a priori, sarebbe in realtà possibile creare la stessa funzionalità anche sul modulo Log Loader, ma i tempi di consegna non hanno permesso questa miglioria (essendo un Requisito Facoltativo non era di vitale importanza).

5.3.6 Coovachilli

Questa pagina è stata creata per permettere agli amministratori di visualizzare in maniera grafica l'utilizzo effettivo del servizio, individuando così i picchi di utilizzo e gli errori di autenticazione avvenuti nell'ultima ora. Il grafico si aggiorna dinamicamente grazie ad un cronjob che rileva i login del server Radius di PadovaWifi e li aggiorna una volta all'ora sul Database utilizzato dalla piattaforma.

Segue lo script orario e parte dello script che, ogni mezzanotte, controlla se nella tabella sono presenti entry che hanno superato l'anno di presenza, diventando quindi superflue, in tal caso la query più vecchia viene rimossa.

Blocco di codice 5.4

```
set -- $(perl /var/www/log_mining/logged_users.pl
***.**.**.** | awk -F"," '{print $1,$2}')
mysql --user=root --password=****** <<!!
update log_mining.logins set logs`date +%H` = '$1',
all_day = '$2' WHERE date = ``date +%Y-%m-%d``;
quit
!!
```

Blocco di codice 5.5

```
mysql --user=root --password=****** <<!!
DELETE FROM log_mining.logins WHERE date <
``date --date='365 days ago' +%Y-%m-%d``;
quit
!!
```

Questo script ricerca ed elimina ogni entry che supera l'anno di anzianità, evitando così di creare una tabella sovraccarica di dati.

All'interno della pagina sono state utilizzate librerie Adobe per la creazione dinamica di grafici in PHP, questa librerie si appoggiano ai dati contenuti nella tabella logins, che, per costruire correttamente il grafico, è stata strutturata utilizzando un campo per ogni ora del giorno, inoltre sono state utilzzate le librerie Php Graph Lib (http://www.ebrueggeman.com/phpgraphlib/). Le righe anomale vengono caricate grazie allo script perl anomalous_lines.pl ed indicano eventuali comportamenti non corretti da parte del server, questi possono essere: richieste troppo grandi, tentativi di accesso non autorizzati ecc..

5.3.7 Configuration

La pagina di configurazione è utile per modificare le impostazioni del programma dal web, le modifiche saranno necessarie in vista di nuove installazioni e modifiche particolari. Solitamente sono presenti molti file di configurazione testuali, in questo caso è stata scelta l'opzione di un file unico 'settings.ini' contenente varie sezioni. Questa struttura facilita la gestione dei file in sè e riduce la possibilità di errori in fase di backup, ogni sezione corrisponde ad un modulo, e all'interno ci saranno tutte le informazioni utili come path, script, nome della tabella e colonne della tabella mysql cui si appoggia il modulo ecc.., per cui ci saranno le sezioni :

- Database
- Iptables
- Coovachilli
- Squid
- Hosts
- Iptrace
- LogLoader

In caso di aggiunta di nuovi moduli è sufficiente aggiungere sezioni anzichè nuovi files di configurazione.

Capitolo 6

Studio problema dell'Auditing

In questo capitolo verrà introdotta la problematica dell'auditing, verrà riportata la normativa del Garante della Privacy, successivamente verranno definiti i requisiti , estrapolandoli della legge e seguendo le richieste espresse dall'area tecnica, e ne verrà definito uno Use Case Diagram.

Infine verrà analizzata una proposta di mercato, da questa offerta è possibile estrapolare informazioni utili circa lo sviluppo della piattaforma, idee e soluzioni open-source.
Infine è stato fatto un Gannt preventivo, questo diagramma servirà come traccia, da seguire in fase di progettazione, in quanto vengono definiti i principali 'palettì da rispettare.

6.1 Disposizioni del Garante della Privacy

Misure e accorgimenti prescritti ai titolari dei trattamenti effettuati con strumenti elettronici relativamente alle attribuzioni delle funzioni di amministratore di sistema, 27 novembre 2008 Gli 'amministratori di sistema' sono figure essenziali per la sicurezza delle banche dati e la corretta gestione delle reti telematiche. Sono esperti chiamati a svolgere delicate funzioni che comportano la concreta capacità di accedere a tutti i dati che transitano sulle reti aziendali ed istituzionali. Ad essi viene affidato spesso anche il compito di vigilare sul corretto utilizzo dei sistemi informatici di un'azienda o di una pubblica amministrazione.

Per questo il Garante ha deciso di richiamare l'attenzione di enti, amministrazioni, società private sulla figura professionale dell'amministratore di sistema e ha prescritto l'adozione di specifiche misure tecniche ed organizzative che agevolino la verifica sulla sua attività da parte di chi ha la titolarità delle banche dati e dei

sistemi informatici.

Le ispezioni effettuate in questi anni dall'Autorità hanno permesso di mettere in luce in diversi casi una scarsa consapevolezza da parte di organizzazioni grandi e piccole del ruolo svolto dagli amministratori di sistema. I gravi casi verificatisi negli ultimi anni hanno evidenziato una preoccupante sottovalutazione dei rischi che possono derivare quando l'attività di questi esperti sia svolta senza il necessario controllo.

Le misure e le cautele dovranno essere messe in atto entro quattro mesi da parte di tutte le aziende private e da tutti i soggetti pubblici, compresi gli uffici giudiziari, le forze di polizia, i servizi di sicurezza. Sono esclusi invece i trattamenti di dati, sia in ambito pubblico che privato, effettuati a fini amministrativo contabile, che pongono minori rischi per gli interessati.

6.1.1 Definizione amministratore di sistema

Con la definizione di 'amministratore di sistema' si individuano generalmente, in ambito informatico, figure professionali finalizzate alla gestione e alla manutenzione di un impianto di elaborazione o di sue componenti. Ai fini del presente provvedimento vengono però considerate tali anche altre figure equiparabili dal punto di vista dei rischi relativi alla protezione dei dati, quali gli amministratori di basi di dati, gli amministratori di reti e di apparati di sicurezza e gli amministratori di sistemi software complessi.

Gli amministratori di sistema così ampiamente individuati, pur non essendo preposti ordinariamente a operazioni che implicano una comprensione del dominio applicativo (significato dei dati, formato delle rappresentazioni e semantica delle funzioni), nelle loro consuete attività sono, in molti casi, concretamente 'responsabili' di specifiche fasi lavorative che possono comportare elevate criticità rispetto alla protezione dei dati.

Attività tecniche quali il salvataggio dei dati (backup/recovery), l'organizzazione dei flussi di rete, la gestione dei supporti di memorizzazione e la manutenzione hardware comportano infatti, in molti casi, un'effettiva capacità di azione su informazioni che va considerata a tutti gli effetti alla stregua di un trattamento di dati personali; ciò, anche quando l'amministratore non consulti 'in chiaro' le informazioni medesime.

6.1.2 Designazioni individuali

La designazione quale amministratore di sistema deve essere in ogni caso individuale e recare l'elencazione analitica degli ambiti di operatività consentiti in base al profilo di autorizzazione assegnato.

6.1.3 Elenco degli amministratori di sistema

Gli estremi identificativi delle persone fisiche amministratori di sistema, con l'elenco delle funzioni ad essi attribuite, devono essere riportati nel documento programmatico sulla sicurezza, oppure, nei casi in cui il titolare non è tenuto a redigerlo, annotati comunque in un documento interno da mantenere aggiornato e disponibile in caso di accertamenti anche da parte del Garante.

6.1.4 Verifica delle attività

L'operato degli amministratori di sistema deve essere oggetto, con cadenza almeno annuale, di un'attività di verifica da parte dei titolari del trattamento, in modo da controllare la sua rispondenza alle misure organizzative, tecniche e di sicurezza rispetto ai trattamenti dei dati personali previste dalle norme vigenti.

6.1.5 Registrazione degli accessi

Devono essere adottati sistemi idonei alla registrazione degli accessi logici (autenticazione informatica) ai sistemi di elaborazione e agli archivi elettronici da parte degli amministratori di sistema. Le registrazioni (access log) devono avere caratteristiche di completezza, inalterabilità e possibilità di verifica della loro integrità adeguate al raggiungimento dello scopo di verifica per cui sono richieste. Le registrazioni devono comprendere i riferimenti temporali e la descrizione dell'evento che le ha generate e devono essere conservate per un congruo periodo, non inferiore a sei mesi.

6.2 Requisiti

Verranno ora definiti i requisiti rilevati durante l'analisi della legge fatta da me e da eltri componenti dell'area tecnica, essendo la legge non molto chiara in alcuni punti è stato necessario definire chiaramente il compito che l'auditing andrà a svolgere.

6.2.1 Obbligatori

1. Ogni account potrà essere utilizzato da una persona soltanto, non si potranno usare account comuni riferiti a macchine invece che a persone,
2. Definire un elenco degli amministratori di sistema,
3. Memorizzare ogni operazione di accesso (con credenziali di amministratore) riguardante gli utenti appartenenti all'elenco sopra definito,
4. Memorizzazione informazioni su login, logout e tentativi di accesso falliti riguardante ogni account appartenente alle persone dell'elenco sopra definito, questi dati riguardano anche accessi in remoto ed in tal caso va rintracciato l'IP della macchina dalla quale proviene la richiesta,
5. Creare un sistema di raccolta dati, con la possibilità di verifica della loro integrità, rendere i dati inalterabili,
6. Mantenimento dei log per almeno 6 mesi,
7. Almeno una volta l'anno vanno verificati i dati raccolti per rilevare eventuali incongruenze o irregolarità, la verifica va fatta da parte del titolare dell'azienda
8. Non creare falle di sicurezza o possibilità di intrusione all'interno della rete interna,
9. Il sistema deve essere monitorabile (Nagios)
10. Il sistema deve inviare mail giornaliere per avvertire di essere ancora in piedi.

6.2.2 Facoltativi

1. Trasparenza del sistema
2. Non creare su ogni server un elenco di username associate alle persone, ciò creerebbe difficoltà in caso di cambio del personale,
3. Bassa incidenza sulle prestazioni delle macchine,
4. Semplificare la gestione degli account

6.3 Usa Case

Per creare il sistema di analisi sono stati creati vari Use Case, quelli più significativi indicano le verifiche che il titolare dell'azienda è tenuto ad effettuare, essendo quelli strettamente necessari, il sistema può essere anche utilzzato dagli stessi amministratori di sistema, per rilevare intrusioni da parte di estranei.

La casistica mostrata si riferisce sempre ad eventi avvenuti entro un determinato lasso temporale, specificabile tramite l'interfaccia che verrà messa a disposizione, le casistiche che ho identificato come più rilevanti sono:

CASE1 : il titolare dell'azienda ha intenzione di venire a conoscenza del totale dei login effettuati su tutto il sistema da ognuno degli amministratori di sistema, di una parte di esso, o di una macchina soltanto, il sistema quindi, ricevuti in ingresso i dati utili, dovrà estrarre i logs e mostrarli, dando la possibilità di effettuare conteggi in base a login, macchina utilizzata o altri parametri

CASE2 : il titolare richiede la verifica dell'integrità dei dati, se questa verifica non avesse successo deve essere possibile reperire dati integri con i quali sostituire quelli corrotti. Il sistema deve quindi implementare la possibilità di verificare la correttezza sia dei logs che dei file md5 contenuti nel server di storage.

CASE3 : La verifica delle attività degli amministratori, che deve avvenire con cadenza almeno annuale, deve fornire dati circa la quantità degli accessi e gli orari in cui essi avvengono, questa operazione deve essere possibile e facilmente eseguibile. Detto ciò, la piattaforma deve implementare controlli standard da effettuare regolarmente, come: verifica del totale degli accessi per ogni amministratore, eventuali grafici che rappresentino l'utilizzo delle macchine in retee conteggio degli accessi possiblimente raggruppati per utente/macchina.

CASE4 : gli amministratori verificano la regolarità degli accessi avvenuti nelle macchine che gestiscono, tentando di scoprire eventuali accessi non autorizzati e prendendo i giusti provvedimenti in merito . La piattaforma deve quindi implementare anche un sistema di filtraggio dei risultati più complesso, per 'power user', per offrire agli amministratori la possibilità di effettuare controlli a piacimento.

Si nota che, se serve avere a disposizione questi dati non sarà sufficiente creare un lettore di logs, ma andrà configurato tutto il sistema di centralizzazione e gestione dei dati.

6.4 Tempistiche previste

Dopo una ricerca di mercato su costi/tempi di altri produttori di software con lo stesso obiettivo, ho preventivato un costo ideale del sistema nel suo complesso basandomi sul tempo che avrei impiegato. Ponendo come limite temporale il 15 Dicembre, e tenendo conto che disponevo di una base per la lettura dei log raccolti, è stato possibile creare in breve tempo un progetto di massima del risultato finale e del tempo che avrei impiegato all'implementazione dell'Analizzatore dei log.

Il tempo previsto per il sistema di raccolta è stato più di ardua valutazione in quanto, all'inizio del progetto, non disponevo delle competenze necessarie alla realizzazione immediata di parte degli stessi progetti che ho proposto, quindi è stato incluso un periodo di apprendimento ed addestramento, la previsione di questo periodo è stata arbitraria e, a metà progetto, rivelatasi errata.

Per la crezione di una pagina dedicata alla lettura sarebbe stato sufficiente prelevare pezzi di codice proveniente dal log mining e ricombinarli con le opportune aggiunte, quindi ho previsto circa 8 giorni di lavoro, saranno da implementare anche una verifica ed un sistema di estrazione delle username dai logs.

La configurazione della raccolta, essendo già stata impostata nel caso del Log Mining, appariva una situazione simile, quindi feci una valutazione, che, al momento, ritenevo pessimistica, e si aggirava attorno alle 3 settimane, in realtà la configurazione si è protratta per quasi un mese, il consuntivo iniziale prevedeva quindi 150 ore circa di lavoro, in realtà ce ne sono volute quasi il doppio includendo il lavoro di perfezionamento successivo al termine del progetto, la migrazione che ho richiesto ed il ripristino di un server down, da dedicare al progetto.

La data dalla quale sono stato impiegato alla risoluzione del problema è stata il 06/11/2009 dopo circa due settimane (23/11/2009) ho fatto notare all'area tecnica che i risultati sperati non arrivavano, è stata quindi convocata una riunione, durante la quale è emersa la necessità di un cambio rotta verso la soluzione più semplice e veloce, con la quale sarei riuscito ad ottenere i risultati richiesti entro i tempi previsti dalla legge.

Come si può notare dal Gannt in figura 6.1, la costruzione effettiva del progetto ha impiegato effettivamente più tempo del previsto.

Il termine ultimo previsto dalla legge era il 15/12/2009, in realtà il lavoro è proseguito ben oltre tale data, ma la milestone principale era il completamento della raccolta entro tale data.

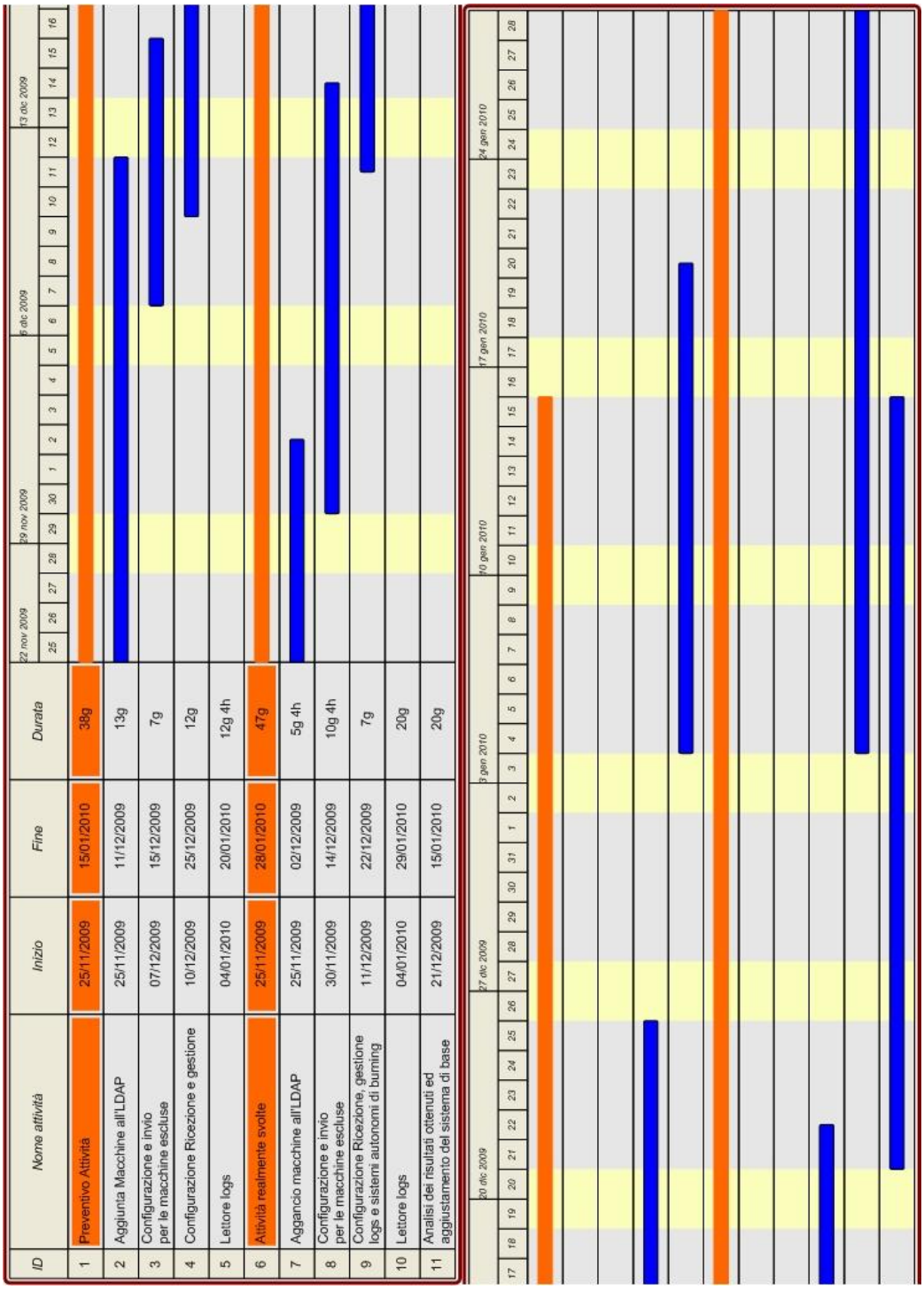


Figura 6.1: Gannt previsto ed effettivo per la realizzazione dell'auditing in azienda

6.5 Soluzioni di mercato

Molti prodotti fungevano unicamente da analizzatori, lasciando all'utilizzatore il compito della raccolta e della centralizzazione dei log, l'unico prodotto valido come rapporto qualità/prezzo è LegalLogger, che ora andremo a descrivere brevemente.

6.5.1 LegalLogger

Legal logger è nu'appliance che si propone di risolvere il problema che il provvedimento del Garante della privacy ha posto, ovvero la raccolta e l'analisi dei logs relativi ai login effetuati degli amministratori di sistemi. I costi da sostenere in caso di acquisto sarebbero:

- Costo attivazione = 1400
- Canone Mensile = 100 euro

Entità Coinvolte :

- Ca Infocamere per il rilascio della Marca Temporale
- Server Centrale con funzione di CA per il rilascio di certificati verso i singoli appliance
- Server centrale con funzione di archivio Informatico con procedure di backup
- Appliance presso il cliente - Agente installato sui server/Client del cliente

Descrizione del processo :

- L'agente installato sul sistema controllato inoltra gli eventi verso l'appliance di logging
- L'appliance di logging conserva i log all'interno di un database strutturato
- Ogni 24 ore l'appliance esporta i file di log in formato XML, li comprime in formato .bz2 e li firma localmente con certificato rilasciato dalla CA centrale.
- Una volta firmati invia il file sulla repository del server centrale attraverso canale sicuro SL

- Il server centrale colleziona i log in una directory differente per ogni cliente e vi appone una marca temporale rilasciata da Infocamere
- Il server centrale mantiene in archivio i file di log firmati e marcati per un periodo che va da un minimo di mesi 6 (previsto dalla legge) ad un massimo di anni 2 se sottoscritto dal cliente.
- Il server centrale permette al cliente di scaricare/visualizzare i file di log mantenuti in archivio.

Appliance L'appliance è costituito da dispositivo Embedded di adeguate capacità computazionali con disco flash da 4GB (di cui 256 MB occupati dal sistema). Sull'appliace sono installati le seguenti applicazioni:

1. Mysql
2. Syslog-ng
3. Openssl
4. Openssh
5. Pppd
6. Apache

Programmi/Procedure da noi prodotte per la gestione dei log

Server Il server installato e residente presso due Server Farm in cluster geografico con accesso loggato e controllato eroga il servizio di :

1. C.A. (Certification Autority)
2. SSL Vpn (RSA 2048 bit)
3. Archiviazione su file system Criptato DSA 1024
4. Browsing Web dei Log Archiviati (con accesso tramite certificato X.509 personale)
5. Marcatura temporale tramite interfacciamento con server Infocamere (viene rilasciata una marca temporale per ogni file di log).
6. Backup di sicurezza dei log dei clienti

Da questa soluzione sono state prelevate alcune idee, che verranno valutate nel capitolo che segue, ovvero lo studio di fattibilità.

Capitolo 7

Proposte Auditing

Come ogni progettista alle prime armi, l'approcio iniziale al problema, era di tentare il classico errore del: 'tutto e subito', ma analizzando le varie problematiche che avrei potuto avere e i rischi da gestire durante lo sviluppo del progetto, ho scelto di utilizzare le nozioni apprese dal corso di Ingegneria del Software per una corretta e sicura gestione del progetto.

Come prima cosa ho analizzato le fasi operative che avrebbero contraddistinto il mio lavoro, successivamente ho effettuato uno studio dei requisiti, sia quelli imposti dalla legge, che sono stati indicati come Obbligatori, che le possibili evoluzioni del sistema interno per garantire maggiore controllo sui sistemi presenti in azienda, una volta raccolti i dati ho creato degli Use Case Diagram circa l'utilizzo della piattaforma finale.

Prima di procedere con lo studio di possibili soluzioni per il soddisfacimento dei requisiti, definiti in base alla legge ed in base alle necessità dell'area tecnica, è stato creato un quadro completo della situazione interna per quanto riguarda le autenticazioni e le gerarchie eventualmente presenti tra le macchine all'interno del dominio.

Per terminare lo studio della prolematica, ho valutato con l'area tecnica la fattibilità di varie proposte, sia strutturali che per la decisione delle utility più adatte per raggiungere gli obiettivi prefissati, definendo infine il sistema più adatto per procedere con lo sviluppo del progetto, ovvero il Modello Evolutivo.

7.1 Studio delle fasi

Le fasi operative, evidenziate dalla pianificazione del lavoro, sono le seguenti:

- descrizione operativa preliminare del lavoro (Capitolo precedente)

- pianificazione: definizione delle fasi di lavoro e delle milestone
- pianificazione: diagramma di Gantt, cartella WBS
- problematiche presenti
- analisi esigenze e necessità (espresse in base alle problematiche presenti)
- individuazione e definizione del problema o dei problemi da risolvere
- stato attuale dell'arte - strumenti, metodologie, servizi offerti, sw, hw (Appendice)
- finalità, obiettivi, specifiche di progetto
- analisi delle competenze, conoscenze, capacità necessarie (know-how)
- vincoli: risorse umane, economiche, tecniche, logistiche, temporali a disposizione
- rischi: analisi e gestione
- soluzioni possibili
- valutazione e scelta della soluzione
- implementazione e messa in opera della soluzione scelta
- testing
- report finale

7.2 Premessa

Seguono alcune osservazioni e le principali considerazioni sorte durante lo studio di fattibilità del progetto.

7.2.1 Pianificazione del lavoro

Nei primi momenti di questo progetto, mi sono avvalso di conoscenze di Project Management al fine di pianificare al meglio le fasi successive.

Purtroppo la mancanza di competenze tecniche e professionali, unita alla scarsa conoscenza delle problematiche legate al progetto in questione, non mi hanno permesso di realizzare un planning preliminare soddisfacente. Per questo motivo, la pianificazione del lavoro si è protratta nel tempo per tutta la durata del progetto; nei mesi successivi ho rivisto e rielaborato progressivamente la pianificazione precedentemente operata, riuscendo nell'ultimo periodo ad avere una visione globale del progetto, nelle sue componenti così come nella sua interezza, la sezione dell'LDAP è stata quella più problematica in quanto, inizialmente ho perso molto tempo per trovare una metodologia per agganciare macchine Linux ad un LDAP Windows, successivamente mi sono ritrovato a dover applicare tale sistema su macchine poco aggiornate e problematiche

7.2.2 Risorse a disposizione

Per quanto riguarda il sottoscritto, le competenze a mia disposizione all'inizio dello stage erano di natura strettamente teorica; non avendo mai operato a livello pratico in ambito tecnico-professionale, le conoscenze che possedevo derivavano solo ed esclusivamente dalla formazione universitaria. Ad ogni modo, in azienda ho trovato numerose risorse umane in possesso di conoscenze approfondite e variegate sulla realtà dell'area tecnica aziendale. L'inconveniente è stato dettato dal fatto che il *know how* è distribuito su numerosi membri del personale, ciascuno molto preparato su alcuni aspetti ma non formato su altri: le difficoltà che ho incontrato nel reperire le informazioni necessarie non sono state poche nè banali. Fortunatamente, buona parte dei dipendenti si è rivelata disponibile (oltre che preparata e professionale) a condividere con me le conoscenze di cui avevo bisogno. Per quanto riguarda le risorse umane (conoscenze) e tecniche (*workstations*) ho trovato il più delle volte, seppur non sempre in tempi brevi, ciò che mi occorreva.

7.2.3 Vincoli di progetto

I vincoli sono dettati unicamente dai requisisti e dalle scadenza imposte dalla legge, in termini economici l'obiettivo era il 'costo zero', quindi :

- Temporalmente il limite ultimo di conclusione del progetto era il 15 Dicembre 2010, se non altro per la sezione riguardante la raccolta, che perciò ha assunto priorità su tutti gli altri task in corso.
- I vincoli economici, come è ragionevole supporre, sono determinati dall'attitudine delle PMI a spendere il meno possibile, o meglio, a spendere solo laddove sia necessario. Non avendo un budget a disposizione per l'acquisto di applicativi software o hardware, ho trovato opportuno pormi il limite di utilizzare solamente software open source e freeware, oltre ad applicazioni da me sviluppate ad hoc in base alle necessità riscontrate.

7.2.4 Rischi

I rischi sono numerosi visto che il progetto è molto esteso e complesso, per cui bisogna cercare di prevedere e gestire i casi in cui un rischio si stia avverando o si sia effettivamente avverato.

Nella seguente tabella vengono elencati i rischi, alcuni preventivati mentre altri imprevisti, che si sono poi verificati durante il progetto e per i quali si sono dovute prendere delle misure allo scopo di mitigarne gli effetti negativi sul buon esito del progetto.

ID	*Rischio*	*Probabilità*	*Conseguenze*
1	Difficoltà gestione progetto	40%	2
2	Allungamento tempi acquisizione delle informazioni	30%	3
3	Scarsità documentazione	30%	2
4	Bassa qualità documentazione	20%	4
5	Problemi configurazione	40%	2
6	Allungamento tempi realizzazione	50%	2
7	Allungamento tempi configurazione	30%	2
8	Allungamento fase di test	30%	4
9	Problemi relativi alla DMZ	20%	3
10	Problemi causati dall'LDAP	15%	2
11	Problemi di sicurezza	20%	2
12	Variazione esigenze	30%	2
13	Interruzione di servizi 24*365	5%	1

Tabella 7.1: Tabella dei rischi

Legenda:

```
Conseguenze
1 : dannose
2 : critiche
3 : lievi
4 : non rilevanti
```

7.2.5 Report

La stesura della documentazione prodotta è stata fatta mano a mano che i progetti procedevano: si tratta di studi di fattibilità, approfondimenti su particolari problematiche, strumenti e metodologie, schemi e rappresentazioni grafiche, documentazioni su software da me sviluppato. Buona parte di questi documenti sono poi serviti per comporre il report finale e la tesi qui presente, dopo opportune rielaborazioni.

L'analisi dei documenti prodotti nel tempo ha aiutato il personale tecnico a fornirmi suggerimenti ed informazioni in merito al progetto ed alla rete interna.

7.2.6 Soluzioni proposte e soluzioni implementate

Per le problematiche analizzate sono state considerate delle possibili soluzioni, presentate nell'arco della tesi, che saranno descritte in maniera più o meno approfondita a seconda del grado di interesse della soluzione stessa. Nei casi in cui una soluzione sia stata scelta in quanto riconosciuta più adatta alla situazione rispetto alle altre, questa verrà descritta con un maggior livello di dettaglio. Alcune soluzioni sono state implementate e ne è stato fatto il deployment; la messa in opera è stata effettuata nei limiti di fattibilità, in base ai vincoli presenti e alle risorse disponibili.

La fase di test delle soluzioni implementate è avvenuta sia costantemente durante la stesura del codice sia in toto al termine della fase di sviluppo.

7.3 Analisi della situazione attuale

Prima di procedere con una ricerca di soluzione del problema è stato necessario efettuare un'analisi della situazione aziedale attuale, andando a reperire informazioni sulle macchine e sul personale utili alla messa in regola rispetto ai requisiti individuati.

L'elenco degli amministratori di sistema verrà omesso in quanto non pertinente con la presente tesi, verranno invece presetate le informazioni reperite sul

comportamento delle macchine e sui sistemi di autenticazione, essendo, grazie ad essi, possibile reperire tutte le informazioni richieste.

Oltre alle autenticazioni è stato utile conoscere la struttura della rete interna e delle reti sotto il la gestione degli amministratori che ho avuto il compito di mettere sotto auditing.

Processi coinvolti I processi coinvolti nella creazione di questo sistema sono unicamente le procedure di login/logout di ogni sistema al quale un amministratore accede, quindi il sistema in sé deve rimanere inalterato, l'unica modifica è trasparente agli utilizzatori. Coinvolge invece l'assegnazione delle username e le relative password individuali, la legge impone che ogni accesso sia associabile ad una persona fisica e non ad un ruolo, ci sarà un successivo salvataggio di ogni operazione di login/logout, inviate poi su di un apposito server di raccolta.

La rete interna contiene tutte le macchine collegate all'LDAP e buona parte dei server Windows e Linux, l'altra rete contenente un buon numero di server è la DMZ, là sono presenti tutti i server che hanno la necessità di fornire servizi disponibili ai clienti da internet; le altre reti riguardano in gran parte dispositivi di rete, solo alcuni server (Radius) sono presenti nella rete Padova Wifi e Monselice Wifi.

In figura 7.1 è mostrato come la macchina che fungerà da server sarà posta fra le due reti. I logs provenienti dalle reti esterne attraverseranno inevitabilmente il Firewall, ma si è riusciti ad evitare un flusso costante di dati entranti verso la rete interna. Saranno ora elencati i software di ambiete che sono stati rilevati essere in uso:

1. Windows XP
2. Windows Server 2000 (Standard e Advanced edition)
3. Windows Server 2003 (Standard e Advanced edition)
4. Linux - Gentoo (In molte versioni differenti)
5. Linux - Ubuntu (Tendenzialmente dalla 6 in avanti)
6. Linux - RedHat
7. Linux - Suse

I software di base (DBMS) utilizzati invece sono i seguenti:

1. Oracle

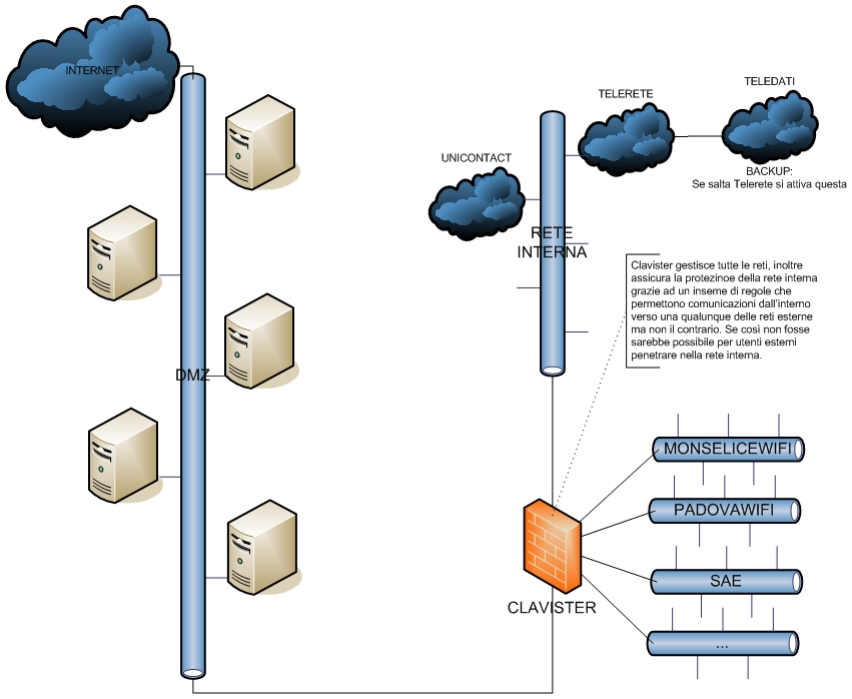


Figura 7.1: Struttura della rete aziendale

2. Mysql
3. Postgres
4. SQLServer
5. Informix - Unicamente per il Tram

7.3.1 LDAP

L'autenticazione per le macchine Windows è centralizzata, ogni macchina all'accesso riceve l'autenticazione dall'LDAP (Windows Server 2000), sono presenti vari LDAP all'interno dell'azienda e questi tra di loro sono visibili ma non accessibili (Telerete vede Unicontact ma non vi può accedere). Lo schema che , al momento dell'inizio dei lavori, definiva il dominio aziendale è rappresentato in figura 7.2, si nota la separazione tra vari domini e la presenza di macchina isolate, inoltre nella DMZ non è presente alcuna struttura 'portante'. Queste caratteristiche mi hanno spinto a proporre soluzioni migliorative in questa strada, quali l'unificazione degli accessi e la creazione di un PDC posto a livello superiore rispetto a quelli già presenti, che permettesse agli amministratori di Rete di controllare i domini sottostanti con maggiore livello di astrazione. Tutte le macchine collegate

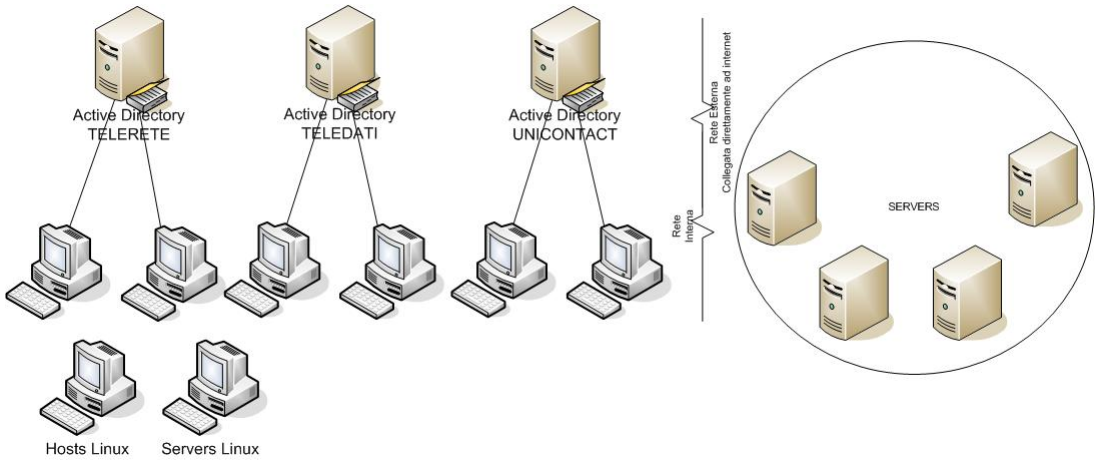


Figura 7.2: Struttura dell'albero LDAP attualmente presente in azienda

all'LDAP dispongono delle utenze relative a tutti i dipendenti e degli amministratori di sistema, grazie alle peculiarità del sistema LDAP è ppossibile accedere ai propri file ed alla propria configurazone personale da qualunque macchina in azienda, essendo caricate sulla macchina che ha il ruolo di Server LDAP e di server Exchange per la posta, in particolare si tratta di un Windows Server 2000, ance se a breve è prevista una migrazione ad un Windows Server 2003.

7.3.2 Host Linux

Sono presenti pochi PC Linux all'interno della Rete Telerete, questi non effettuano l'accesso al dominio, per la precisione sono macchine utilizzate da amministratori di sistema, su queste macchine sono presenti unicamente dati personali degli amministratori o lovo lavori in corso, e non sono necessariamente da mettere sotto auditing, in ogni caso i dati sulle macchine personali restano di responsabilità propria.

7.3.3 Server Linux

In azienda sono presenti numerosi server Linux, la maggior parte delle macchine nel DataCenter, tutti questi server utilizzano il sistema di autenticazione nativo di Linux, ovvero una username ed una password salvate localmente all'interno della macchina stessa, inoltre è uso comune accedere unicamente attraverso l'utente root, non permettendo al sistema di ottemperare alle cogenze di legge, le quali richiedono di poter identificare l'identità dell'amministratore che ha effettuato l'accesso.

I dati contenuti sui server Linux sono svariati e molti di questi rientrano nei requisiti dettati dalla legge, dato il numero di server e il tipo di dati che essi contengono, questi saranno parte integrante del progetto.

7.3.4 Server Windows isolati

Molti dei server Windows, indipendentemente dalla rete in cui si trovano, sono scollegati dall'LDAP, questa scelta è stata fatta per evitare che utenti non autorizzati possano accedervi o anche solo essere a conoscenza della loro presenza, in sostanza per motivi di sicurezza, come per le macchine Linux, anche qua è uso comune utilizzare unicamente l'utente Administrator, creando difficoltà nell'identificazione della persona reale che ha effettuato l'accesso.

I dati presenti su questi server rientrano quasi sempre nell'insieme definito dalla legge, andranno quindi monitorati anche gli accessi a queste macchine.

7.3.5 Database

I database sono presenti sia sulle macchine Windows che sulle macchine Linux, in entrambi i casi le macchine coinvlote non sono incluse nell'LDAP. Attualmente l'accesso ai database è garantito dall'inserimento di user e password direttamente da console, o inviando i dati tramite connessioni remote, l'elenco di utenti è contenuto all'interno del database stesso. Per ogni databse solitamente è presente un unico utente Root, non è quindi possibile risalire al nome dell'amministratore che sta effettuando l'accesso.
Va tracciato ogni accesso ai Database.

7.3.6 Attività coinvolte

Un vicino cambio di Windows Server (da 2000 a 2003) sarà da tenere presente sia durante che dopo i lavori, oltre ad aggiungere al dominio tutti i pc della rete andranno riagganciati anche i pc linux della rete interna. Il nuovo server è già presente ed è momentaneamente situato nella DMZ, questo per evitare conflitti col Windows Server correntemente in uso.

7.4 Soluzione proposta

La soluzione proposta prevede:

- Server di raccolta centrale: Zabbix-interno, già attivo per la raccolta dei log riguardanti PadovaWiFi, verrà leggermente modificato per implementare l'inalterabilità richiesta,

- Client Windows: mantenere la struttura attuale (LDAP) attivando solamente il logging sul domain controller, cosa al momento già attiva e funzionante,

- Client Linux: configurazione di Samba che con l'ausilio di WinBind e Kerberos è in grado di agganciarsi ad un dominio preesistente (nel nostro caso al Windows Server 2000) : la procedura per effettuare questo agganciamento è già pronta sia per macchine Gentoo che Ubuntu,

- Database: Attivazione logging built-in e lettura di questo tramite syslog-ng, lo stesso syslog potrà inoltrare i log utili verso un database dove verrà effettuata la stessa operazione che sui PDC,

- Creazione di un Domain controller Samba dedicato ai server esterni (DMZ),

- Attivazione del logging sul Primary Domain Controller (PDC): effettuata e (dopo una settimana che era stata attivata) funzionante

- Configurazione di un Database per la raccolta giornaliera dei log sui Domain Controller e sulle macchine contenenti DBMS da controllare, questo DB ogni 24 ore produrrà un file che verrà firmato ed inviato al server centrale di raccolta.

Per quanto riguarda l'unificazione della rete DMZ con l'active directory locale, il pericolo è più alto dei benefici che si otterrebbero, la gestione degli account per i server riguarda unicamente gli amministratori, essendo gli unici che accedono a queste macchine. La parte DMZ della rete aziendale è fisicamente separata, e tramite Clavister lo è anche logicamente, unificarla avrebbe come beneficio semplificare gli accessi (peraltro non molto frequenti) e la gestione degli account degli amministratori, ma di contro alzerebbe il rischio del resto della rete interna, il rischio è dovuto al fatto che un utente esterno che riuscisse a bucare un server, si ritroverebbe all'interno del dominio e potrebbe accedere alla rete interna, creare un server Samba apposito è sembrato più opportuno per evitare intrusioni nella rete interna.

Nella figura 7.3 è rappresentata la struttura che verrà proposta. Si notano le due strutture separate, infatti, per motivi di sicurezza, la preferenza è andata

sull'isolamento logico della DMZ, per evitare che eventuali intrusioni siano in grado di superare il Firewall, che nell'immagine è rappresentato dalla linea posta fra le due strutture. Essendo già in corso d'opera la configurazione di un nuovo Active Directory si può sfruttare il server temporaneo per effettuare testing e possibilmente impostare l'active Directory futuro già da ora, prima di terminare la migrazione degli utenti, al momento sono stati agganciati alcuni PC linux all'Active Directory attuale, questo perchè non sono riuscito a forzare l'aggancio verso l'esterno essendo i nomi di dominio, di server e di workgroup, identici tra i due domini (WS 2000 sulla rete interna e temporaneamente WS 2003 sulla rete DMZ).

I Logfiles saranno situati unicamente sugli Active Directory, dopo la raccolta, ogni 24 ore, verranno cifrati ed inoltrati, per essere inoltrati sul server di raccolta Zabbix verrà installata un'istanza di OpenSSL e di Stunnel sull'Active Directory e su Zabbix, questo per comunicare in modo sicuro tra le due entità. I log verranno certificati prima dell'invio mediante Certificati Infocamere (durante la fase di sviluppo/testing verrà usato un certificato openssl).

7.4.1 Benefici

Questa soluzione porterebbe una centralizzazione degli account e dei domini della rete interna, l'accesso tramite account personali alle macchine linux interne e quindi una generale sicurezza interna aumentata, i log verrebbero raccolti unicamente dai server Active Directory (Windows o OpenLDAP) ed inviati al server di raccolta ogni 24 ore già firmati.

Altra centralizzazione riguarda gli accessi (anche se rari) degli amministratori ai server DMZ, aumenterebbe la sicurezza di questi ultimi in quanto, così facendo, modificare regolarmente le password sarebbe molto più semplice che agire su ogni macchina singolarmente.

7.4.2 Problemi

Mettere in comunicazione l'Active Directory della DMZ con la rete interna per la trasmissione dei log crea una falla di sicurezza, tramite la quale malintenzionati possono penetrare nella rete interna.
Capire come generare il Database 'sicuro' dove caricare i log ed ogni 24 ore firmarli ed inviarli al server di raccolta.
Opzionale: Installazione di SWAT per la configurazione da remoto (in http usando ssl) del PDC posto nella DMZ, altrimenti un eventuale 'man in the middle'

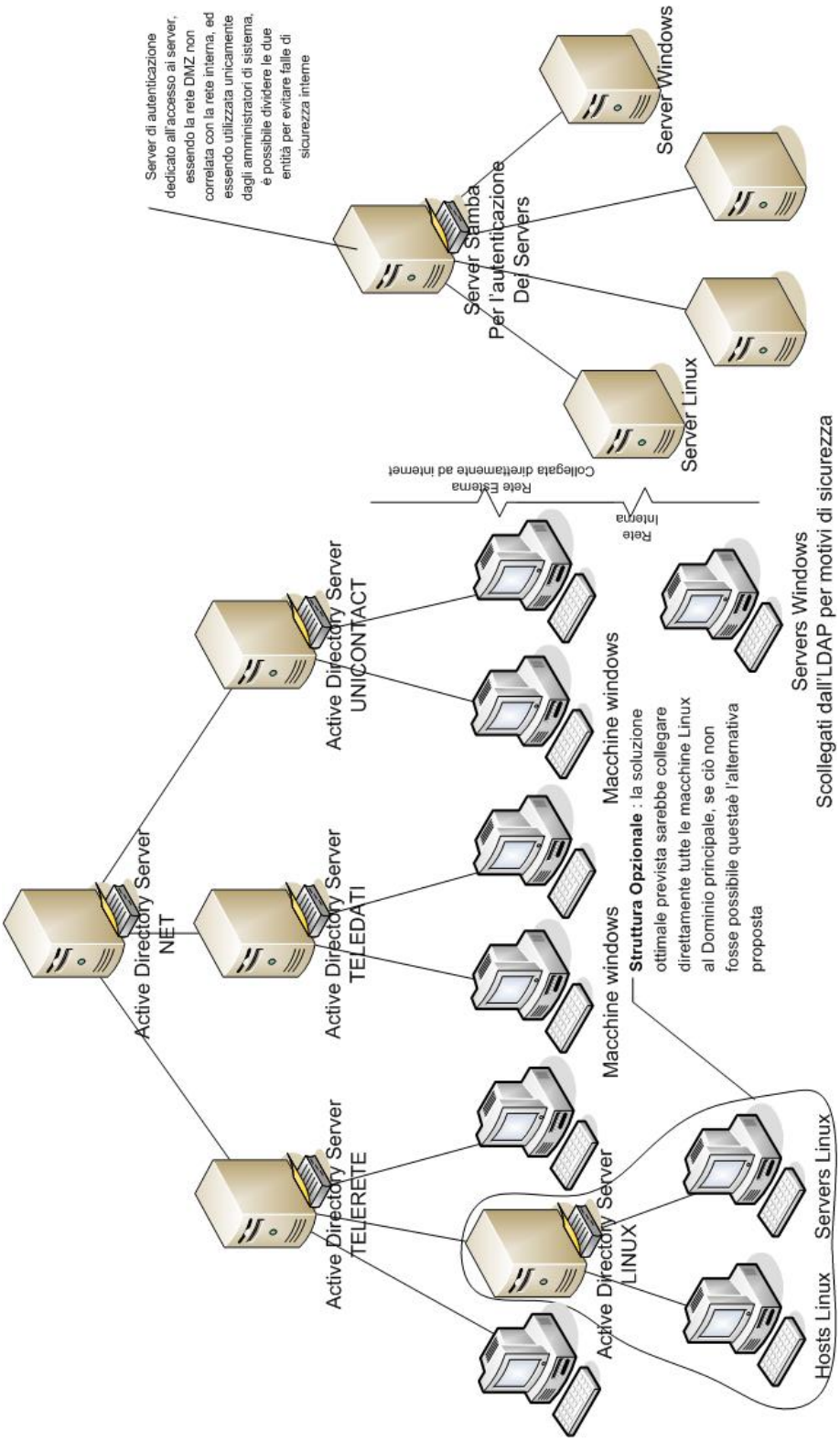


Figura 7.3: Struttura LDAP con 2 alberi distinti, Active Directory e Samba

sarebbe in grado di recuperare dati.
Comunicazione OpenLDAP-Windows Server

7.4.3 Tempi

Creazione ed impostazione dell'Open LDAP samba per la DMZ > 1 settimana, studio di Samba compreso.
Installazione di OpenLDAP(Winbind) su ogni macchina linux e configurazione > 4 ore per macchina (1/4 d'ora per macchina se non vengono riscontrati problemi riguardo installazione o aggiornamenti), ma, che vada tutto liscio, è un'eventualità molto rara.
Impostazione dell'elenco di amministratori su ogni DBMS ed impostazione Syslog per l'invio > 1/2 ora per DBMS

7.4.4 Rischi

Questa soluzione è stata proposta in quanto pone meno rischi delle altre, maggiore sicurezza interna ed esterna (DMZ), ma mantiene separate tra di loro le due reti (in verità sono presenti altre reti ma sono tenute isolate tra loro grazie al Clavister), il rischio che permane in ogni impostazione di lavori sta nella comunicazione dei server nella DMZ col server di raccolta.
Non terminare i lavori per i tempi previsti dalla legge (15 Dicembre 2009).
Due soluzioni per gestire il rischio di non riuscire a terminare il lavoro in tempo sono :

1. Soluzione semplificata
2. Acquisto soluzione esterna.

Soluzione Semplificata

Effettuare l'agganciamento su ogni macchina della rete interna seguendo la procedura che al momento non crea problemi, mentre sulla DMZ la creazione di un Server per l'autenticazione (configurare Samba) può essere problematica, questo principalmente per due motivi:

1. Impossibile verificare la fattibilità dell'allacciamento di macchine Window (presenti nella DMZ) ad un Server Samba, su varie guide trovate su internet viene detto che è sufficiente attivare una funzione di supporto a window sul server samba, ma per verificarne la veridicità andrebbe testata questa

funzionalità, quindi il test sarebbe fattibile unicamente dopo che il server samba è funzionante (4/5 giorni).

2. Il server andrebbe costruito blindato in quanto sarebbe sulla DMZ, zona a rischio intrusioni.

Al contempo questa soluzione, più semplice a prima vista, potrebbe produrre problemi per quanto riguarda la gestione dei logfiles, andrebbe implementato lo stesso meccanismo di inoltro/certificazione di log su ogni macchina invece che unicamente sul Server di autenticazione; cosa che andrebbe fatta comunque sulle macchine che contengono DBMS.
Se invece ci fosse a disposizione una macchina unicamente per l'auditing, questa potrebbe essere utilizzata per effettuare tutte le operazioni senzxa che nessuno possa accedervi,risolvendo gran parte dei problemi.
La soluzione di emergenza che è stata proposta è la seconda tra quelle che verranno descritte nella sezione sottostante (7.5.2).

Acquisto soluzione esterna: LegalLogger

Il termine ultimo per adottare una soluzione di auditing a norma di legge è il 15 Dicembre, ed il prolungamento dei lavori è probabile dati i problemi riscontrati su alcune macchine, una soluzione per contenere il problema sarebbe ricorrere a una soluzione esterna, ovvero comprare un pacchetto SW che gestisca il sistema di Auditing, il costo della versione Enterprise è di 1400 Euro per l'acquisto ed un canone di 100 euro mensili. L'offerta è stata descritta nel precedente capitolo.

7.5 Soluzioni alternative proposte

Prima di procedere sono state proposte svariate soluzioni, alcune unicamente di carattere teorico, altre invece fornite di esempi pratici, in ogni caso tutte fattibili nel contesto aziendale, alcune di queste sono state parzialmente implementate nella soluzione reale, lo studio quindi è stato utile all'identificazione di soluzioni veloci da adottare in caso di problemi nell'applicazione della soluzione principale.

I metodi di raccolta riguardano la struttura del sistema che si andrà ad implementare, le modifiche ai sistemi di autenticazione o alla struttura di LDAP presente in azienda attualmente. La raccolta relativa alle macchine interne non presenta problematiche organizzative di rilevanza, sono presenti più Windows Server, sui quali è possibile impostare l'auditing ed inviare i log al server centrale

tramite un qualunque servizio di invio log (i sistemi di trasmissione dei log verranno discussi nel seguente capitolo), le macchine Linux non collegate al dominio verranno collegate tramite OpenLDAP.

7.5.1 Tracciamento degli IP

Questa soluzione organizzativa è la meno invasiva sul sistema generale, ma al contempo non offre miglioramenti, unicamente la raccolta dei log, questa soluzione non è completamente sicura nella raccolta, essendo relativamente semplice nascondere la provenienza fisica di una richiesta di login remoto.

La soluzione consiste nell'effettuare un tracciamento degli Ip di provenienza delle richieste; dentro ogni log relativo ad accessi remoti macchine (server o host) è presente l'ip da cui proviene la richiesta, prelevando questa informazione è possibile capire da che PC proviene la richiesta, essendo le macchine all'interno dell'ufficio assegnate a determinate persone, si dovrebbe riuscire a determinare l'utente (l'amministratore) che ha effettuato l'accesso.

Benefici Raccolta centralizzata dei log con relativamente poco lavoro, il sistema Log_mining verrà aggiornato e sarà implementata la funzione di lettura dei log riguardanti gli accessi, sfruttando una risorsa già disponibile.
La situazione di Username e password resterà inalterata, il che sarebbe un vantaggio essendo un sistema già assodato e non vengono richieste modifiche ad applicativi o altri sistemi che necessitano accessi a questi server.

Problemi Il soddisfacimento della legge sarebbe parziale, infatti non sempre è possibile ricavare il nome di una persona tracciando l'IP delle varie connessioni.
Sicurezza del Firewall aziendale ridotta a causa della necessità di aprire porte verso l'interno

Tempi É la soluzione più veloce: attivazione dell'auditing e installazione di Snare sui Windows Servers 1 giorno per server, attivazione di Syslog-ng per l'auditing su ogni macchina Linux 30 minuti per macchina, configurazione syslog-ng sul Server di raccolta 1 giorno.
I tempi che saranno allungati saranno quelli riguardante l'analisi dei log prodotti, starà infatti all'analizzatore effettuare il tracciamento degli ip memorizzati.

Rischi *Condivisione user/password*
Se più persone condividono lo stesso set di user e password, e qualcuno ne viene

a conoscenza, non si è in grado di determinare chi le ha conservate male.

Relativo alla DMZ

La rete esposta ad internet è soggetta ad attacchi, questi attacchi restano localizzati all'esterno grazie a Clavister al momento, ma se si desidera inviare Log dalla rete DMZ verso la rete interna per la raccolta, è necessario aprire almeno una porta verso l'interno, il che crea una falla di sicurezza, in quanto, attraverso questa porta, un utente malintenzionato sarebbe in grado di penetrare nella rete interna (la porta è rilevabile in molti modi, ad esempio esiste nmap: un comando che controlla le porte di un firewall/router/pc..); in pratica si lascia un passaggio libero all'interno del Clavister (già molti sono presenti per necessità organizzative e per i servizi forniti), in figura 7.4 è mostrato sia il problema che la sua risoluzione. Il problema è causato dal modo di operare di Syslog (e di ogni altro sistema

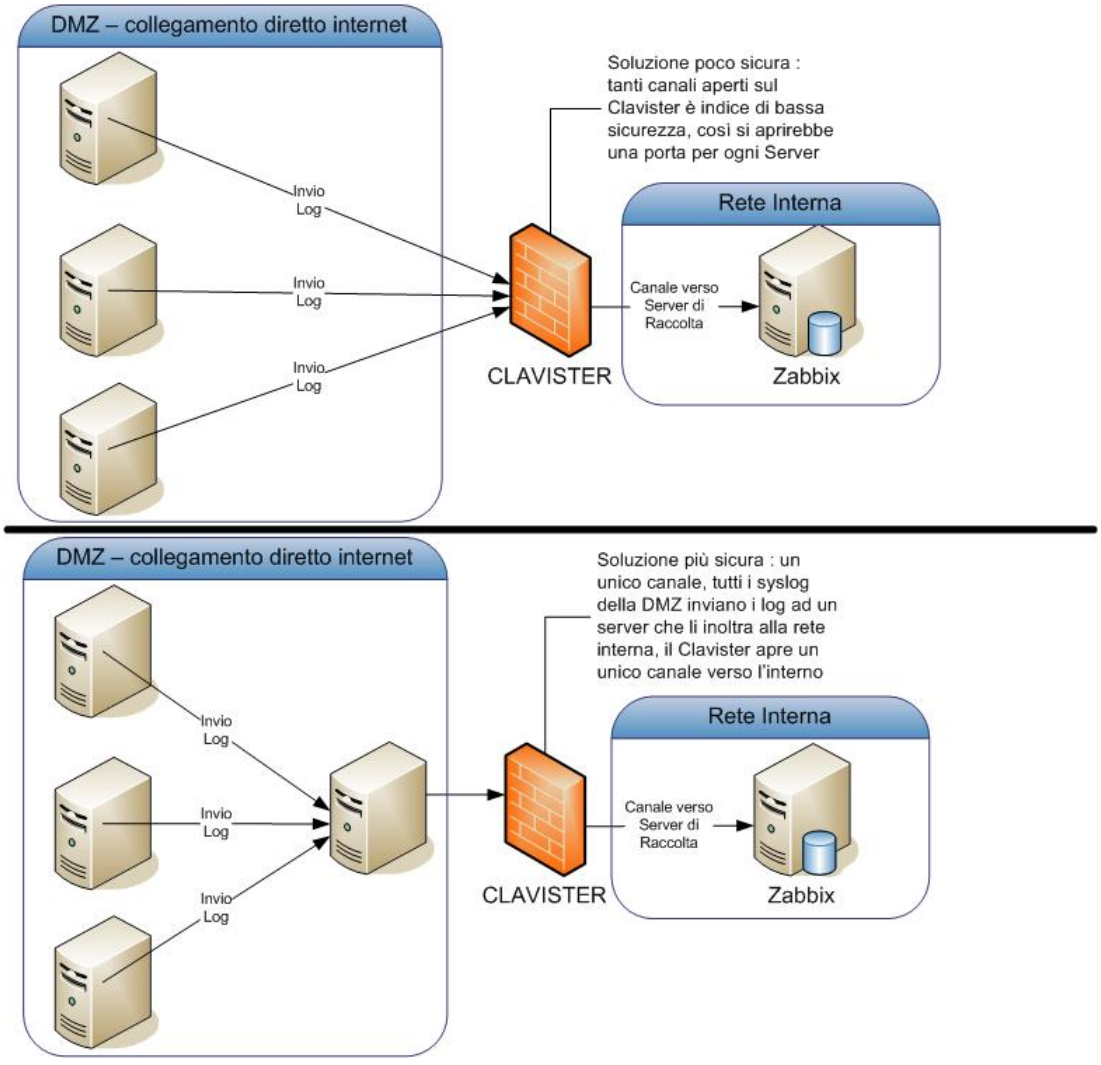


Figura 7.4: Problema relativo alla DMZ

di invio log), il syslog centrale (Zabbix) è in ascolto PASSIVO, i vari syslog sui server inviano attivamente pacchetti verso la rete interna, quindi il problema risultante è che Clavister deve restare aperto per traffico diretto verso l'interno. Sono stati trovati due sistemi di raccolta e tra questo è stato scelto il secondo essendo più sicuro del primo ma comunque crea un rischio.
Incertezza della provenienza delle richieste
I login remoti possono essere fatti in sequenza, creando catene di sessioni ssh aperte, normalmente è sufficiente osservare gli orari e gli ip di provenienza per determinare chi ha effettuato un certo accesso e da dove. Ma se venissero effettuati accessi sequenziali, utilizzando redenziali comuni (come root o administrator), non è possibile determinare chi ha effettuato un accesso già a partire dalla seconda sessione remota in sequenza.

7.5.2 Assegnazione Username personali

Questa soluzione coincide in gran parte con la precedente, l'unica differenza è che su ogni server presente nella DMZ (e sulle macchine della rete interna ancora con autenticazione locale) sarà presente un elenco di user e password personali, ovvero ogni amministratore avrà un proprio set di user/password.

Benefici Risoluzione dell'incertezza della provenienza delle richieste, un'eventuale smarrimento di password da parte di un amministratore sarebbe ora associabile non più all'insieme di amministratori ma al solo possessore delle sue credenziali personali.
Raccolta centralizzata: vedi proposta precedente.

Problemi La creazione di un set di user e password per ogni amministratore avrà come conseguenza la modifica di ogni database 'users' sulle macchine coinvolte (Server e tutte le macchine non autenticate attraverso il Domain Controller).
Vedi problemi soluzione precedente.

Tempi Creazione di un set di credenziali su ogni server nella rete DMZ e su ogni server nella rete interna.
Vedi tempi soluzione precedente.

Rischi Problemi causati dalla modifica delle user/password in alcuni sistemi che utilizzano accesso automatico
Relativi alla DMZ : vedi proposta precedente.

7.5.3 Creazione albero LDAP unico

Questa soluzione utilizza gli alberi LDAP descritti nel capitolo precedente, al momento gli alberi sono attivi e visibili tra loro, ma non formano un albero unico, la struttura che si propone con questa soluzione è rappresentata in figura 7.5 :

Nella quale anche le macchine linux presenti nella rete interna verranno collegate al Domain Controller tramite OpenLDAP, la raccolta log sarà affidata unicamente ai vari LDAP presenti, oppure all'LDAP NET in cima all'albero, mediante Snare per i server Windows e Syslog per i server Linux (OpenLDAP).

Benefici Centralizzazione degli accessi senza modificare gran parte della struttura già presente, i vari Windows Server verranno a capo di un unico Domain Controller posto alla radice di un albero di Active Directory, la gestione degli account risulta semplificata e centralizzata. I server nella rete interna e nella DMZ non sarebbero più a sé stanti (come autenticazione).

Problemi Il fatto di collegare la rete esterna con la rete interna può creare falle di sicurezza, una volta che un malintenzionato entra all'interno di un server nella DMZ sarebbe in grado di accedere alla rete interna, cosa che al momento è molto più complicata. Andrebbe incrementata la sicurezza.
Creare un albero unico avrà la necessità di ridefinire un elenco degli amministratori : quali solo di un ramo e altri che avranno diritti di root dell'active directory.
Mettere in comunicazione OpenLDAP con Windows Server crea spesso problemi causati dal formato della comunicazione, un sistema alternativo per comunicare col Windows Server è WinBind, grazie al quale è possibile definire un template per la comunicazione col Server, è una soluzione meno pulita rispetto all'OpenLDAP, ma funzionante, quindi verrà presa in considerazione in caso di problemi di interazione OpenLDAP-Windows Server (un'alternativa è installare un hotfix fornito da Windows appositamente per questo bug, ma essendo il Domain Controller attuale datato, ogni modifica potrebbe creare problemi).

Tempi Installazione di OpenLDAP su ogni Server DMZ, creazione di un Server OpenLDAP (slapd) per i server esterni $>$ 2 settimane.
Installazione di OpenLDAP sulle macchine Linux dentro la rete interna e collegamento di queste col Window Server della rete interna $>$ 2 settimane.
Creazione dell'albero di LDAP, ovvero collegare quelli già presenti+quello creato,

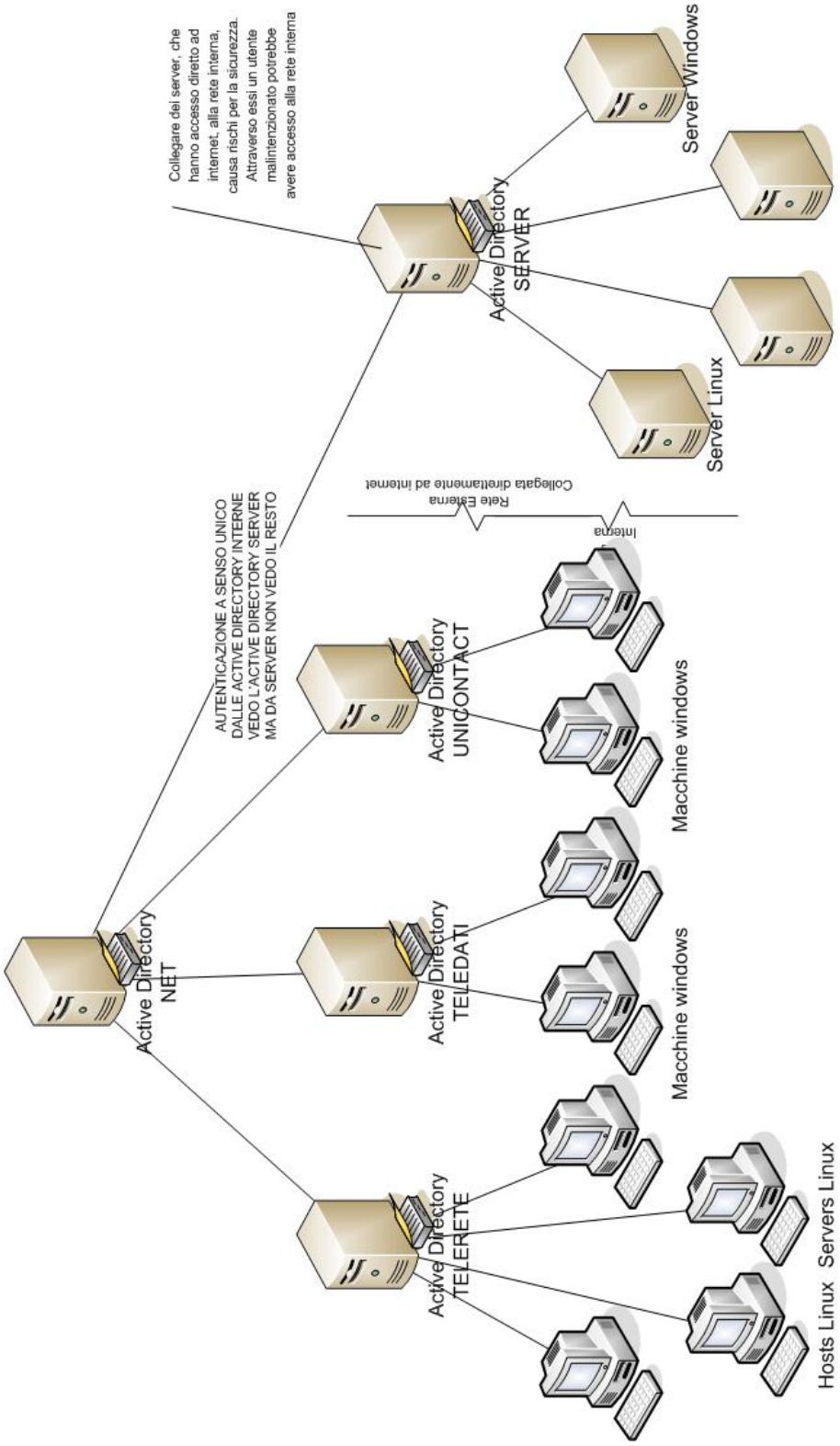


Figura 7.5: LDAP unico col ramo DMZ escluso parzialmente dal dominio

con un'altro LDAP (da creare) posto in cima alla gerarchia che avrà il ruolo di root.

Rischi Collegare mediante Active Directory la rete esterna con la rete interna rappresenta un rischio sulla sicurezza.
Problemi su applicativi che utilizzano il sistema attuale di autenticazione.

7.5.4 Single Sign On

Creare una macchina che funga da autenticazione centralizzata, questa poi invierà certificati a tutte le altre macchine, i certificati saranno accettati e il livello di accesso garantito da un raggruppamento: ogni certificato potrà essere di tipo amministratore, utente, poweruser, ecc..
É una soluzione migliorativa, centralizzare completamente gli accessi, come illustrato in figura 7.6, renderà la vita più facile agli amministratori, in quanto non sarà richiesto di autenticarsi nuovamente ad ogni accesso alle macchine esterne, una volta effettuato l'accesso ad inizio giornata il successivo scambio di certificati tra macchina centrale di autenticazione e tutte le altre, le autenticazioni risulteranno trasparenti.
Capire chi è connesso e dove, sarà più semplice, è sufficiente che la macchina centrale memorizzi ogni azione login/logout a che utente è associata e verso che macchina è diretta. Questa macchina centrale sarà l'unica che avrà il diritto di autenticarsi, verrà eliminata la possibilità di autenticazione tramite ssh, telnet, ecc..,questo per evitare catene di autenticazione da una macchina all'altra, cosa che renderebbe più complesso rintracciare il vero utente che ha compiuto l'accesso già dopo il secondo login sequenziale sulla stessa shell.

Si potrebbe applicare solo parzialmente, ovvero creare un albero interno di LDAP ed un SSO esterno relativo unicamente alle macchine presenti sulla DMZ e mantenerli separati, i log raccolti dal LDAP verranno inviati direttamente allo Zabbix, quelli raccolti dalla macchina SSO attraverseranno il Calvister (creando la falla di sicurezza prima descritta).

Benefici Centralizzazione degli account,
Sicurezza interna aumentata,
controllo accessi semplificato,
Riduzione del numero di password e di utenti attivi su ogni macchina (un user a

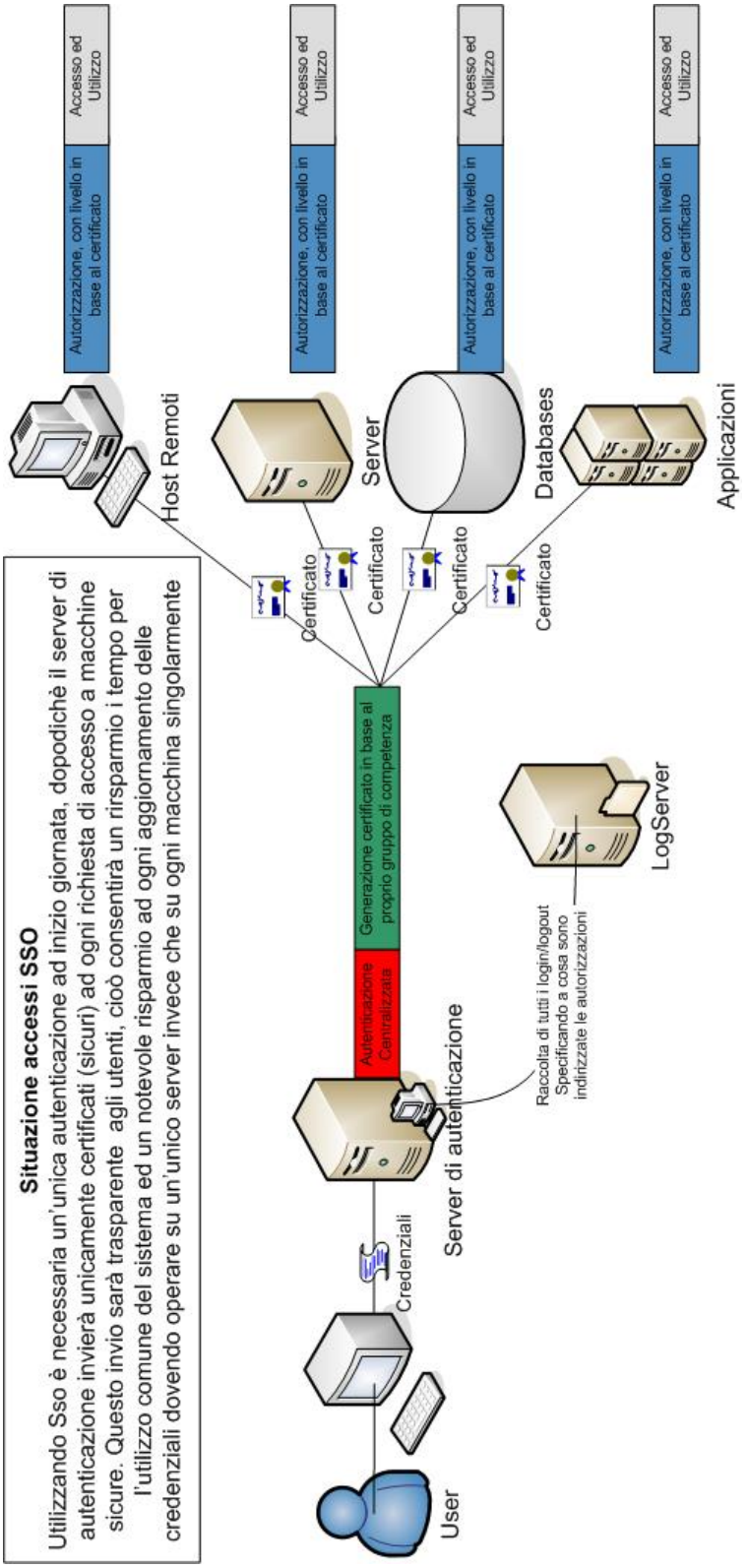


Figura 7.6: SSO

persona e ad ogni user verranno associati determinati permessi, questi permessi verranno comunicati alle altre macchina in caso di richiesta di autenticazione).

Problemi Modifica radicale del sistema di accessi per gran parte delle macchine, questo richiede uno studio delle macchine sia linux che windows e dei loro sistemi di autenticazione, una modifica del sistema di accesso, la creazione di un sistema centrale di autenticazione attraverso il quale è necessario passare per accedere alle altre macchine.
Gli accessi in locale su ogni macchina vanno gestiti tramite Snare e Syslog, rendendo questa soluzione sovrapponibile alla seconda.
Creazione di un sistema centralizzato che va a sostituire l'LDAP ha poco senso, meglio aggiungere SSO separato dove non c'è e creare nuove gerarchie di autenticazione.
Difficoltà tecniche considerevoli.

Tempi Studio dei vari sistemi di autenticazione $>$ 3 giorni (a sistema).
Creazione di un sistema centralizzato o modifica del sistema attuale $>$5 giorni.
Implementazione del sistema sui client e sul server centrale, modifica incrementale, una volta creato il sistema di autenticazione/generazione certificati, si opererà su di una macchina alla volta, obbligandola a ricevere autenticazioni unicamente dal server : 4 ore a macchina.
Valutazione di un sistema di accesso che non crei falle di sicurezza per richieste destinate alla rete DMZ.

Rischi Creare un collo di bottiglia a causa della necessità di continuare a creare certificati per tutti gli utenti dell'azienda.
Modificare un sistema LDAP già funzionante creerebbe instabilità iniziale e problemi di aggiornamenti futuri (dall'esperienza di passaggio dai Windows Server 2000 a Windows Server 2003).
Unificare reti interne alla DMZ potrebbe creare facilitazioni per le intrusioni esterne, non solo alla DMZ ma anche alla rete interna.

7.6 Sistemi per la Raccolta dei log

Sul mercato sono presenti Sw nati di recente appositamente per soddisfare le esigenze della nuova normativa sul controllo degli accessi degli amministratori. Non verranno prese in considerazione soluzioni a pagamento in questa proposta.

I requisiti imposti sopra richiedono una gestione centralizzata dei log riguardanti le operazioni di logon eseguite degli amministratori.
La comunicazione tra punti di raccolta log e server centrale di mantenimento dei log, e la certificazione dei log raccolti avverrà tramite OpenSSL e Stunnel, entrambi presenti sia per Windows che per Linux.

7.6.1 Server di Raccolta

Il server di raccolta proseguirà la strada intrapresa per il progetto di log mining, si utilizzerà il Syslog-ng su macchina Linux con un cron job atto alla compressione dei log giornalieri.
Syslog-ng è appunto già attivo per la raccolta dei log riguardanti PadovaWifi e MonseliceWifi. Al momento dell'ingresso nel server ai dati verrà aggiunto un hashing riga per riga successivamente verranno prima inoltrati su una PIPE gestita da stunnel, il quale effettuerà una crittografia rapida ed infine salverà il tutto all'interno di un file giornaliero. Rispettando così la caratteristica di integrità dei log, modificare dati cifrati non produce dati sensati e l'hashing garantirà l'inalterabilità, la chiave privata sarà mantenuta unicamente dal responsabile asegnato, i dati saranno visualizzabili tramite l'utility Log Mining, dopo la creazione di un modulo separato creato appositamente per l'analisi e la visualizzazione dei log dedicati all'auditing.
La completezza dipende dalle informazioni inviate dai client.

7.6.2 Client Windows

L'auditing, sulle macchine windows, è già presente ed è sufficiente attivarlo, per risolvere il problema dell'invio al server centrale di raccolta sono stati rilevati vari servizi da installare, molti di questi free, tra quelli testati i migliori in termini di log prodotti, interfaccia e leggerezza sono:

1. Snare
2. NTSyslog
3. KiwiSyslog
4. WpSyslog2

Snare

L'agent da installare sui client è open source e serve a prelevare una serie di eventi da un sistema e, se impostato correttamente, li invia come log ad un server, documentazione ampia e completa, molto utilizzato. Si può installare anche sulle macchine linux, è già installato e in fase di testing su una macchina all'interno dell'azienda, rileva tutti i dati di interesse senza creare fastidi di alcun genere.

Benefici : Soluzione presa in considerazione, snare offre un servizio pulito e leggero, l'interfaccia per l'aggiornamento è web e molto veloce. Una funzionalità indicata nel nostro caso, è la possibilità di creare pacchetti di installazione per windows preconfigurati, così da evitare l'installazione e la configurazione sequenziale di molte macchine, cosa che aumenterebbe di molto la probabilità di commettere errori. Una volta prodotto e testato un pacchetto di installazione è sufficiente procedere con l'installazione in remoto su tutte le macchine interessate senza la necessità di configurare più nulla.
Semplicità di installazione e manutenibilità, aggiornamento costante del software da parte di 'InterSectAlliance',

Problemi : I log prodotti non sono modificabili, produce dati non pertinenti allo scopo, dati che saranno ignorati in fase di analisi dei log, ma che occuperanno spazio sul disco.

Tempi : Creazione e configurazione di un pacchetto .msi, tempo previsto = circa 4 Ore; installazione semplice circa 10 minuti.
Installazione di ogni pacchetto con testing , tempo previtso = 5 minuti per macchina.
La differenza di tempo non è sufficiente all'impegno inziale di creare un pacchetto apposito, verrà installata ogni istanza singolarmente.

Rischi : In caso di modifica dell'elenco degli amministratori verrà necessariamente modificato l'elenco su ogni macchina su cui è installato (i Windows Server);

NTSyslog

Servizio di logging dedicato a Windows XP, 2000, NT, non essendo riuscito a reperire una documentazione ampia e casi d'uso promettenti non è stato preso in considerazione - http://ntsyslog.sourceforge.net/

Benefici : Sistema di logging creato appositamente per inviare in remoto i log da clien Windows verso un server Linux, che è il nostro caso.

Problemi : Scarsa documentazione

Tempi : Installazione + configurazione >2 ore per macchina

Rischi : Probabilmente prodotto obsoleto In caso di modifica dell'elenco degli amministratori verrà necessariamente modificato l'elenco su ogni macchina su cui è installato (i Windows Server);

KiwiSyslog

Offre una soluzione completa di server di raccolta, il prodotto costa 200 euro , ma la parte del server non è interessante in quanto svolge le stesse funzioni di Syslog ma in versione grafica, il prodotto utile è il log forwarder per Windows, che produce gli stessi risultati di altri prodotti free.

WpSyslog2

Sistema sviluppato da Word Press, buono ma ha la pecca di non essere impostabile per inviare i log raccolti ad un server syslog remoto, avrebbe la necessità di avviare su ogni macchina Windows due sistemi di logging : WPSyslog2 per la raccolta e l'invio ad un demone syslog adibito all'invio al server remoto.
Appesantisce la macchina più di altri sistemi, per maggiori informazioni vedere: http://www.bufferoverflow.it/2009/09/26/

Benefici : Si dota il sistema di un applicativo di cui è disponibile il codice e che svolge le funzioni necessarie allo scopo.

Problemi : Essendo legato unicamente a Windows richiederebbe di avviare più servizi sui PC sotto controllo, appesantendo la macchina più del dovuto.

Tempi : Installazione + configurazione >2 ore per macchina.

Rischi : Incompatibilità con alcune versioni di Windows.
In caso di modifica dell'elenco degli amministratori verrà necessariamente modificato l'elenco su ogni macchina su cui è installato (i Windows Server);

7.6.3 Client Linux

Per le macchine linux le soluzioni in prima battuta erano sembrate più facili e meno problematiche in termini di installazione e configurazione rispetto alle soluzioni Windows.
Effettuando dei test sono state riscontrate delle problematiche sia per le necessità dei software per effettuare l'installazione che nelle soluzioni provate in quanto non rispettano il carattere di completezza che i log devono avere. Le soluzioni trovate sono le seguenti :

1. Snare 0.98
2. Snare 1.5.0
3. Syslog

Snare 0,98

Soluzione presa in considerazione inizialmente per le stesse motivazioni del caso windows. Ipotesi scartata in fase di analisi a causa della richiesta di effettuare un patch di ogni kernel, operazione lunga e dispendiosa, che potrebbe inoltre causare problemi ad applicativi più importanti per l'azienda.

Snare 1,5,0

Descrizione : Soluzione presa in considerazione, la validità del Software è indubbia, inoltre questa versione, unicamente per Linux, offre più personalizzazione della precedente, richiede l'aggiornamento delle macchine al kernel 2,6,13

Benefici : Software compatibile con la versione Windows. Stessi vantaggi di windows

Problemi : Richiede l'aggiornamento di ogni macchina al kernel 2,6,13 o superiori, operazione molto lenta e spesso pericolosa a causa dell'incompatibilità di alcuni pacchetti con versioni nuove degli stessi.

Tempi : L'aggiornamento può richiedere più giorni per ogni macchina.
La successiva installazione circa un'ora per macchina(con testing)

Rischi : Creare problemi ad applicativi Core per l'azienda a causa dell'aggiornamento generale che viene richiesto, come è successo per una macchina Gentoo quando è sorta la necessità di aggiornarla a causa delle librerie obsolete che conteneva.

Syslog Scelta attuale

Soluzione più intuitiva, non richiede l'installazione di nessun software aggiuntivo, quindi il requisito di leggerezza sarebbe ampiamente soddisfatto, sarebbe necessaria una configurazione singola e personalizzata per ogni macchina, con conseguente aggiornamento (macchina per macchina) ad ogni eventuale modifica dell'elenco degli amministratori.

Benefici Attesi : Semplicità della soluzione, non necessita prodotti esterni ed è un sistema built-in di linux

Problemi : Struttura dei log raccolti completamente diversi tra Linux e Windows; Necessita aggiornamenti delle librerie presenti per l'autenticazione e una modifica del contenuto di queste. Richiede un aggiornamento di massa oppure la creazione di un modulo apposito per le librerie PAM (la libreria utilizzata per effettuare l'autenticazione). Questo modulo va creato compatibile con tutte le versione passate, presenti e possibilmente future della libreria PAM, oppure uno differente per ogni versione, ma non assicura la risoluzione del problema.
Difetto rilevato in fase di testing : I login vengono correttamente rilevati e raccolti, il logout, se la sessione è di login remoto e proviene da una macchina Windows, non viene notato dalla macchina, questo viene risolto aggiornando le librerie PAM di sistema.
Il problema è riscontrato unicamente in macchine non aggiornate da più di un anno

Tempi : Variabili, dipendentemente dai problemi che si riscontreranno in testing.
In prima battuta era sembrata la soluzione ideale, rapida ed efficiente, ma si è rivelata incompleta.
Se tutto va liscio si configura ogni macchina in circa mezz'ora (testing compreso).

Rischi : Creare problemi nell'autenticazione se si modificano le librerie PAM

7.6.4 Database

Seguono i due sistemi identificati come validi per memorizzare gli accessi eseguiti nei Database.

Trigger

Ogni DBMS offre la possibilità di implementare Trigger, in questo caso si attiverebbero all'accesso e alla disoconnessione, un problema è stato rilevato verso i trigger di Mysql, il quale rileva correttamente i login ma non i logout, una possibile soluzione era : creare una tabella nella quale salvare per mezzo dei trigger l'ora di inizio di ogni sessione e ad ogni query inserita aggiornare la data di fine sessione, questa modalità avrebbe se non altro memorizzato la data dell'ultima azione compiuta, che solitamente è equivalente alla disconnessione.

Benefici : Risoluzione del problema.

Problemi : Per ogni operazione eseguita (in mysql) si attiva il trigger che modifica la riga nella tabella degli accessi, a lungo andare raddoppiano le operazioni eseguite, con conseguente rallentamento delle base dati.
Una volte che il DBMS inserisce i dati nella tabella va trovato il sistema di estrarre i dati dal DB ed inoltrarli via syslog al server di raccolta.

Tempi : Dipendenti dai problemi che si riscontreranno.
Mediamente la creazione dei 4 trigger 5/6 ore per ogni DBMS, a causa della necessità di identificare gli eventi specifici di ogni DBMS e risolvere eventuali problemi.

Rischi : Tabella accessibile a tutti prima dell'invio.
Rallentamenti eccessivi.
Problemi con la lettura del Database e nell'invio tramite syslog.

Acces Log : Scelta Attuale

Questa soluzione prevede di attivare su ogni DBMS l'archiviazione di ogni operazione all'interno di log, e affidare a syslog-ng la la lettura di questo logfiles e l'invio delle righe utili al server di raccolta

Benefici : Questa soluzione al contrario dei Trigger non cera rallentamenti, salva ogni comando che riceve all'interno di un logfile, coprendo, senza effettuare operazioni non necessarie, le cogenze di legge.

Problemi : Crea all'interno di ogni macchina grandi quantità di log che non servono allo scopo, si può risolvere con un CronJob atto all'eliminazione giornaliera dei logfiles.

Tempi : Per ogni DBMS è sufficiente modificare il file di configurazione aggiungendo questa opzione, e syslog-ng per l'invio al server di raccolta.

Rischi : Rallenamento di Syslog a causa del controllo continuo sugli access log di ogni DBMS.

7.7 Modello Progettuale

La gestione del progetto richiedeva una scelta metodologica per procedere, questa scelta avrebbe influenzato il modo di procedere nel resto del progetto, anche se l'obiettivo resta quello di tentare di soddisfare tutti i requisiti entro i tempi prefissati. Il tempo a disposizione era in larga misura insufficente per riuscire a ricoprire tutti i requisiti in tempo, quindi andava scelta una strategia che garantisse, entro i tempi previsti dalla legge, il raggiungimento degli obiettivi 'core' per il sistema di Auditing che si andava a creare.
Le scelte utili per procedere erano ristrette, tra queste solo due potevano essere prese in considerazione :

- Modello Sequenziale lineare a cascata
- Modello Iterativo Prototipale
- Modello Evolutivo a spirale

Il primo certamente non avrebbe garantito il raggiungimento del successo, ovvero avrei impiegaro troppo tempo nella fasi di analisi generale, e sarebbe stato impossibile riuscire a terminare la fase di raccolta in tempi ragionevoli, per questo motivo è stata subito scartata. L'alternativa del modello prototipale non era adattabile al caso in questione, infatti venendo a mancare un cliente, che provasse ogni prototipo, perde gran parte della sua utilità, talvolta è stato utilizzato questo sistema in fase di sviluppo del log Analyser, utilizzando come 'cliente' un membro

dell'area tecnica, il quale porgeva domande circa l'utilizzo ed eventualmente ne evidenziava mancanze.

Il modello evolutivo a Spirale fornisce le caratteristiche utili alla corretta prosecuzione del progetto, rilasciado man mano che si sviluppavano, le nuove funzionalità, come si può notare dall'immagine 7.7, che rappresenta questo modo di procedere, ogni fase inizia con una comunicazione e termina con un rilascio.

Nel caso in questione i rilasci, o le fasi che si sono susseguite sono 3 :

1. Nel primo rilascio, da effettuare obbligatoriamente entro i termini imposti dalla legge, la piattaforma ha la sola necessità di memorizzare tutti i logs necessari, quindi vanno configurate tutte le macchine per effettuare l'allacciamento all'LDAP oppure, in caso di problemi, l'impostazione delle credenziali personali e l'invio via Syslog/Snare al server di raccolta; il centro di raccolta, a tal proposito, ha la necessità di essere configurato con priorità massima, in quanto, se le altre macchina inviano logs ma il server non è in grado di riceverli, si perdono dati inutilmente. Quindi sono da impostare le regole per la corretta ricezione ed organizzazione dei logs.

2. Nella seconda fase vanno gestiti i logs appena memorizzati, quindi, una volta impostate tutte le macchine, il server di raccolta ed il Domain Controller, sul server di raccolta è necessario impostare correttamente la gestione dei dati a lungo termine. Inoltre vanno ricordati i requisiti circa il mantenimento dei dati, che devono essere immodificabili ed inaltrebili, inoltre data la mole di dati vanno configurati sistemi per la compressione giornaliera dei dati e l'eliminazione dei logs non più utili.

3. La terza fase consiste nel fornire un sistema di analisi dei log raccolti, questa fase si suddividerà a sua volta in varie sotto-fasi, ognuna delle quali fornisce una versione del prodotto. Le caratteristiche che potrebbero contraddistinguere le varie versioni sono le seguenti :

   (a) Lettura semplice dei dati,

   (b) Migliorare il livello di sicurezza,

   (c) Rendere il sistema user friendly,

   (d) Ottimizzare il sistema e rendere grafici i risultati,

   (e) Standardizzazione delle funzioni.

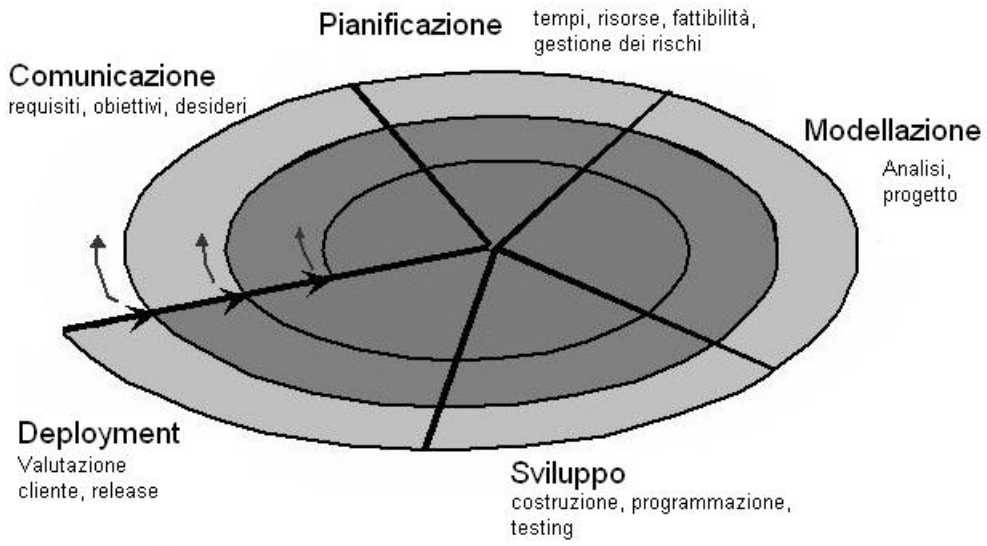


Figura 7.7: Modello evolutivo a spirale

Le varie fasi descritte nel grafico 7.7 saranno le seguenti:

1. Comunicazione: solitamente con l'area tecnica o col Professor Filira valutavo eventuali necessità o problemi nelle varie versioni prodotte, mi venivano comunicate le priorità in base alle quali procedere con l'evoluzione del sistema, le nuove funzionalità e le problematiche riscontrate,

2. Pianificazione: terminata la stesura dei micro requisiti per la versione successiva veniva prodotto un Gannt temporaneo e veniva definita la fattibilità delle nuove caratteristiche richieste,

3. Modellazione: dopo aver fissato i 'paletti' temporali e progettuali veniva effettuata una breve fasi di analisi e di progettazione, per definire le metodologie da utilizzare, la struttura che il codice da stendere dovrà avere (funzioni, strutture ecc..) ed il design del progetto al prossimo step

4. Sviluppo: in questa fase la nuova versione del progetto viene effettivamente sviluppata e viene effettuate un testing parziale per ogni modifica apportata.

5. Deployment: al termine del lavoro viene rilasciata la release corrente e viene mostrato all'area tecnica il lavoro fatto, si verifica la soddisfazione dei membri dell'area tecnica e degli utilizzatori del sistema, al termine .

Capitolo 8

Implementazione Auditing

Terminata la fase di analisi del problema e di progettazione è stata è avviata al più presto la fase di implementazione, la scelta del progetto da seguire è stata guidata dalla situazione interna in azienda, andando ad optare per un LDAP separato e misto (Activer Directory/OpenLDAP), utilizzando la gestione dei rischi ogniqualvolta venissero riscontrati problemi.

La tecnica risolutiva intrapresa consisteva nel procedere con la configurazione predefinita, ovvero l'aggancio di macchine all'LDAP, in caso di problemi, dopo aver fissato un tempo massimo per la loro risoluzione, ed aver oltrepassato tale limite, si procede con la soluzione di sicurezza, ovvero la più semplice ed inuitiva, pur di portare a termine il progetto nei tempi previsti.

8.1 Aggancio macchine Linux all'LDAP aziendale

L'aggancio di macchine Linux ad alberi LDAP Windows è stato l'aspetto più sensibile del lavoro, infatti, come ben si sa nell'ambiente, i due sistemi non sempre vanno molto d'accordo, come appunto è successo durante la mia esperienza.

L'approccio al problema è iniziato con uno studio teorico riguardante l'LDAP e le varie implementazioni o utility che potevano essermi di aiuto per risolvere la problematica della comunicazione Linux-Windows in un ambiente delicato come l'LDAP, i software cui mi sono avvicinato maggiormente sono stati:

- LDAP: ho innanzituto studiato l'LDAP nativo, definito per Windows e per strutturarne i domini, tra macchine windows é relativamente semplice creare gruppi di lavoro e Domini aziendali;

- OpenLDAP: la struttura fornita da OpenLDAP era inizialmente la strada che intendevo percorrere, essendo molto simile, almeno a prima vista, a quella che i sistemi Windows utilizzano; non é stata la scelta definitiva in quanto ha una struttura predefinita che non crea problemi in presenza di altre macchine OpenLDAP, ma quando si trova a dover comunicare ed autenticarsi con sistemi Windows, le differenze di implementazione dei due sistemi vengono allo scoperto, non sono stato infatti in grado di collegare in maniera stabile una macchina Linux al dominio Windows presente in azienda utilizzando unicamente OpenLDAP, era necessario affiancare questo prodotto a Winbind per portare a termine il mio obiettivo;

- Winbind: è stata presa in considerazione dopo qualche tentativo, in quanto forniva una soluzione meno elegante rispetto all'utilizzo del solo OpenLDAP, ma l'unica soluzione valida, a suo vantaggio offriva una maggiore semplicità d'uso rispetto alla riconfigurazione dei pacchetti samba; abbinato a Kerberos ed a Samba é stato il mezzo che mi ha permesso di allacciare alcune macchine all'LDAP aziendale, il suo ruolo in questo caso é di tramite fra le metodologie windows e quelle Linux;

- Samba: è uno standard nel mondo Linux, praticamente ogni volta che si parla di condivisione Samba è impiegato, era inevitabile che anche nell'LDAP non venisse chiamato in causa. Samba è una implementazione del protocollo SMB/CIFS per sistemi UNIX che fornisce il supporto per condividere file e stampanti fra piattaforme diverse come Microsoft Windows, OS X e altri sistemi UNIX. Samba può anche funzionare da domain controller in stile NT4 e può integrarsi sia con domini NT4 e realm Active Directory come membro server. Questa ultima caratteristica è molto importante, infatti Samba tende a diventare il PDC, nella configurazione va specificato di mantenere la macchina come client del Dominio.

Allo studio teorico ho affiancato esperimenti di 'aggancio' all'albero, verificando, volta per volta, le affermazioni e gli approcci studiati, spesso, però, le cose non erano 'semplici' come venivano descritte nella teoria studiata. Le difficoltà maggiori sono arrivate al momento dell'installazione dei pacchetti in quanto, a causa della diversitá che é presente fra ogni macchina della rete aziendale, non riuscivo a creare una combinazione di pacchetti 'standard' che fosse installabile ovunque.

I problemi erano causati soprattutto dalla datazione delle macchine, alcune non venivano aggiornate dal 2006, tutte le altre andavano da due a tre anni di

'anzianità', e non sempre ero in grado di trovare pacchetti adatti alla situazione, alcune volte si tentava un aggiornamento generale della macchina, ma, sempre a causa della mancanza di compatibilità tra pacchetti nuovi e kernel vecchi, non era sempre possibile aggiornare le macchine, in quanto non esistevano nuove versioni dei servizi o le nuove versioni andavano configurate in maniera differente; ho quindi deciso che, in caso di difficoltà causate da queste problematiche, avrei scelta la strada più semplice e sicura, anche se non coincidente col progetto iniziale.

La via più semplice che avevo a disposizione, per perseguire l'obiettivo, è chiaramente quella che incide meno sulla macchina, quindi ho scelto di modificare leggermente le librerie di autenticazione Linux, le librerie PAM, aggiungendo opzioni per il logging, successivamente ho creato credenziali personali, una per ogni amministratore presente, ed infine ho configurato l'istanza di Syslog della macchina problematica per l'invio verso il centro di raccolta. Le modifiche fatte riguardano il file centrale della librerie, ovvero system-auth, in particolare ho modificato il livello di logging delle sessioni:

Blocco di codice 8.1

```
session required pam_unix.so audit try_fist_pass authok
```

Con ciò si indica al sistema di effettuare un logging più intenso del normale, infine, dove fosse necessario è stato aggiunto il servizio `auditd`, il quale, se propriamente configurato, effettua un servizio di logging di livello leggermente superiore, anche in questo caso il pacchetto è stato applicato dove non creasse problemi all'albero delle dipendenze della macchina.

Il software utilizzato per il collegamente è costituito da un insieme di programmi che forniscono diverse funzionalità, tra questi spiccano in particolare tre moduli: smbd, nmbd e winbindd. I primi due, sostanzialmente, consentono la condivisione dei file utilizzando il protocollo SMB: da una qualunque macchina Windows si potrà avere accesso ai file della macchina Linux in modo del tutto trasparente. Winbind consente la condivisione di account e l'autenticazione sulla rete. In pratica rende visibili ad un sistema Linux gli utenti del Dominio Windows.

8.1.1 Sincronizzazione

Per prima cosa le macchine vanno sincronizzate fra loro, in modo da evitare che si scambino dati e messaggi incoerenti. Ciò avviene grazie ad un demone

che abbiamo già visto nei capitoli precedenti, ma stavolta le macchine vanno sincronizzate unicamente con il Primary Domain Controller

Blocco di codice 8.2

```
NTPCLIENT_CMD="ntpdate"
NTPCLIENT_OPTS="-s -b -u PDC.dominio.net"
```

Inoltre è stata assegnata ad ogni macchina una lista di alias che la identificano, la lista di hostname è stata specificata nel file /etc/hosts

Blocco di codice 8.3

```
127.0.0.1   nome.dominio.net nome localhost
::1   localhost
```

8.1.2 WinBind

Winbind è un componente della Suite di programmi Samba, il compito di questo modulo è di risolvere di le problematiche riguardanti il logon unificato. Winbind utilizza un'implementazione Unix delle RPC calls di Microsoft, dei moduli per le librerie PAM (Pluggable Authentication Modules), e infine il Name Service Switch, per permettere a utentu appartenenti a domini Windows NT di apparire ed operare come utenti Unix su di una macchina Unix.

Per utilizzare il pacchetto sulle macchine Gentoo vanno specificate le Use Flags, nel nostro caso è necessario indicare quanto segue:

Blocco di codice 8.4

```
vi /etc/portage/package.use
  -->  +net-fs/samba ldap ads winbind
vi /etc/conf.d/samba
  -->  daemon_list="smbd nmbd winbind"
```

Nel nostro caso non andava configurato il componenete, è sufficiente installarlo con le corrette flags, in modo che Samba lo utilizzi.

8.1.3 Kerberos

Kerberos è un protocollo di rete per l'autenticazione tramite crittografia che permette a diversi terminali di comunicare su una rete informatica insicura provando la propria identità e cifrando i dati. Kerberos previene l'intercettazione e i replay attack, e assicura l'integrità dei dati. I suoi progettisti mirarono soprattutto ad un modello client-server, e fornisce una mutua autenticazione, sia l'utente che il fornitore del servizio possono verificare l'identità dell'altro.

Kerberos si basa sulla crittografia simmetrica e richiede una terza parte affidabile.

La prima cosa da fare è quindi accertarsi che la propria macchina venga accettata e comunichi esattamente col PDC, per fare ciò vanno installati sulle macchine i pacchetti che implementano il protocollo Kerberos, vanno valutate eventuali problematiche di installazione sulle macchine, nel caso del nostro server ed esempio ho risolto un conflitto dei pacchetti aggiornando e reinstallando i pacchetti necessari come gli ultimi due nell'estratto che segue:

Blocco di codice 8.5

```
emerge sys-auth/pam_krb5 app-crypt/mit-krb5 util-linux e2fsprogs
etc-update
```

Seguiranno le righe piú rappresentative della configurazione riguardante Kerberos, sono state omesse tutte le righe lasciate di default e quelle riguardanti le tempistiche e le dimensioni dei pacchetti che Kerberos utilizzerá:

Blocco di codice 8.6

```
[logging]
default = FILE:/var/log/krb5.log
# specifico il dominio di default
[libdefaults]
   default_realm = DOMINIO.NET
# qua si specificano i 'reami' di azione
[realms]
   DOMINIO.NET = {
      kdc = PDC.DOMINIO.NET:port1
      admin_server = PDC.DOMINIO.NET:port2
      default_domain = DOMINIO.NET
```

```
   }
#qua indico gli alias di dominii
[domain_realm]
   .DOMINIO.NET = DOMINIO.NET
   DOMINIO.NET = DOMINIO.NET
```

Questa configurazione definisce unicamente il server col quale si ha intenzione di 'parlare' ed il relativo dominio di cui la macchina andrá a fare parte, per poter iniziare l'effettiva comunicazione é prima necessario creare un ticket Kerberos e verificarne la validitá, cosa che andremo a fare dopo aver terminato la configurazione del client Samba, per permettere la condivisione di risorse.

8.1.4 Samba

Samba é un'implementazione SMB/CIFS che permette condivisione ed interazione tra macchine differenti in ambienti che non siano Windows. Le caratteristiche che garantiscono il successo di Samba sono le seguenti :

- Samba fornisce servizi di condivisione di file e stampanti a client SMB/CIFS,
- al contrario di altre implementazioni SMB/CIFS, è liberamente disponibile,
- permette di ottenere interoperabilità tra Linux, Unix, Mac OS X e Windows,
- è un software che può girare su piattaforme che non siano Microsoft Windows, per esempio, UNIX, Linux, IBM System 390, OpenVMS e altri sistemi operativi,
- Samba utilizza il protocollo TCP/IP utilizzando i servizi offerti sul server ospite,
- quando correttamente configurato, permette di interagire con client o server Microsoft Windows come se fosse un file e print server Microsoft agendo da Primary Domain Controller (PDC) o come Backup Domain Controller,
- può inoltre prendere parte ad un dominio Active Directory,

Questo ultimo punto é la parte che noi andremo ad esplorare e sarà l'obiettivo della configurazione. Nella configurazione fa sempre ricordato che Samba nasce

anche come PDC o BDC, quindi bisogna fare attenzione a non creare conflitti con il PDC già presente in azienda.

All'interno di un dominio e di una rete è possibile avere un PDC ed uno soltanto, è concesso un DC secondario, detto Backup Domain Controller, che diventa utile in caso il primo abbia problemi, ma non è il nostro caso, essendo già presente anche il secondario, nella configurazione qui riportata sono state omesse le righe che definiscono se utilizzare la macchina come PDC, BDC o semplice client, se impostare le stampanti di rete, il sistema di rilevamento delle password e l'algoritomo utilizzato per crittografarle, e altre caratteristiche riguardanti il dominio.

La seguente configurazione, effettuata sul file /etc/samba/smb.conf, imposta una macchina per entrare all'interno del dominio dell'azienda Telerete, ho mantenuto i punti che indicano a Samba di sfruttare winbind:

Blocco di codice 8.7

```
workgroup = WGDOMINIO
# Nome visualizzato nel dominio
server string = Samba client %v - %h
# IP address dl dominio
hosts allow = **.**.**. 127.
# indico Active Directory Service come security
security = ADS
# Dominio di appartenenza
realm = DOMINIO.NET
# Server dal quale verificare le credenziali
password server = **.**.**.**
winbind enum users = yes
winbind enum groups = yes
# Creazione Home dirs
template homedir = /home/%D/%U
# Shell utilizzata al login
template shell = /bin/bash
# Indica il tipo di autenticazione che verrà
# utilizzata, no = UNIX, yes = Windows
client NTLMv2 auth = yes
# Specifica a winbind di utilzzare o meno il
# dominio di default
```

```
winbind use default domain = yes
```

8.1.5 OpenLDAP

Il software OpenLDAP è una implementazione libera, open source del Lightweight Directory Access Protocol (LDAP) sviluppato nell'ambito dell'OpenLDAP Project. E' stato rilasciato mediante una propria licenza in stile BSD denominata OpenLDAP Public License. LDAP è un protocollo indipendente dalla piattaforma e nel nostro caso è servito sulle macchine client per gestire il protocollo AD, sfortunatamente le implementazioni Microsoft sono totalmente incompatibili con le versioni di OpenLDAP, è stato quindi necessario sfruttare le potenzialità di Samba associato a Winbind.

La modifica al file di configurazione di OpenLDAP (/etc/openldap/ldap.conf) è opzionale, da fare in caso si riscontrassero problemi durante l'accesso al dominio, Samba solitamente riconosce queste informazioni e le inoltra all'OpenLDAP, ma, se si dispone di una versione datata o se sono presenti bug sulla macchina, può far comodo impostare l'OpenLDAP separatamente, stando attenti a farlo in modo coerente rispetto alla configurazione fatta su Samba :

Blocco di codice 8.8

```
BASE  dc=dominio, dc=net
URI   ldap://*****.dominio.net
SIZELIMIT   12
TIMELIMIT   15
DEREF    never
```

8.1.6 Test

Net Join Una volta configurato Samba si può procedere all join della macchina nel dominio, innanzitutto verrà inviato un ticket Kerberos per far identificare la macchina tra quelle affidabili dal Kerberos del Domain Controller, fatto ciò va aggiunta la macchina al dominio effettuando il net join, infine, se il net join dà l'ok, aggiungo samba alla run list di default, questo implica che, se il servizio Samba venisse interrotto, il sistema lo ripristinerà automaticamente.

Blocco di codice 8.9

```
kinit kszabo@DOMINIO.NET
```

```
klist
# se klist fornisce informazioni sul dominio si puó procedere.
net ads join -U administrator
/etc/init.d/samba start
#restart se è già avviato/
rc-update add samba default
```

Prima di effettuare ulteriori test, va aggiunto il demone winbind come fonte di dati per l'autenticazione, sul file /etc/nsswitch.conf si indica:

Blocco di codice 8.10

```
passwd:      compat winbind
shadow:      compat winbind
group:       compat winbind
```

Questo indica al sistema di ricercare le credenziali da confrontare anche dal servizio Winbind.

Prima di terminare la procedura è necessario verificare che il procedimento sia, finora, corretto, per fare ciò si effettuano due brevi test che, se riportano i risultati corretti, garantiscono che la macchina sia allacciata al dominio Windows. Per prima cosa verrà testata l'effettiva visibilità del dominio rispetto alla macchina che si sta utilizzando, successivamente le credenziali:

Blocco di codice 8.11

```
wbinfo -u
wbinfo -g
getent passwd
getent group
```

I risultati ottenuti da questi comandi devono includere, oltre all'elenco delle credenziali locali presenti sulla macchina, anche tutti i dati e le username del dominio, se così fosse vuol dire, per i primi due comandi, che la macchina è stata inserite nel dominio da parte del PDC, mentre per gli ultimi due significa che la macchina appena configurata riconosce di essere all'interno del dominio e verificherà le credenziali inserite anche con gli strumenti appena installati, ciò è possibile nonostante manchi un ultimo passaggio al completamento della configurazione, arrivati fin qui le credenziali verrebbero confrontate ed eventualmente

confermate da parte di OpenLDAP e Samba, ma non è possibile accedere alla macchina finchè i dati inseriti non vengono riconosciuti ed accettati anche dalle librerie PAM.

Andremo intanto a creare preventivamente una cartella di Dominio nella quale verranno poi inserite tutte le cartelle personali degli utenti che effettueranno l'accesso, la creazione delle cartelle personali viene effettuata da Samba, solo se confermata dalle librerie PAM:

Blocco di codice 8.12

```
mkdir /home/WGDOMINIO
chmod 777 /home/WGDOMINIO
```

A questa cartella assegneremo permessi 777, che in Linux significa praticamente nessun controllo, tutti scrivono e leggono in quella cartella, successivamente il sisterma creerà automaticamente le cartelle personali all'interno della cartella di cominio.

8.1.7 Librerie PAM

La modifica più sensibile al sistema è l'ultima della catena, qua si definiscono i metodi con i quali l'autenticazione avverrà all'interno della macchina, ovvero come verranno gestiti i dati creati dalle precedenti sezioni.

Modifiche errate in questi moduli possono portare all'esclusione dell'operatore dalla macchina e all'impossibilità di effettuare accessi, dipendentemente dalla gravità dell'errore, si può venire esclusi salla sola autenticazione via ssh, oppure anche da console, il che sarebbe molto grave, vanno quindi effettuate unicamente modifiche di cui si è certi, soprattutto in caso di macchine vitali per i servizi che l'azienda offre. In caso di esclusione completa dalla macchina è necessario spegnare la macchina, smontare l'Hard Disk e leggerlo da un'altra macchina, modificando le librerie in modo da consentire nuovamente l'accesso, fatto ciò sarebbe da verificare se le modifiche hanno avuto effetto o meno. Fortunatamente questa eventualità non si è mai presentata.

Riporto di seguito le linee più significative che sono state aggiunte nei moduli system-auth e sshd di una macchina Gentoo, le modifiche su altre distribuzioni variano di poco, ed il significato resta invariato.

Blocco di codice 8.13

```
# creazinoe della home personale
session  required   /lib/security/pam_mkhomedir.so
                    skel=/etc/skel/ umask=0022
# garantisce la visibilità delle credenziali
# di Dominio su locale
account  required   pam_winbind.so
# permette l'autenticazinoe attraverso le
# credenziali sopra definite
auth     sufficient pam_winbind.so
# gestione password di dominio
password sufficient pam_winbind.so
# gestisce le sessioni create dal modulo
# pam_winbind.so
session  required   pam_winbind.so use_first_pass
```

Al termine della configurazione va riavviato il servizio relativo all'autenticazione per rendere effettive le modifiche. A questo punto, il miglior test possibile, è tentare di effettuare un accesso, se il test precedente si è svolto correttamente, e se le configurazioni sono state fatte seguendo le indicazioni, si presume che questo passaggio non crei problemi, si termina così l'inserimento della macchina all'interno dell'LDAP aziendale.

8.2 Sistema di Raccolta

Il sistema di raccolta e centralizzazione dei log è basato sul protocollo syslog, utilizza pacchetti UDP diretti alla porta 514 nella rete interna, invece la porta utilizzata per la DMZ è la **** (per evitare di fornire una porta in ricezione ad un eventuale malintenzionato esterno), la raccolta va fatta considerando più l'eterogeneità dei log in arrivo piuttosto che la mole di dati, infatti, al contrario del progetto relativo all'antiterrorismo, non è un fattore decisivo per questa funzione in particolare, questa supposizione si è, tra l'altro, rivelata corretta dopo una verifica delle dimensioni dei log al termine dei primi tre mesi di funzionamento del sistema, le dimensioni medie dei logs compressi sono inferiori ai 20 MB al mese.

Il problema è la struttura della rete, che fornisce log da zone sicure come la rete interna, da zone a rischio come la DMZ, e altre fonti isolate. Questo

crea la necessità di suddividere i log in base a più regole al momento del loro arrivo, innanzitutto la divisione Linux/Windows, poi in base all'IP, alla data e alla facility quando possibile.

8.2.1 Struttura del sistema

La ricezione di log da macchine distribuite su due reti è stata molto facilitata dalla presenza di un server con 2 schede di rete, così facendo si evita il passaggio di dati attraverso il Firewall aziendale, infatti il server compare in entrambe le reti, per evitare il rischio di fare da ponte tra la DMZ e la rete interna sono state fatte regole apposite su Iptables, la macchina accetterà tutti i pacchetti tranne quelli che hanno l'obiettivo di essere inoltrati in una direzione o nell'altra.

Si nota dalla figura 8.1 che la macchina è in grado di accedere direttamente alle reti con un maggior numero di host interessati.

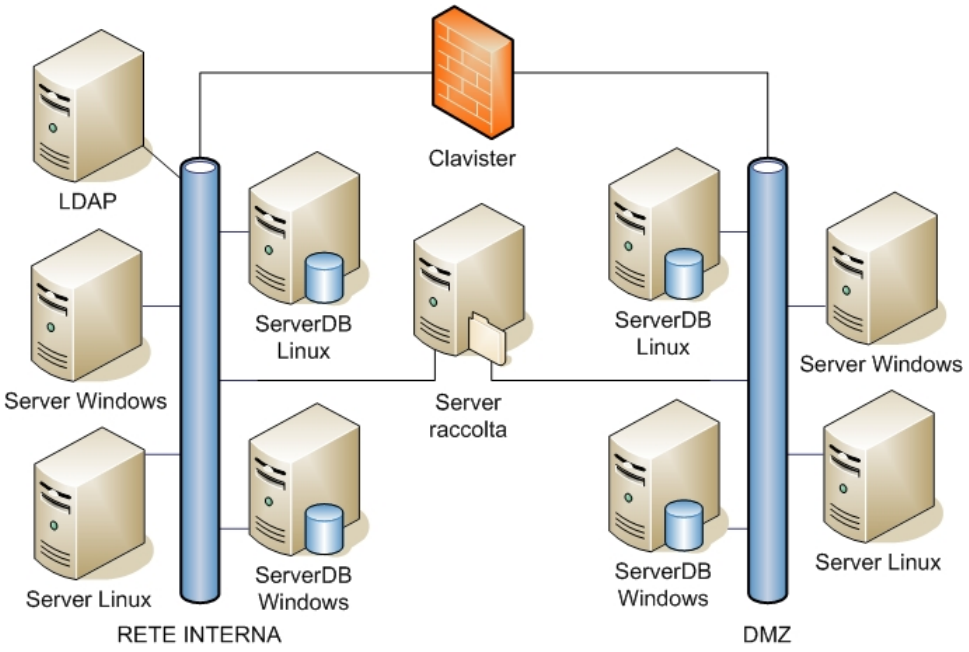


Figura 8.1: Struttura del sistema di raccolta dei logs rispetto alle reti in cui opera

8.2.2 Centro Raccolta

Inizialmente la raccolta era stata affidata alla stessa macchina che gestiva la grande mole di dati provenienti dai server Radius, essendo il flusso di dati previsto per l'auditing parecchio inferiore rispetto ai dati relativi all'anti terrorismo, quindi non di grande impatto sulle prestazioni generali della macchina.

Il problema è sorto quando, per ordinare e suddividere i logs in arrivo, mi sono trovato di fronte alla necessità di implementare regole più precise e definire filtri nella configurazione dell'istanza Syslog-ng presente sulla macchina. Non appena sono state applicate le nuove regole, la macchina non era più in grado di gestire il flusso dati, le regole ora prevedevano di leggere, oltre all'header syslog del pacchetto, anche il contenuto del log stesso, e filtrare/smistare i logs in base ai dati filtrati.

La valutazione dimensionale per la creazione delle regole era stata fatta sulla mole dati relativa all'auditing, mentre la macchina si è trovata di fronte a nuovi filtri per una mole molto più grande, causando un notevole rallentamento nello smaltimento dei logs in arrivo, fino al blocco del servizio Syslog-ng e la conseguente perdita di alcune entry, che sono poi state recuperate sul server Radius di Padova WiFi. La macchina quindi non è stata più considerata adatta, in quanto era in grado di gestire il flusso di dati, ma con regole troppo semplici per applicarle a pacchetti così eterogenei quali quelli del'auditing.

La raccolta è stata quindi affidata ad una macchina della quale gli amministratori non hanno password (così da ottenere l'immodificabilità dei log da parte loro), è stata scelta la macchina fisica 'dlt', presente sia sulla rete interna che sulla Dmz, in caso di necessità di manutenzione essa verrà eseguita da un amministratore sotto la supervisione di un responsabile. La caratteristica di presenza in entrambe le reti è molto comoda in quanto si evita di aprire porte su Clavister, altrimenti sarebbe necessario aprire un canale di comunicazione dalla DMZ verso la rete interna per inoltrare i log dei vari server, oppure si potrebbe far transitare ogni log attraverso una macchina sola, col difetto di rallentarla se si modificano tutti gli header dei pacchetti, oppure di 'sporcare' i pacchetti, rendendo disordinata la loro organizzazione finale (si va a perdere l'IP di provenienza).

Il sistema utilizzato è syslog-ng, ogni file creato identifica la coppia di dati `giorno:ipmacchina`, mentre la struttura di cartelle che si viene a creare avrà la seguente struttura: Mese/Giorno/Facility/Macchina.annomesegiorno. L'anno è stato omesso dalla struttura di cartelle, in quanto la legge chiede che i log vengano mantenuti per un minimo di 6 mesi, quindi ogni 9 mesi circa verranno eliminati. La facility usata per indicare i Database Linux è: local1, quella per le autenticazioni Linux è auth (o authpriv) mentre per Database presenti sulle macchine Windows e per l'autenticazione sulle macchine stesse è stata usata la facility local0

8.2.3 Crontab

Il crontab della macchina è stato configurato per eseguire una serie di funzioni di routine, tra queste le principali sono:

1. Compressione dei logs
2. Calcolo degli md5 giornalieri
3. Invio della mail automatica al titolare dell'azienda
4. Calcolo dell'MD5 mensile di tutti i logs
5. Masterizzazione mensile dei logs su DVD/CD
6. invio della situazione della macchina agli amministratori, per eventuale manutenzione
7. Mail Annuale di alert per cambiare CD/DVD

8.2.4 Sistema di archiviazione

Il sistema di archiviazione, come tutti gli altri script chiamati in automatico dal sistema, avviene tramite una riga del crontab. In particolare, il comando utile alla compressione dei logs, viene eseguito tramite lo script /opt/log_zip.sh, avviato ogni notte alle 1.00, il compito di questo script è quello di ricercare tutti i logs corrispondenti alla giornata appena trascorsa e di comprimerli utilizzando l'algoritmo GZip.

Mensilmente un altro script effettua una ricerca generale in tutta la cartella dedicata all'auditing, per comprimere tutti i logfiles eventualmente rimasti in formato RAW e comprimere tutti i file contenenti gli md5, che non vengono compressi dallo script giornaliero. La compressione è effettuata ad opera del comando :

Blocco di codice 8.14

```
find  /root/skalog/‘/bin/date --date=’1 days ago’
    +%m/%d/‘ -depth -type f -exec gzip {} \;
```

In realtà nello script sono indicate altre operazioni che, essendo di routine, non è necessario esplicitare. Va ricordato di riportare i permessi all'utente associato ad Apache invece che a 'root'. Per lo script mensile il comando è lo stesso, si elimina solo il parametro %d/ dallo script.

8.2.5 Calcolo MD5 giornalieri/mensile

L'esecuzione di questo script : log_dailymd5.sh , avviene ogni notte dopo la compressione ad opera dello script precedentemente descritto, questa scelta è stata fatta per permettere una verifica immediata dell'integrità del file, se venisse calcolato prima della compessione, per verificare la validità dei dati, bisognerebbe, ad ogni verifica, estrarre tutti i file compressi, calcolarne il Digest ed infine rimuovere i file temporanei creati.

Il compito dello script è di selezionare e successivamente calcolare il digest di ogni logfiles prodotto nella giornata appena trascorsa, i risultati verranno scritti su di un files col formato [digest path_del_file] per ogni riga, il comando principale dello script è il seguente :

Blocco di codice 8.15

```
find  /root/skalog/‘/bin/date --date=’1 days ago’
  +%m/%d/‘ -depth -type f -exec md5sum {} \;
  > /root/skalog/‘/bin/date --date=’1 days ago’
  +%m/%d/MD5/%Y%m%d‘.md5
```

I file creati serviranno all'utilizzatore del sistema per verificare il corretto funzionamento del sistema e l'integrità dei logfiles.

Una volta al mese viene invece eseguito lo script log_monthlymd5.sh, il quale compito è di calcolare l'md5 di ogni file creato dallo script log_dailymd5.sh nel mese appena trascorso (il 01/01/2009 lo script calcolerà l'md5 di tutti i files presenti nelle cartelle e sottocartelle /root/skalog/12/), nell'ordine del cronjob, questo script, verrà avviato per ultimo, alle 03.00 di notte,così da includere anche l'md5 dell'ultimo giorno del mese.

8.2.6 Mail system

Per poter effettuare verifiche rapide da parte del titolare dell'azienda è stato implementato un sistema di mailing automatico il quale ogni notte invia gli md5 calcolati.

Questo sistema è stato creato per vari motivi:

1. il primo è che rappresenta una sorta di Heartbeat del sistema, ovvero un segno di vita, infatti, essendo il sistema creato indipendente e su di una macchina isolata, nessuno potrebbe accorgersi di un'eventuale malfunzionamento o addirittura dello spegnimento della macchina, invece ci si può

facilmente rendere conto della mancanza la mail giornaliera, che dovrebbe allertare gli amministratori per provvedere al ripristino della macchina;

2. altra motivazione fondamentale è il controllo dell'integrità dei logs: sia i log presenti sul cd (coi relativi MD5) che gli MD5 arrivati sulla casella di posta, sono immodificabili, è sufficiente implentare un controllo rapido da effettuare tra i digest della mail e quelli memorizzati sulla macchina, in caso di problemi sarà facile sostiuire i logs corrotti con quelli presenti sul supporto ottico;

3. infine funge da promemoria al titolare dell'azienda epr effettuare il controllo dell'operato degli amministratori.

Il sistema di mailing utilizzato è sendmail, un programma leggero che non richiede interfacce e quindi è adatto all'utilizzo in automatico negli script o in pagine internet, il comando che, nello script giornaliero, invia le e-mail è il seguente :

Blocco di codice 8.16

```
echo "Subject: Mail automatica-Auditing" |
cat - /root/skalog/dailyheader -
  /root/skalog/‘/bin/date --date=’1 days ago’
  +%m/%d/MD5/%Y%m%d‘.md5* | /usr/lib/sendmail
  -F auditing@dominio.net -t ##@dominio.net
```

L'indirizzo di destinazione è stato oscurato per motivi di privacy. Il soggetto è stato inserito per permettere un automatismo per la classificazione delle e-mail in arrivo sulla casella di posta, il corpo del testo è presente nel file dailyheader, le righe sopra descritte sono un comando unico, costruito grazie alle pipe di Linux.

8.2.7 CD Mensile

Le richieste del Garante prevedevano un tipo di conservazione dei dati molto sicura, non ottenibile in altro modo che :

1. Outsourcing, con invio dei dati in tempo reale verso altre aziende,

2. Marche temporali Infocert, che garantiscono la genuinità dei file (ma se il file viene modificato risulta impossibile recuperarlo ,quindi risolve solo parzialmente gil obblighi legali), con l'aggiunta di NAS con permessi particolari.

3. Scrittura su supporto ottico, in modo automatico, possibilmente con scrittura giornaliera.

Chiaramente la scelta è ricaduta sul CD automatico, in quanto i costi delle altre due soluzioni erano estremamente elevati a confronto (la nostra scelta è quasi costo 0), a questo punto è stato necessario installare un masterizzatore sul Server di raccolta. Una volta installato ho effettuato una ricerca su tutti i prodotti Linux in grado di scrivere multisession su supporto ottico, la scelta era ristretta ai soli prodotti che non richiedessero interazione con l'utente, in quanto sono utilizzati da soli scripts, inoltre, agli amministratori, vengono inviate e-mail automatiche sul rapporto che il software di burning produce, così da capire se tutto è andato a buon fine o meno (se un file non è raggiungibile, o il cd haun graffio non si potrà terminare con successo la masterizzazione, ed il programma riporterà l'errore).

Il software scelto per questo scopo è *growisofs*, è un modulo delle cd-utils di Linux, che fornisce la possibilità di creare immagini ISO da scrivere successivamente, scrivere CD o DVD di vario formato e specificando un'ampia varietà di parametri, la scelta migliore è, nel nostro caso, di far rilevare a growisofs il filesystem del CD prima di effettuare la scrittura, senza imporne uno di default, così nel caso venga erroneamente inserito un CD UDF o ISO9001 non vengono creati errori.

La scrittura deve avvenire sempre in multisessione, ed inviando le mail sopra descritte agli amministratori, preferibilmente è meglio inseire un CD con sopra già scritto un file README, indicante nuove indicazioni o aggiornamenti della struttura della rete aziendale. Ogni primo gennaio un secondo script, dopo l'esecuzione del primo, effettua la chiusura del CD e l'eject, invia inoltre una serie di e-mail di alert agli amministratori, con le indicazioni necessarie per effettuare l'inserimento del CD per l'anno successivo e lo stoccaggio del CD appena chiuso. Da ricordare il fatto che, a causa di un bug di cd-utils, dopo ogni scrittura la periferica viene smontata, quindi prima di ogni scrittura è da effettuare il mount :

Blocco di codice 8.17

```
# Montaggio CD, logging stato dischi
mount /dev/scd0 /media/cdrom
df -h -T > /var/log/stato.log
# Specifico velocità, periferica, tipo e filesystem
growisofs -speed=1 -M /dev/scd0 -R -J -graft-points
```

```
# Sorgente
/root/skalog/‘/bin/date --date=’1 days ago’ +%m/‘
# Destinazione
=/root/skalog/‘/bin/date --date=’1 days ago’ +%m/‘
# Logging
1>/var/log/growout.log 2>/var/log/growerr.log
> /var/log/growisofs.log
# Mailing
echo "Subject: Messaggi_CD_Burning_auditing_e stato dischi" |
  cat  - /root/skalog/monthlyheader - /var/log/stato.log -
  /root/skalog/errheader - /var/log/growerr.log -
  /root/skalog/outheader - /var/log/growout.log |
  /usr/lib/sendmail -F auditing@dominio.net -t #@dominio.net
```

Inizialmente il rate di scrittura doveva essere giornaliero, ma dopo una breve ricerca su Internet ho scoperto che il numero massimo di sessioni scrivibili (MAX TRACK NUMBER) su di un CD nuova generazione è 99 (altrimenti molte meno), mentre su di un DVD è 124 se DVD+R 156 se DVD-R, in ogni caso insufficienti al mio scopo, quindi era stato pensata la soluzione del burning settimanale, ipotesi scartata per due motivazioni:

1. la determinazione del giorno corretto per l'eject a fine anno era di discreta complessità, nulla di impossibile, ma ogni cosa complessa aumenta la probabilità di errori, sia nella progettazione che eventuali modifiche fatte da altri in futuro, c'era il rischio che il CD uscisse prima del termine dell'anno o più tardi del dovuto;

2. l'aumento del numero di tracce potrebbe creare problemi di lettura in sistemi datati.

Segue parte dello script che effetuta la chiusura e l'eject del DVD a fine anno, per brevità ometto anche qua l'invio delle e-mail, essendo la struttura simile allo script mensile:

Blocco di codice 8.18

```
mount /dev/scd0 /media/cdrom
df -h -T > /var/log/stato.log
```

```
growisofs -M /dev/sr0=/dev/zero
  1>/var/log/growout.log 2>/var/log/growerr.log
eject /dev/scd0
```

8.3 Raccolta dalle macchine

Il sistema di raccolta è stato modificato rispetto alla proposta fatta, questo a causa di molteplici problemi riscontrati nelle macchine Linux (soprattutto di aggiornamento), problemi che, per altro, hanno causato una notevole perdita di tempo, quindi la scelta è stata quella di ricorrere alla gestione del rischio preventivata (il rischio in questo caso era andare oltre i tempi massimi previsti), anzichè collegare tutte le macchine presenti all'LDAP principale e crearene uno secondario, si è scelto il sistema più semplice per portare avanti il lavoro, ad ogni macchina esterna all'LDAP sono stati creati 4 account (uno per ogni amministratore presente in azienda) dotati di permessi di amministratore/root, ad eccetto delle macchine dove non sono stati riscontrati problemi nell'aggiunta dei pacchetti necessari all'allacciamento all'LDAP

8.3.1 Linux

Alcune macchine linux sono state collegate all'albero LDAP, quindi ogni accesso viene tracciato tramite la security attivata alla radice dell'LDAP, ovvero su pdc.dominio.net. Per tutte le altre macchine, per le quali non è stato possibile installare pacchetti o configurarli correttamente (sarebbe stato necessario aggiornarle completamente) sono stati aggiunti 4 account col diritto di effettuare il comando 'su', per passare dal proprio user a root (sarebbe stato preferibile installare su ogni macchina, dove già non fosse presente, il pacchetto 'sudo', così da poter evitare di usare la password di root per effettuare il passaggio da utente normale a root); l'invio verso il centro di raccolta avviene tramite il logger di sistema presente: syslog-ng, rsyslog oppure syslog, il compito del logger in questo caso è di filtrare i messaggi di sistema ed inoltrare via UDP all'indirizzo IP del Server di raccolta solo quelli provenienti dalla facility auth oppure authpriv.

8.3.2 Windows ed LDAP

Per le macchine Windows il problema dell'invio dei dati, dall'event log al server di raccolta, è stato risolto mediante il Client SNARE, la funzione che svolge è di leggere l'event log di sistema di Windows, filtrare gli eventi interessanti tramite il

loro Event ID ed inoltrarli al centro di raccolta apponendovi l'header di syslog, ho scelto di segnalare questi pacchetti con una facility differente rispetto a quelli di Linux, essendo i log molto differenti tra loro, sia nella forma che nella dimensione: la facility utilizzata è local0.

Le macchine da analizzare erano molte, ma la maggior parte di esse è collegate all'albero LDAP, quindi il problema non si è posto per quelle macchine. Al contrario , per i server che non vanno collegati al dominio per ragioni di sicurezza e per evitare accessi indesiderati, si è intrapresa la stessa strada delle macchine Linux, creare 4 account nominali ed impostare l'invio da ogni macchina, impostando lo Snare presente volta per volta, in base alle necessità della macchina (se è presente Oracle, Postgresql o SQL Server sono necessari Event ID differenti). Sul PDC è stata attivata la security di base, attivando quindi il logging degli accessi di ogni utente presente in azienda, anche se nel nostro caso verranno filtrati questi dati e memorizzati unicamente gli accessi relativi agli amministratori.

Gli Event Id di windows associati alle sessioni (inzio e fine) per le macchine normali sono :
515, 528, 529, 538, 540, 552, 576, 577, 646, 672, 673, 674, 675, 680, 6006, 6005, 5805.

Mentre per monitorare i login effettuati tramite l'LDAP è necessario impostare, sul PDC, anche gli eventi :
681, 682, 683, 1009, 1016, 1013, 1029.

8.4 Raccolta dai Dababase

Gli accessi ai Database, col ruolo di amministratore, sono stati loggati utilizzando sistemi built-in dei DBMS, come per i Sistemi Operativi sono state create le user personali (una per amministratore), la facility scelta per i log provenienti da macchine linux è local1, mentre per le macchine Windows non è stato possibile differenziare l'invio dei log per facility, sarebbe stato possibile creare dei meccanismi di selezione cartella filtrando più approfonditamente i pacchetti in arrivo, ma ho scelto di scartare questa ipotesi di lavoro a causa di un crash del Syslog avvenuto nel momento in cui il volume dei pacchetti in arrivo è aumentato assieme alle regole per la loro memorizzazione, per rimediare ho sempilficato le regole e così non si sono più verificati problemi, quindi ho ridotto più possibile i filtri e le regole in generale, rimandando così questo filtraggio al momento della lettura dei log (alla suite 'auditing').

8.4.1 Mysql

Mysql è presente unicamente su macchine Linux, l'attivazione del logging è relativamente semplice, è sufficiente decommentare la riga relativa al logging sul file di configurazione mysql/my.cnf, e successivamente riavviare il daemon di mysql.

Mysql, sfortunatamente, non ha la capacità di inviare i log direttamente al logger di sistema, ma unicamente su file, si è creata la necessità di fornire al logger di sistema ogni riga utile (connect e disconnect) che mysql scriveva sul proprio logfile; ho quindi studiato il funzionamento delle pipe in Linux ed ho creato un daemon che ha il compito di leggere il file di log e di inoltrare ogni evoluzione di questo file ad un `egrep` che seleziona unicamente le righe relative ad inizio o fine sessione di un utente all'interno del DB, queste righe vengono poi inoltrate al logger di sistema utilizzando come facility:local1 e tag identificativo dell'applicazione:mysql. Il daemon creato è stato inserito nel runlevel di default, e all'avvio del sistema, per assicurarsi che anche in caso di riavvio il servizio parta, e che, anche se viene fermato in qualche modo, di default Linux lo riavvia.
Riporto qui lo script che effettua tale controllo :

Blocco di codice 8.19

```
tail -f /tmp/mysql | egrep 'Connect|Quit' |
        logger -p local1.info -t mysql &
```

Lo script funziona grazie alle Pipe di Linux, con le quali è possibile creare sequenze di comandi, il primo passa il proprio output al secondo, e così via fino all'ultimo, ad ogni passaggio è possibile creare, leggere o modificare file, inviare Mail o effettuare altri comandi, nel nostro caso si legge, si filtra e si passano i dati al loggeri di sistema. Il logger riceve questi messaggi e li inoltra a sua volta al centro di raccolta, indicando propriamente l'header del pacchetto.

8.4.2 PostgreSQL

É stato attivato il logging di base, che invia le informazioni relative ad accessi e disconnessioni al logger di sistema (sia in Linux che in Windows), da qua Linux legge il messaggio e lo inoltra al centro di raccolta attraverso le corrette inmpostazioni di Syslog mentre, in windows, SNARE ha impostato l'id relativo agli eventi Postgres, oltre a questo, in entrambi i sistemi di logging sono stati impostati termini di ricerca : Connect o Disconnect, così da inviare solo dati effettivamente utili e non inondare la rete con ogni query eseguita, soprattutto perchè spesso queste sono molto voluminose.

8.4.3 SQL Server

Sui SQL Server, presenti unicamente su macchine Windows, è stata configurata la security di base, attivando il logging di tutti gli accessi effettuati, inoltre sono state genereate le credenziali che rendono identificabile l'accesso al DBMS. Su SNARE, preventivamente installato sulla macchina per il logging della security, sono stati impostati gli eventi corretti nel filtro e l'inoltro verso il centro di raccolta. Il riavvio dei SQLServer è stato pianificato, infatti questi Database sono collegati a servizi che devono essere forniti 24 ore su 24, l'interruzione di questi, anche se solo per brevi periodi e senza una valida motivazione, avrebbe causato problemi all'azienda, è stato quindi pianificato un riavvio durante il cambio sede che a breve dovrà avvenire.

Snare rileva questi dati nell'Eventlog di sistema, in particolare nella sezione riservata alle applicazioni, da là è stato impostato il filtraggio di tutti gli eventi riguardanti SQL Server, per rilevare eventi provenienti dai Database è sufficiente creare una regola su Snare che monitori unicamente il modulo Application dell'EventLog.

L'event ID associato a SQL Server è il 17055, con match del codice di login: 18453 o di logout: 18454.

8.4.4 Oracle

Per i database Oracle, presenti in macchine Windows, sono stati impostati, 4 account di accesso e con le relative policy di auditing necessarie, gli acount sono stati dotati di privilegi amministrativi, agendo quindi come l'utenza Administrator, ma associabile ad una persona fisica. L'istanza di Snare installata sulla macchina é stata configurata per effettuare l'invio degli eventi Oracle verso lo Zabbix.

Snare comunque rileva gli accessi degli amministratori grazie alla peculiarità della security dei Databse Oracle, infatti questi, di default, effettuano il logging di ogni accesso effettuato con credenziali da 'Super User', includendo nell'insieme amministratori e gestori del database, cioè chiunque abbia la possiblità di effettuare operazioni di modifica strutturale o addirittura eliminare tabelle, non vengono invece rilevati gli accessi che hanno le credenziali per effettuare unicamente caricamento dati o visualizzazione.
L'event ID associato ai login effettuati nei database Oracle è il 34.

8.4.5 Exchange

L'attivazione della policy di sicurezza su Exchange è da effettuare utilizzando gli strumenti messi a disposizione da Windows : in modo grafico In particolare si effettua con la seguente sequenza di comandi : Gestore sistema Exchange

1. Gruppi amministrativi
2. servere
3. NomeServer
4. tasto dx sul serverdi interesse
5. Proprietà
6. registrazione Diagnostica
7. MSExchangeIS
8. private o cassetta postale
9. Accessi = minima; Controllo accessi = minima

La selezione degli eventi corretti è avvenuta analizzando l'event Log di Windows ed impostandoli nell'istanza di Snare presente.

8.5 Lettore log

Una volta sviluppata e consolidata la raccolta dei dati ci si trova sempre di fronte alla necessità di fornire un sistema per l'analisi o anche la sola visualizzazione delle informazioni, possibilmente si desidera un sistema in grado di estrapolarne variabili utili e sul quale si possano definire statistiche.

Le soluzioni di mercato che sono state analizzate non erano altro che configurazioni particolari di programmi già implementati, di uso commerciale, che solitamente forniscono supporto per dati parecchio più complessi, come analisi di mercato, business intelligence ecc.., forniscono quindi funzionalità in eccesso rispetto alle richieste dettate della legge, ed un prezzo non indifferente.

La soluzione implementata si basa sulla piattaforma di Log_mining descritta nella prima parte della presente tesi, questa scelta era quasi obbligata, il riutilizzo del codice è una 'regola non scritta' della programmazione, il progetto precedente effettuava parte delle funzionalità di analisi qua descritte, forniva un'interfaccia

grafica e l'inserimento all'interno del database, partendo da log puramente testuali. La piattaforma creata in questa sezione è differente rispetto alla più 'anziana', più prestante in alcune parti, meno in altre, le modifiche più rilevanti in senso logico sono le seguenti :

- il codice è stato ottimizzato eliminando cicli e ridondanze non necessarie,
- codice semplificato in molte parti che erano eccessivamente complesse per la mole dati da analizzare,
- sono stati eliminati gli script Perl per la selezione della forchetta temporale, essendo di poca utilità quando si lavora su poche centinaia di righe,
- sono stati utilizzati più indici all'interno del Database,
- sono state i aggiunti script ulteriori per migliorare la costruzione automatica delle query,
- è stato eliminato il modulo per la configurazione.

Comunque, essendo la base di partenza modulare ed espandibile, si è riusciti ad ottenere dei risultati molto in fretta rispetto alla creazione di una soluzione ad-hoc, domostrando quindi la riutilizzabilità del software creato.

La personalizzazione chiaramente ha richiesto del tempo, ed anche le successive modifiche strutturali per ottimizzare perdendo leggermente controllo in certi aspetti e migliorare la gestione in altri al costo di una leggera perdita di perfomance, queste scelta, come già spiegato, sono dovute adun differente utilizzo della piattaorma stessa, e ad una ridotta mole di dati da analizzare.

8.5.1 Versioning

Ho scelto di dividere le due piattaforme in quanto, strutturalmente, differenti tra loro, in caso di riutilizzo futuro è comodo disporre di strumenti per analizzare dati che, in termini di quantità e contenuti, abbiano le caratteristiche che seguono:

- Pochi e Semplici, in questo caso è indifferente
- Voluminosi e semplici, una buona base di partenza è la piattaforma Log Mining
- Complessi ma meno voluminosi, in questo caso è meglio partire con la piattaforma dedicata all'Auditing

- Complessi e Voluminosi, anche se in questo caso serve una macchina adeguata e una attenta selezione dei metodi da utilizzare, prendendo parti dell'uno e parti dell'altro

Nella fase terminale dello stage mi è stato chiesto di implementare molte delle funzioni presenti nella versione Auditing anche sulla piattaforma di base, per rendere lo strumento più user friendly, quindi attualmente sono molto somiglianti per quanto riguarda le interfaccie. Sono comunque presenti tutte le versioni pubblicate, attualmente la versione della piattaforma Auditing è la 3.6, le varie versioni sono state suddivise in più sezioni, riportiamo di seguito le modifiche principali di ogni versione rilasciata :

- Nella versione 0 venivano utilizzati i sistemi derivanti dal vecchio sistema, sia a prima vista che analizzando le prestazioni ci si è resi conto che andava modificata la struttura interna dell'estrazione e della ricerca dei logs, la personalizzazione iniziale si limitava al semplice funzionamento, alla creazinoe del Database di supporto (compresa l'installazione dei vari componenti) ed alla configurazione degli script che permettono un corretto funzionamento della piattaforma nel lungo periodo.

- Verificando il funzionamenteo della versione precedente si è ritenuto opportuno passare ad una nuova versione; nella versione 1 sono state implementate le nuove funzionalità definite testando la prima versione, ottenendo maggiori controlli sui dati inseriti e migilori prestazioni a livello di inserimento dati ed estrazione di informazioni; è stato aggiunto il controllo automatico degli md5 coi file e la sezione Bluetooth, dedicata all'analisi dei logs provenienti dalle apparecchiature Net-Blue situate nelle varie mostre, fiere, ed eventi in generale, questa analisi server per fini commerciali ed analisi di mercato, stime sui possibili utilizzatori dei servizi offerti e riscontro di trend su visitatori abituali ed occasionali.

- La versione 2 è la release che aveva lo scopo di aggiungere funzionalità, è presente la creazione automatica delle query per effettuare le analisi standard dei logs, le ricerche automatizzate sono quelle solitamente più utili ed utilizzate, funzioni di aggregazione, count, condizioni sui contenuti, sugli utenti ecc..; altro punto importante è la creazinoe della pagina dedicata al check dei logs, il controllo automatico è sempre presente all'interno della pagina principale, questa seconda pagina è utile per verificare che gli MD5 presenti sul server non siano stati corrotti, si effettua il controllo inserendo

in una Text-Box le coppie Digest-Path dei logfiles che si desidera controllare, la coppia è prelevabile dalle e-mail automatiche che il sistema invia al titolare dell'azienda, che è la persona tenuta ad effettuare il controllo perdiodico sull'operato degli amministratori.

- Nella versione 3, l'ultima prodotta, sono presenti svariati grafici che rappresentano i dati caricati sul Database, una serie di controlli e di script che ottimizzano la gestione dei dati ed un maggior numero di controlli.

Future versioni Ulteriori espansioni prevedono il controllo imposabile direttamente da CD, e l'eventuale sostituzione dei logfiles corrotti direttamente da pagina web (ottimizzando ed automatizzando l'upload che nella sezione Bluetooth ho implementato), migliorare la grafica e, se possibile, la velocità di esecuzione.

8.5.2 Auditing

La sezione Auditing è la parte principale della nuova piattaforma, il suo ruolo è di permettere al titolare dell'azienda di fruire delle informazioni raccolte in maniera chiara e sicura, senza che venga data la possibilità di creare danni reali ai dati in analisi ma consentendo di ottenere unicamente le informazioni interessanti da ogni log.

Inserimento Dati

La prima schermata permette l'inserimento del lasso temporale di interesse, e di un indirizzo IP, senza molta fatica si potrebbe implementare anche una data ed una durata, la funzione *playtime* è già presente e funzionante, ma si è scelto di far specificare le date esplicitamente per far sì che l'utilizzatore del sistema abbia la reale percezione dei giorni che andrà ad analizzare. Il lasso temporale permette di specificare Date ed ore di interesse, o unicamente le date, l'ora non sempre è interessante, almeno in fase di estrazione dei dati, piuttosto è comoda per effettuare l'analisi dei dati e ricerche più precise. L'IP si può specificare anche solo parzialmente, in modo da permettere agli utenti di indicare anche solo la parte iniziale dell'IP, ottenendo il risultato di caricare sul Database tutti gli IP della subnet indicata, o degli indirizzi simili, infatti se viene inserita solo la prima porzione di un IP, il sistema caricherà sul DB tutti i dati riguardanti macchine con IP che iniziano come indicato, se invece si desidera indicare un IP singolo è sufficiente terminiare la stringa con un '.' (questo per evitare che, indicando l'IP x.x.x.2 vengano selezionati, ad esempio, anche gli ip x.x.x.21, x.x.x.206 ecc..).

Fase di Querying

Selezionati i log di interesse, questi vengono caricati sul Database, dopo aver passato il Test sull'integrità. Il test sull'integrità effettua un parsing tra i digest contenuti nei vari files *.md5 giornalieri ed un calcolo effettuato al volo su tutti i files di interesse, il risultato che si ottiene se tutto è andato bene è unicamente una stringa che conferma la veridicità dei logs, altrimenti compare un elenco dei file che sono stati identificati come corrotti. Sarà cura del titolare dell'azienda far sostituire questi logs da un amministratore sotto la propria supervisione, i logs originali vanno prelevati dal CD prodotto mensilmente.

Una volta caricati i dati sul Database, e verificata la loro integrità, si può procedere con l'analisi e la ricerca delle informazioni utili; viene proposta la possibilità di analizzarli per estrarne informazioni interessanti, verificare la correttezza delle username ed eliminare entry inutilizzabili che il logging talvolta inserisce, utili come dato da mantenere nel sistema per capire eventuali danni, ma non per verificare l'operato degli amministratori. Questi controlli ed estrazione diinformazioni avvengono grazie ad una serie di Regular Expression e CASE inseriti nei comandi SQL.

Qui viene proposta la soluzione per 'Power User', ovvero scrivere la propria query manualmente, in modo da ottenere esattamenete i risultati desiderati, oppure la soluzione automatizzata, quindi inserire dei dati e costruire automaticamente la query. La costruzione automatica delle query avviene attraverso delle funzioni in Javascript, ho scelto di cotruire pagine che lavorino sia lato server che lato client per alleggerire il lavoro svolto dal Server. Per effettuare ricerche è possibile:

1. restringere il lasso temporale,
2. scegliere i campi che si desiderano visuliazzare,
3. gli eventuali campi sui quali effettuare il raggruppamento,
4. selezionare il campo i base al quale ordinare i risultati (sempre che faccia parte della selezione dei campi visulizzati),
5. indicare un elenco di parole da ricercare,
6. indicare un elenco di IP di host da analizzare (che siano nell'elenco definito durante il caricamento),

7. indicare se effettuare il conteggio dei risultati in caso di grouping o di nessun campo selezionato,

8. limitare la visualizzazione dei risultati.

In particolare il campo relativo alle parole da ricercare ha implementate delle funzionalità avanzate, sono inseribili caratteri speciali per effettuare ricerche più particolareggiate, seguono le tre regole implementate :

- Parole senza caratteri speciali : vengono ricercate tutte, quindi ogni parola inserita senza l'aggiunta di modificatori sarà presente nei risultati estratti,

- Parole preceduta dal carattere '!' : Rientrerà nell'elenco di parole da non visualizzare nei risultati, nessuna delle entry contenenti anche solo una delle parole inserite con il ! inizialmente verrà selezionata,

- Parole preceduta dal carattere '?' : rappresentano un elenco di parole che vengono cercate in maniera opzionale, almeno una delle parole di questo elenco dovrà essere contenuta in ogni entry risultante.

Al termine dell'inserimento verrà chiesto all'utente se la query creata è corretta e se si desidera utilizzarla veramente, andando a perdere quella precedentemente inserita.

Grafici

La sezione contenente i grafici ottiene i dati che rappresenta dai logs caricati precedentemente dal modulo auditing, prima di procedere alla visulizzazione dei grafici è necessario assegnare le username analizzando i logs, operazione automatica ma, se confrontata al caricamento dei dati, di durata considerevole.

Questi dati vengono selezionati attraverso varie query, dove vengono raggruppati per ora (selezionando solo parte del campo datetime) e suddivisi per i vari utenti da controllare. Da questi dai vengono visualizzati vari grafici, uno per ogni elemento dello staff più un grafico che include tutti i login avvenuti. Questo sistema grafico non permette la selezione dei dati, vengono presi tutti i login dei logs caricati sul Database, questa scelta è stata fatta per evidenziare eventuali login fuori orario, che dovrebbero allertare il possessore delle credenziali, se non altro ad effettuare accurate analisi riguardo ai login anomali.

8.5.3 Check

La pagina relativa al check implementa una serie di controlli per verificare la validità dei logs presenti sul Server. Inizialmente viene presentata una pagina con nua grande TextBox, nella quale vanno inserite le coppie di dati digest:path relative ai file da verificare, per semplificare il compito, i dati sono da inserire così come sono sulle e-mail inviate, senza richiedere nessuna modifica; il controllo che viene effettuato è sequenziale, innanzitutto vengono verificati i logs con gli md5 presenti sul Server, è lo stesso controllo che avviene durante l'esecuzione di ogni caricamento dati sul Database, il risultato può essere :

- Stringa che afferma correttezza dei logs
- elenco dei log corrotti

In sequenza viene effettuato il controllo dei dati inseriti nella Text-Box con gli md5 corrispndenti presenti nel server, la verifica avviene aggiornando i dati caricati nella fase precedente sul Database. Il risultato anche qua può consistere in un elenco di file considerati corrotti o una stringa che afferma che il controllo è l'avvenuto con successo. Infine al momento della sostituzione dei dati vengono verificate le eventuali differenze con gli md5 calcolati al volo dal sistema, così da poter presentare all'utente un quadro completo degli eventuali errori e non solo una lista di file corrotti.

8.5.4 Bluetooth

Grazie al lavoro svolto in precedenza nel modularizzare il progetto Log Mining, sono stato in grado di creare il Log Analyser per l'auditing in tempi molto brevi, avanzando qualche ora libera rispetto al progetto pianificato, nel quale prevedevo di terminare questo lavoro un giorno dopo di quanto in realtà ho terminato.

Durante il tempo 'libero' che mi si è presentato, mi sono interessato ad altri problemi presenti, proponendomi di risolverne uno in particolare che si adattasse al lavoro svolto da me fino ad allora; alla fine ho impiegato le ore che avevo a disposizione per creare un modulo ad-hoc utile ad un'altro progetto che si sta sviluppando in azienda.

Degli apparati Bluetooth sono stati installati in alcuni luoghi pubblici, fiere, mostre d'arte, ecc.., questi apparati registrano dei logs che, di per sè, non forniscono informazioni leggibili, ma, analizzandoli ed estraendone solo le sezioni interessanti, si ricavano dati riguardanti il pubblico, lo stato degli apparati stessi,

e sono rilevabili trend che si verificano negli eventi dove gli apparecchi (Net-Blue) vengono installati.

Il modulo propone un'interfaccia unica per la gestione dei logs, il loro caricamento sul server e la selezione manuale per il caricamento, non è stato effettuato un logging automatico in quanto gli apparati Net-Blue non sempre sono collegati alla rete, si è scelto di memorizzare localmente i logs e caricarli al termine dell'evento per effettuare un'analisi per preparare al meglio gli eventi futuri. Una volta selezionato e caricato il log l'interfaccia aggiunge una serie di bottoni che forniscono risultati standard, conteggi, e la solita costruzione automatica di query, inoltre, per completezza, c'è la possibilità di scrivere query manualmente per permettere agli utenti che conoscano l'SQL, di effettuare controlli personalizzati e più evoluti. E' possibile scegliere di visualizzare un grafico che fornisce i trend di affluenza all'evento, si possono visualizzare, suddivise in base all'orario :

- il totale delle persone presenti,
- il numero dei presenti che ha accettato la comunicazione bluetooth,
- il numero di quanti hanno rifiutato,
- il numero delle richieste che non hanno avuto successo.

Il modulo Graphics, utile a rappresentare in modo intuitivo i dati caricati, è stato ottenuto grazie all'utility PHPGraphLib B.10, sfruttata anche per la realizzazione del modulo Coovachilli della precedente piattaforma. Il funzionamento è collegato alla sezione relativa al caricamento, quindi i grafici visualizzati rappresentano, mediante opportune Select, i dati già caricati nel Database nella pagina Bluetooth.

In particolare una delle query utilizzate in questa sezione è la seguente:

Blocco di codice 8.20

```
SELECT HOUR(logging_time) as ora,count(*) as nr
FROM blue WHERE result rlike 'OK'
group by HOUR(logging_time)
```

Questa query raggruppa tutti i dati in base all'ora, sommando così tutti i logs, anche di giorni differenti, che sono stati memorizzati nella stessa ora, fatto ciò è semplice creare un array associativo mediante PHP e, da questo, ottenere una rappresentazione grafica del tutto.

8.6 Testing

Terminata l'implementazione è iniziata una fase di testing dell'intero progetto che ha portato allo scoperto problematiche riguardanti la stesura del codice, controlli non rigorosi a sufficienza e tempi di caricamento eccessivi. I test eseguiti spaziano dall'inserimento di dati errati, la ricerca di collegamenti tra login e la ricerca di intrusi nella rete, dopo ogni test venivano corretti i problemi trovati fino al raggiungimento di una soluzinoe soddisfacente e corretta, l'aggiunta di funzionalità e la costruzione di query sempre più complesse. I vari problemi sono stati risolti alla seconda versione prodotta, come descritto all'inizio del capitolo, nella terza versione sono state aggiunte molte funzionalità e controlli di integrità dei dati e sulla correttezza dei form inseriti, infine sono stati aggiunti grafici che rappresentino i dati inseriti.

Capitolo 9

Futuri Sviluppi

Terminato lo stage è necessario valutare futuri sviluppi del proprio operato, da affidare all'area tecnica o ad altri stagisti. L'obiettivo certamente più vicino è la creazione della piattaforma di monitoraggio degli apparati di rete, basata su Nagios e che utilizzi la piattaforma di analisi dei logs per reperire informazioni precise circa eventuali problematiche.
Le basi per completare questo lavoro sono già presenti, infatti:

- I log utili sono già stati identificati ed associati alle problematiche che sono in grado di identificare,
- Nagios è già presente e funzionante in azienda, ma fornisce un supporto limitato e, soprattutto, non è possibile ottenere infrmazioni circa le cause o gli eventi che hanno creato eventuali problemi,
- La piattaforma di analisi dei logs è molto modulare, e, come si è visto, è applicabile in vari contesti, il punto di forza sta nell'incrociare i dati forniti.

Le operazioni fondamentali per il raggiungimento di questi obiettivi saranno quindi

- modificare l'interfaccia di Nagios integrando in essa la piattaforma di Log Analysis,
- attivare e centralizzare i logs di Nagios per i vari apparati di rete, ove possibile,
- creare plugin di Nagios ad-hoc per gli apparati ove non fosse possibile attivare il logging di Nagios,

- implementare una piattaforma che visualizzi, su megaschermo, tutte le macchine e il loro stato, cliccando su di una macchina si avrà accesso all'anlisi dei log relativi a quell'apparato.

Conclusioni

In questo documento ho descritto le attività principali che mi hanno impegnato durante il periodo di stage presso *ne-t by Telerete Nordest*.

Sono state ritratte le prime mansioni che ho svolto, da una parte allo scopo di ambientarmi nel nuovo contesto lavorativo, dall'altra e sopratutto in ottica di formazione. Durante tutto l'arco del mio tirocinio, mi sono documentato sulle varie tecnologie ed i vari sistemi con cui mi sono ritrovato a lavorare, ed ho accresciuto notevolmente il mio bagaglio di conoscenze tecniche, migliorando in maniera particolare dal punto di vista dello sviluppo di software e della correlazione tra discipline differenti.

Successivamente, siamo passati alla pianificazione del lavoro ed alla descrizione delle fasi operative, anche queste attività che mi hanno accompagnato per tutta la durata dell'esperienza.
Ho studiato in maniera particolare i sistemi di rete tra cui i server di autenticazione, i firewall, i dispositivi di switching, la struttura della rete interna/esterna dell'azienda, il funzionamento e le mancanze del progetto che mi è stato affidato, e infine la creazione di più approcci per la risoluzione di un problema.

Dopo aver delineato i problemi che ci si proponeva di risolvere, nei capitoli 3 e 6 sono passato ad esporre i progetti e i sistemi di logging che avrebbero portato alla corretta raccolta dei logs necessari per ognuno dei problemi, nei capitoli 4 e 7, per concludere, nei capitoli finali delle due sezioni, ho descritto le modalità secondo le quali ho implementato i log Analyser, rispettivamente nei capitoli 5 e 8. Queste sono le soluzioni che ho effettivamente messo in atto durante il periodo di stage. Possiamo citare il server di monitoraggio, per rispettare la legge riguardante l'auditing degli amministratori, e l'estensione di un applicativo PHP per l'estrazione di dati di interesse a partire da file di log. In modo particolare, quest'ultimo è stato applicato anche per estrarre in maniera efficiente le informazioni richieste dalle leggi per il contrasto del terrorismo internazionale, dopo che è stata effettuata una selezione dei logs utili e configurata la centralizzazione di questi logs. Queste soluzioni sono attualmente utilizzate quali strumenti di ausilio per il personale dell'area tecnica e per il monitoraggio degli accessi nelle macchine

aziendali e per evadere rapidamente eventuali richieste provenienti dalla polizia postale.

I sistemi che ho implementato durante questo Stage servono, fondamentalemente, a rispondere alla domanda CHI, DOVE e QUANDO, ma i dati che sono stati identificati possono essere utili per terminare i progetti futuri, ovvero la creazione di una piattaforma per l'analisi degli apparati di rete e dei loro logs; questo secondo obiettivo può essere raggiunto grazie alla modularità della piattaforma 'Log Analyser', che, come abbiamo visto, può essere utilizzata per fornire informazioni in situazioni molto differenti.

Appendice A

Manuali

Seguiranno le indicazioni per l'utilizzo e le nuove installazioni degli applicativi sviluppati in questo periodo di stage, la configurazione della raccolta dati verra' descritta in maniera minimale in quanto gia' esposta nei capitoli precedenti.

A.1 Utilizzo Log Mining

La piattaforma relativa alla legge anti-terrorismo fornisce diversi strumenti, innanzitutto e' necessario effetuare un login, dopodiche' si avra' accesso ai dati che la piattaforma gestisce.

A.1.1 Ip Trace

La sezione principale e' Iptrace, utilizza i logs forniti dal netfilter di sistema, ovvero iptables. Inizialmente il modulo pone l'utilizzatore di fronte ad un lasso temporale da definire e la possibilita' di specificare:

- Nessun altro dato,
- Una username,
- L'IP di un server,
- Sia l'IP che l'username.

Una volta caricati i logs desiderati all'interno del Database e' possibile iniziare ad effettuare le ricerche vere e proprie, per facilitare questo compito e' stata creata la possibilita' di creare query in modo automatico. Sono state messe a disposizione delle Text-Box da riempire coi valori relativi, in caso si desideri o meno che certe entry vengano selezionate dalla query in costruzione.

L'elenco dei risultati, nel caso un cui la richiesta sia di tipo HTTP (ovvero rivolta alla porta 80 di un web server), fornisce un link diretto alla pagina Squid, utile per fornire dettagli maggiori rispetto a quelli di Iptrace.

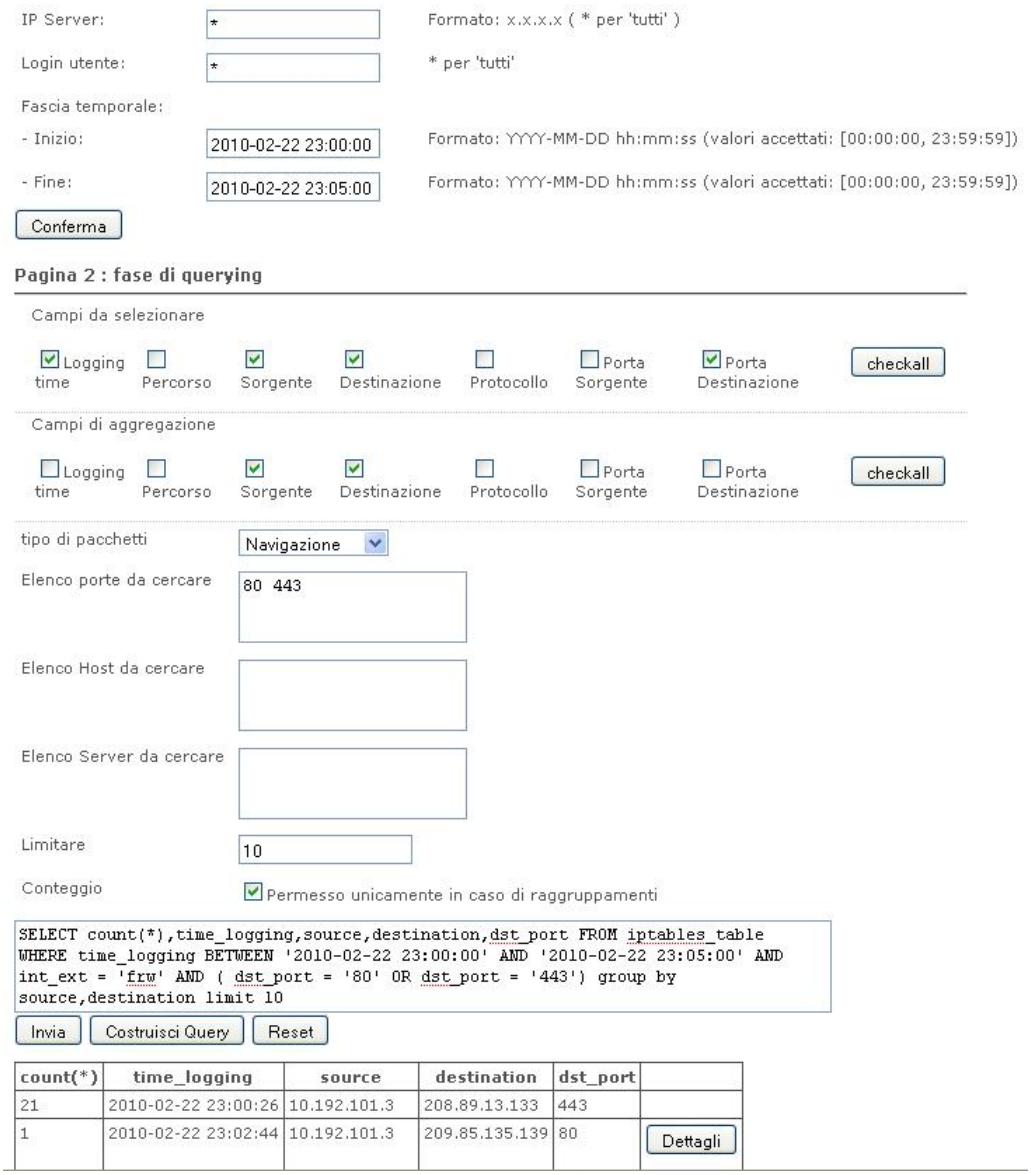


Figura A.1: Screenshot del modulo IpTrace, senza specificare user nè server

Se si desidera specificare una username per sapere quali server ha visitato si ottiene la pagina di figura A.2, sarà inoltre possibile cliccare sul bottone Dettagli per sapere all'interno del server cosa è stato visitato.

Se invece si specifica un server vengono visualizzati tutti gli utenti che l'hanno visualizzato nel lasso temporale indicato, come da figura A.3.

1 file di log selezionato per l'inserimento nel database.
All 3163 rows from /root/local/stefanolog/temp/kern.20100304_11:30:00_11:31:00.gz in database!

IP Server: 212.48.106.75 — Formato: x.x.x.x (* per 'tutti')

Login utente: * — * per 'tutti'

Fascia temporale:

- Inizio: 2010-03-04 11:30:00 — Formato: YYYY-MM-DD hh:mm:ss (valori accettati: [00:00:00, 23:59:59])

- Fine: 2010-03-04 11:31:00 — Formato: YYYY-MM-DD hh:mm:ss (valori accettati: [00:00:00, 23:59:59])

Conferma

Pagina 2 : fase di querying

Chi ha richiesto il server all'IP 212.48.106.75 nella fascia tra 2010-03-04 11:30:00 e 2010-03-04 11:31:00 ?

L'indirizzo IP 212.48.106.75 corrisponde all'host name : 212.48.106.75 .
1 utenti hanno richiesto il server all'IP 212.48.106.75 nella fascia temporale selezionata.
I dati di questi utenti sono riportati nella tabella seguente, estratta da "public.cliente" del database di FreeRadius:

id	idutentebase	nome	cognome	indirizzo	cap	citta	provincia	idnazione	datanascita	luogonascita	codicefiscale
10112	sabau	Karoly Albert	Szabo	via Monte dei santi 5	37036	San Martino Buon Albergo	VR	122	1985-07-20 00:00:00	Verona	szbkly85l20l781m

Figura A.2: Screenshot del modulo IpTrace, specificando solo user

1 file di log selezionato per l'inserimento nel database.
All 3163 rows from /root/local/stefanolog/temp/kern.20100304_11:30:00_11:31:00.gz in database!

IP Server: * — Formato: x.x.x.x (* per 'tutti')

Login utente: sabau — * per 'tutti'

Fascia temporale:

- Inizio: 2010-03-04 11:30:00 — Formato: YYYY-MM-DD hh:mm:ss (valori accettati: [00:00:00, 23:59:59])

- Fine: 2010-03-04 11:31:00 — Formato: YYYY-MM-DD hh:mm:ss (valori accettati: [00:00:00, 23:59:59])

Conferma

Pagina 2 : fase di querying

Quali Server ha visitato l'utente sabau nella fascia tra 2010-03-04 11:30:00 e 2010-03-04 11:31:00 ?

Nella fascia tra 2010-03-04 11:26:27+01 e 2010-03-04 11:30:43+01 gli stato assegnato l'indirizzo IP 10.192.8.10

IP address richiesto	Host name corrispondente	Istante della prima richiesta	Istante dell'ultima richiesta	Porta sorgente (prima richiesta)	Porta destinazione (prima richiesta)	
174.129.241.144	ec2-174-129-241-144.compute-1.amazonaws.com	2010-03-04 11:30:02	2010-03-04 11:30:29	38844	443	
209.85.135.102	mu-in-f102.1e100.net	2010-03-04 11:30:14	2010-03-04 11:30:39	42834	80	Dettagli
10.192.0.1	10.192.0.1	2010-03-04 11:30:43	2010-03-04 11:30:43	51763	3990	
209.85.135.139	mu-in-f139.1e100.net	2010-03-04 11:30:43	2010-03-04 11:30:43	40725	80	Dettagli
74.125.101.80	74.125.101.80	2010-03-04 11:30:43	2010-03-04 11:30:43	55449	80	Dettagli

numero di IP richiesti da sabau nella fascia richiesta : 14
numero di indirizzi IP assegnati a sabau nella fascia richiesta : 1

Figura A.3: Screenshot del modulo IpTrace, specificando il server

Se invece si specifica sia la user che il server verrà indicato se e quando l'utente indicato ha visitato il server, è possibile cliccare sul bottone richeiste se laporta utilizzata è la 80.

Se invece e' stata effettuata la ricerca in base a Username o IP o entrambi, verra' visualizzato l'elenco delle persone che hanno visualizzato un dato server o l'elenco dei server visitati da una certa persona, oppure, se vengono inserite entrambe le possibilita', appare l'elenco delle visite che l'utente ha, eventualmente, fatto.

Al termine delle operazioni e' consigliabile effettuare il logout, per evitare l'accesso ai dati da perte di personale non autorizzato alla visione di tali informazioni.

A.1.2 Squid

Il modulo squid, si presenta in maniera molto simile ad IPTrace, fornisce dati prelevati dai logs dell'applicazione omonima Squid. La differenza è che qua non si offre la possibilità di specificare la login e il server, solitamente, infatti, le richieste riguardanti le username vengono fatte su Iptrace e passate al modulo Squid della piattaforma, sempre che riguradino la porta 80.

Una volta caricati i logs nel database, è stata implementata la possibilità di creare automaticamente le query per ricercare dati e parole contenute negli URL da includere od omettere dai risultati.

Un tipico utilizzo del modulo è riportato in figura A.4.

Il modulo fornisce la possibilità di risalire alla username associata ad ogni richiesta, passando i dati utili alla pagina IPTrace.

A.1.3 Coovachilli

La sezione di Coovachilli sfrutta i logs dell'omonima applicazione per fornire dati relativi ai tentativi di accesso eseguiti sulla rete Padova Wifi, sottolineando tutte le righe anomale, ovvero contenenti errori e pacchetti rifiutati, un esempio di ciò che viene visualizzato è riportato in figura A.5.

Inoltre viene generato un grafico che rappresenta l'andamento della rete giornaliero ed annuale, il primo grafico indica il numero di login effettuati con successo per ogni ora della giornata, mentre il secondo raccoglie il numero di login effettuati ogni giorno per 365 giorni (terminata la finestra temporale l'entry di 366 giorni prima viene eliminata da una script bash).

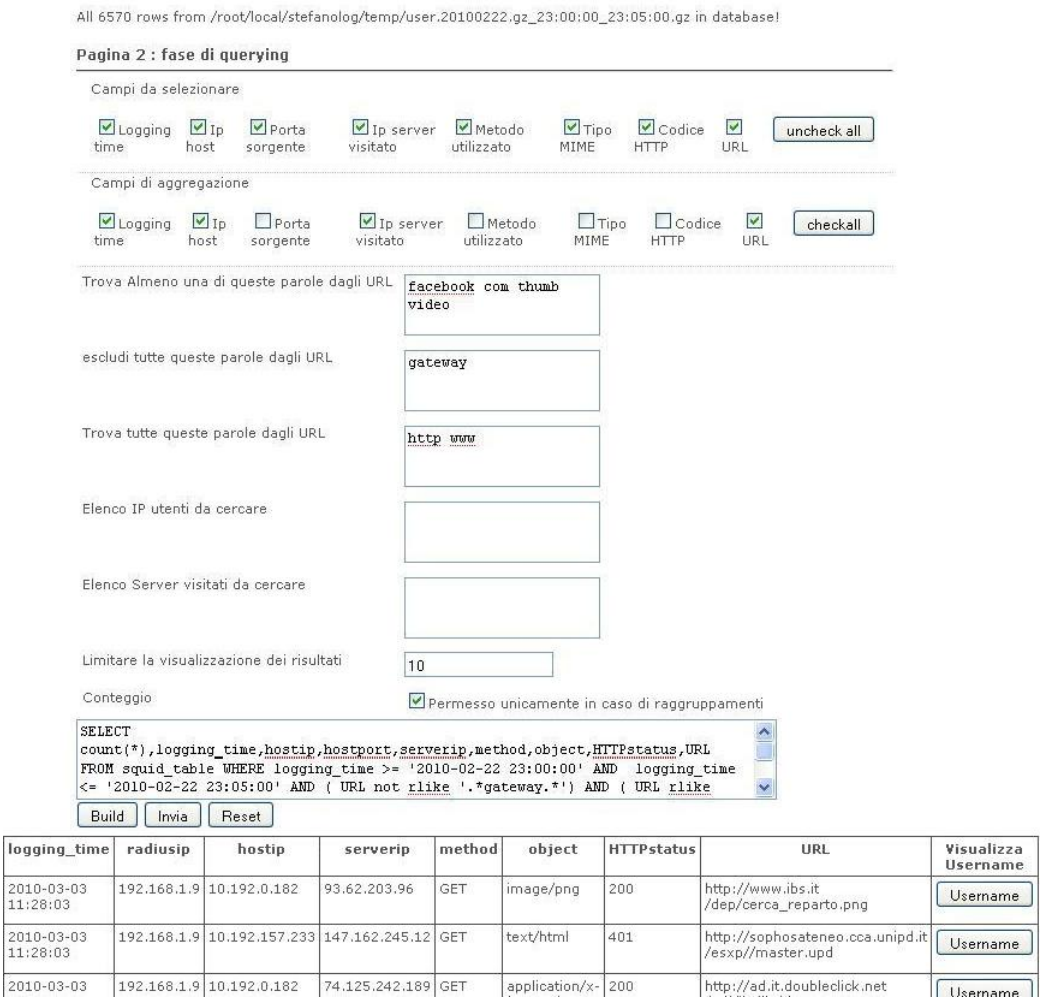
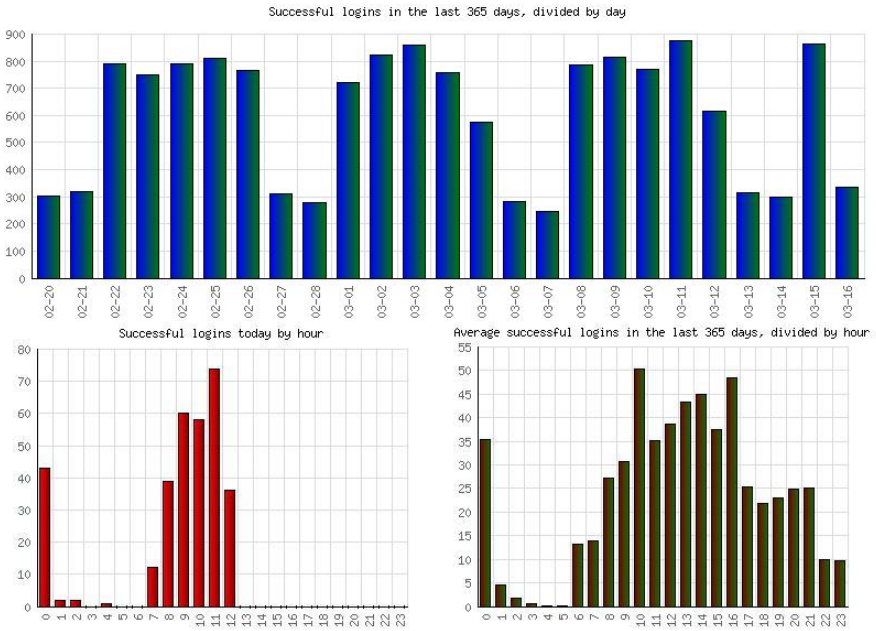


Figura A.4: Screenshot del modulo Squid




```
Anomalous messages in coovachilli log files:
Mar 16 08:05:52 192.168.1.9 coova-chilli[21659]: redir.c: 2134: Required: 'wwwdir' (in chilli.conf) and 'file' query-string param
Mar 16 10:30:40 192.168.1.9 coova-chilli[1572]: redir.c: 2134: Required: 'wwwdir' (in chilli.conf) and 'file' query-string param
Mar 16 10:31:05 192.168.1.9 coova-chilli[1592]: redir.c: 2134: Required: 'wwwdir' (in chilli.conf) and 'file' query-string param
Mar 16 10:31:45 192.168.1.9 coova-chilli[1625]: redir.c: 2134: Required: 'wwwdir' (in chilli.conf) and 'file' query-string param
Mar 16 10:32:01 192.168.1.9 coova-chilli[1637]: redir.c: 2134: Required: 'wwwdir' (in chilli.conf) and 'file' query-string param
Mar 16 10:32:15 192.168.1.9 coova-chilli[1643]: redir.c: 2134: Required: 'wwwdir' (in chilli.conf) and 'file' query-string param
Mar 16 10:32:40 192.168.1.9 coova-chilli[1660]: redir.c: 2134: Required: 'wwwdir' (in chilli.conf) and 'file' query-string param
```

Figura A.5: Screenshot del modulo Coovachilli

A.2 Nuove installazioni Log Mining

Per replicare la piattaforma il di analisi e' necessaria innanzitutto l'installazione dei seguenti pacchetti:

1. php
2. apache
3. mysql
4. interfacce tra gil applicativi (apache/php, php/mysql)

Il passo successivo consiste nel ricreare le tabelle necessarie al corretto funzionamento del login e dei vari moduli della piattaforma:

- logins : per la tabella di coovachilli
- users : per permettere il login corretto
- active_users - per mantenere traccia degli utenti correntemente sul sistema
- logged_users - per sapere chi ha effettuato il login
- squid_table : tabella utile al caricamento dei dati da parte della piattaforma, i dati caricati sono data e ora, due IP (in varchar), la porta di origine (la destinazione e' sempre la 80), l'URL visitato, il tipo MIME oggetto della comunicazione ed il codice HTTP ritornato dal server.
- iptables : tabella utile al caricamento dati provenienti da iptables, il modulo che utilizza questa tabella e' iptrace, i valori caricati sono porte ed IP (destinazione ed origine), data e ora, interfaccia attraversata e protocollo utilizzato.

Fatto cio' si prosegue l'installazione copiando semplicemente la cartella contenente il modulo log_mining nella cartella /var/www/ della macchina di destinazione. Si configurano correttamente i logs da analizzare e le cartelle di destinazione. Per farlo si puo' organizzare Syslog come per questa implementazione oppure sistemare e personalizzare la configurazione dell'applicativo.

Da tenere presente che sono da aggiornare anche il crontab e copiare gli script, eventualmente personalizzati, nella cartella /opt/.

Gli script fondamentali al funzionamento del plotting dei grafici per Coovachilli sono coovaday.sh, coovamonth.sh. Quelli utili alla gestione generale dei logs memorizzati sono log_find.sh e log_zip.sh, che utilizzano funzioni Perl per ottimizzare le operazioni.

A.3 Utilizzo Auditing

La piattaforma relativa all'auditing degli amministratori di sistema ha un funzionamento simile a quella relativa al Log Mining, essendo la base di partenza la stessa.

La differenza sostanziale consisite nel Check dei logs, che avviene ad ogni caricamento di dati su Database e la velocità di esecuzione, che grazie alla mole dati ridotta è diminuita, ma a causa della complessità di questi ultimi è aumentata in certi frangenti che ora andremo ad illustrare.

L'utilizzo di questa piattaforma, come per la precedente, inizia col login, fatto ciò si ha accesso alle varie sezioni di analisi, infine si termina col comando logout.

A.3.1 Auditing

La pagina dedicata all'analisi dei logs degli amministratori è la sezione principale di questa piattaforma, il suo compito, come spiegato precedentemente, consiste nel fornire indicazioni circa l'utilizzo delle macchine da parte degli amministratori di sistema, durante la prima fase, ovvero la selezione dei logs utili, si ha a disposizione un lasso temporale, da specificare indicando data e ora, o, se interessano tutti i dati di una giornata, è possibile indicare unicamente la data, in questo caso verranno selezionate tutte le righe relative a quella giornata. Il lasso non ha limiti di ampiezza, ma è consigliabile specificare date non troppo distanti, per evitare rallentamenti eccessivi.

L'altro paramametro specificabile è l'IP, si può omettere parte dell'IP ed indicare unicamente la parte iniziale, questo per permettere all'applicativo di caricare unicamente, ad esempio, tutti i logs di una determinata sottorete, se invece si desidera specificare un IP solo, si termina la stringa con un punto.

Una volta caricati i dati compare l'interfaccia che permette di effettuare le ricerche e fare statistiche sui dati raccolti. Le ricerche riguardanti le caratteristiche più comuni sono state automatizzate per rendere più user friendly l'applicativo, in particolare la costruzione automatica delle query riguarda :

- L'analisi dei logs per estrarre le username
- La selezione dei campi desiderati
- La selezione dei campi da utilizzare per aggregare dati
- La selezione del campo in base al quale ordinare le entry

- Una textbox per indicare le parole che vanno trovate/escluse secondo le regole :

    1. elenco parole da escludere, nessuna di questa sarà presente in nessun log visualizzato (NAND), modificatore utilizzato '!'
    2. almeno una delle parole indicate comparirà in ogni log (OR), modificatore utilizzato '?'
    3. ogni parola senza modificatori dovrà comparire in ogni log (AND), nessun modificatore utilizzato
    4. L'opzione NOR non è stata presa in considerazionein quanto fornisce risulatati poco significativi, secondo la regola : almeno una delle seguenti parole non comparirà in tutti i logs.

- Specificare un elenco di hosts da analizzare

- Porre un limite alla visualizzazione, per evitare di rallentare l'apertura della pagina

- Effettuare un conteggio, utile in caso si scela di usare campi per l'aggregazione di dati.

Un esempio di ricerca è riportato nell'immagine A.6.

Durante il caricamento dei dati, come già detto poco fa, vengono confrontati gli md5 presenti sul Server con un calcolo effettuato 'al volo' dell'md5 dei files prima che questi vengano caricati, in caso di incoerenze viene visualizzato l'elenco dei logs corrotti, in modo tale da permettere a chi sta utilizzando l'utility di sostituire tali files con quelli presenti sul supporto ottico. Alternativamente è possibile verificare la correttezza di questa verifica inserendo nella pagina Check i digest opportuni, inviati giornalmente per e-mail appositamente per questa eventualità.

A.3.2 Auditing Graphics

Il modulo relativo alla visualizzazione grafica degli accessi permette di visualizzare unicamente dati pre-caricati sul Database, in termini pratici, questo modulo è visualizzabile unicamente dopo aver effettuato il caricamentodei dati ad opera del modulo precedentemente descritto. I cinque grafici visualizzati rappresentano gli accessi, e le principali operazioni riguardanti la security, avvenute nell'arco di tempo specificato nel modulo precedente.

Figura A.6: Screenshot del modulo Auditing

I dati, per chiarezza, vengono compressi in 24 ore, anche nel caso in cui riguardino più giornate, come mostrato in figura A.7 il grafico principale conterrà un numero di login parecchio superiore a quelli dei singoli amministratori, comprendendo anche le operazioni automatiche e i login esterni.

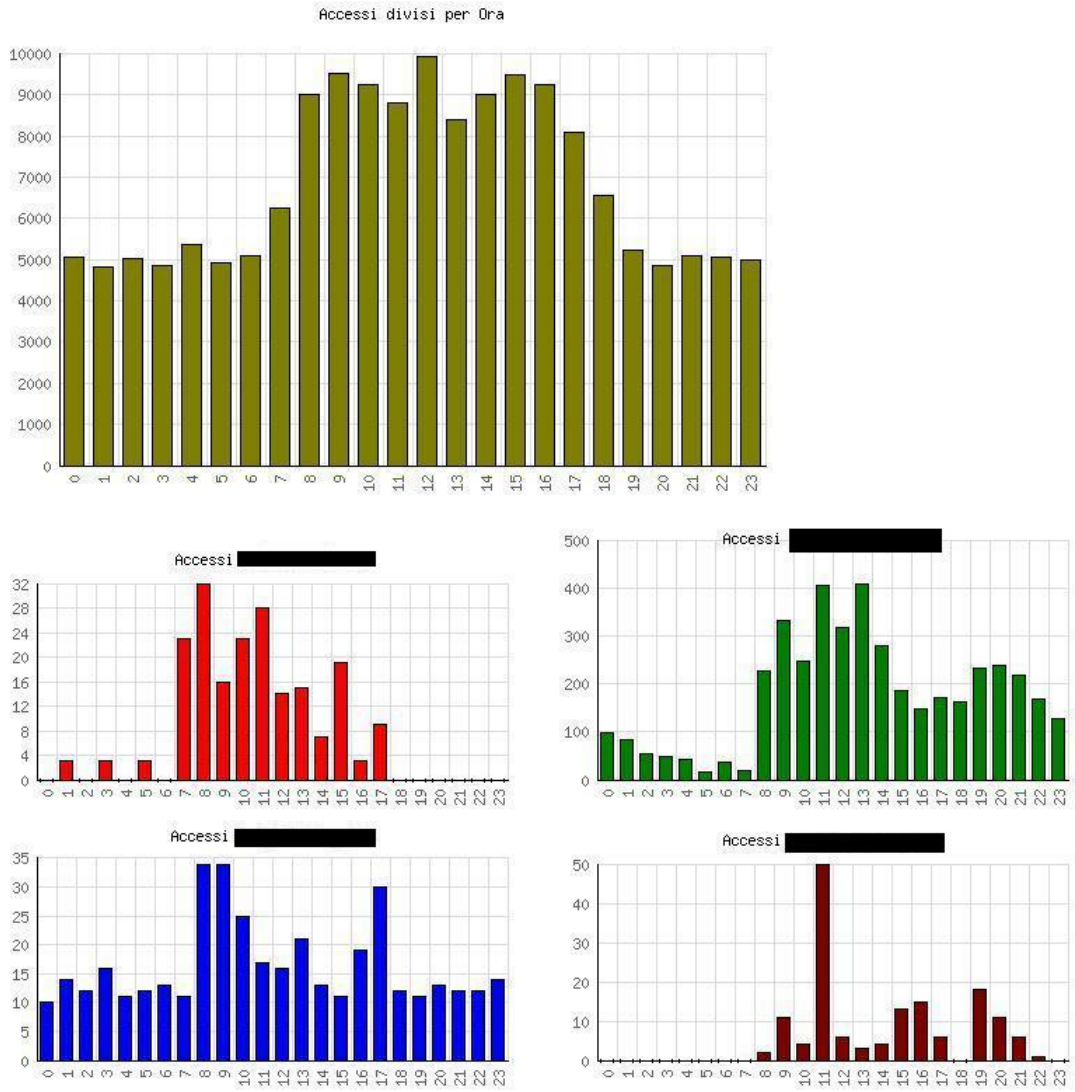


Figura A.7: Screenshot dei grafici sull'auditing

A.3.3 Check

La pagina Check dà la possibilità agli utenti di verificare rapidamente lo stato dei logs, la loro integrità e, se necessario, provvedere alla sostituzione dei logs rilevati corrotti.

Il funzionamento è relativamente semplice, è presente un unica TextBox, in questo campo vanno inserite tutte le coppie [Digest Path_del_file] così come sono nell'e-mail, sono solo da copiare i digest arrivati nella text box.

E' possibile inserire digest provenienti anche da più e-mail, andando così a coprire più giornate, o eventualmente solo alcune righe per ogni e-mail,.

Il risultato sarà un elenco delle giornate che si stanno verificando e, se vengono rilevati files corrotti, o MD5 modificati, due possibili elenchi, il primo indica se gli md5 presenti sul Server corrsipondono al calcolo effettuato 'al volo' sui file interessati; la seconda lista che, eventualmente, compare, è ottenuta dal confornto tra i digest presenti nella text-box e quelli presenti nei files contenenti gli MD5.

Di seguito riporto un esempio, A.8 di Check dei logs, nell'esempio sono presenti logs integri, md5 corretti e digest della mail modificati appositamente per il test, come si può notare ho evidenziato le modifiche da me effettuate.

Figura A.8: Screenshot della sezione Check

A.3.4 Bluetooth

La sezione relativa al Bluetooth offre tutte le sue funzionalita' in un'interfacia unica, e' possibile effettuare l'upload di files sul server, una volta caricati vengono visualizzati nell'elenco di files disponibili all'analisi.

Per l'analisi è possibile selezionare più di un file alla volta e caricarli sul Database, dopodichè comparirà la finestra nella quale inserire query di interrogazione e i bottoni utili ad ottenere informazioni quali :

- Numero di telefoni OK
- Numero di telefoni FAIL
- Numero di telefoni in totale

L'estrazione dei dati avviene identifica per ogni riga un OK/FAIL/RETRY, un MAC, un codice e, se presente, il nome 'friendly' del telefono, assegnato dal proprietario.

Anche qua è possibile effettuare le ricerche più articolate scrivendo le proprie query a mano, ma è presente la possibilità di automatizzarne la costruzione, almeno per le formule più usate, per permettere ad utenti meno esperti di utilizzare anche questo modulo. Un esempio di utilizzo del modulo è rappresentato dall'immagine A.9.

A.3.5 Bluetooth Graphics

Il modulo è dotato anche di un sistema per visualizzare in maniera grafica l'afflusso di pubblico negli eventi. Questi grafici, mostrano i dati che, nella parte principale dalla piattaforma, vengono restituiti come numeri, ovvero:

- Numero di telefoni in totale nel grafico principale
- Numero di telefoni OK, in verde
- Numero di telefoni FAIL, in rosso

Come si vede dalla figura A.10 i tre moduli indicano l'afflusso, raggruppando tutti i giorni e dividendoli unicamente per le ore, altre versioni erano state fatte ma risultavano poco leggibili, quindi è stata scelta la versione raprresentante le 24 ore.

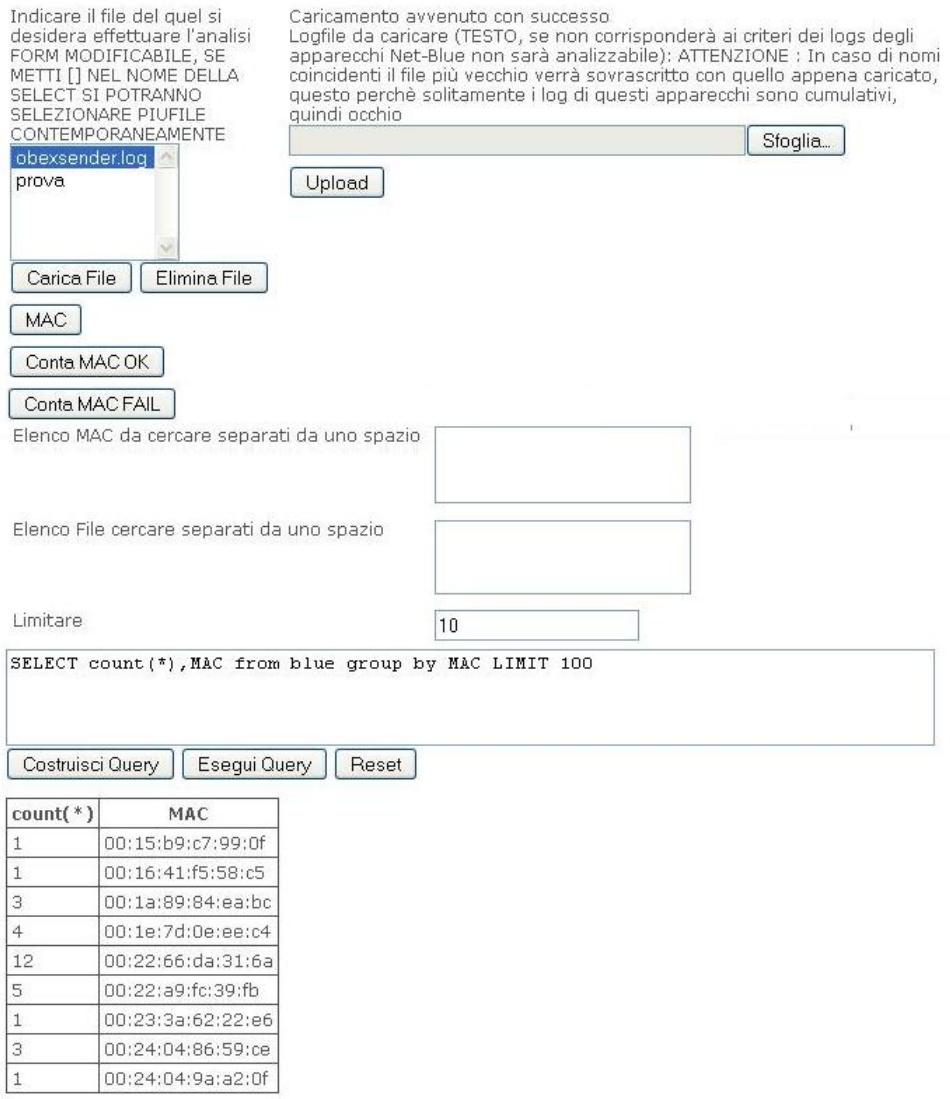


Figura A.9: Screenshot del modulo Bluetooth

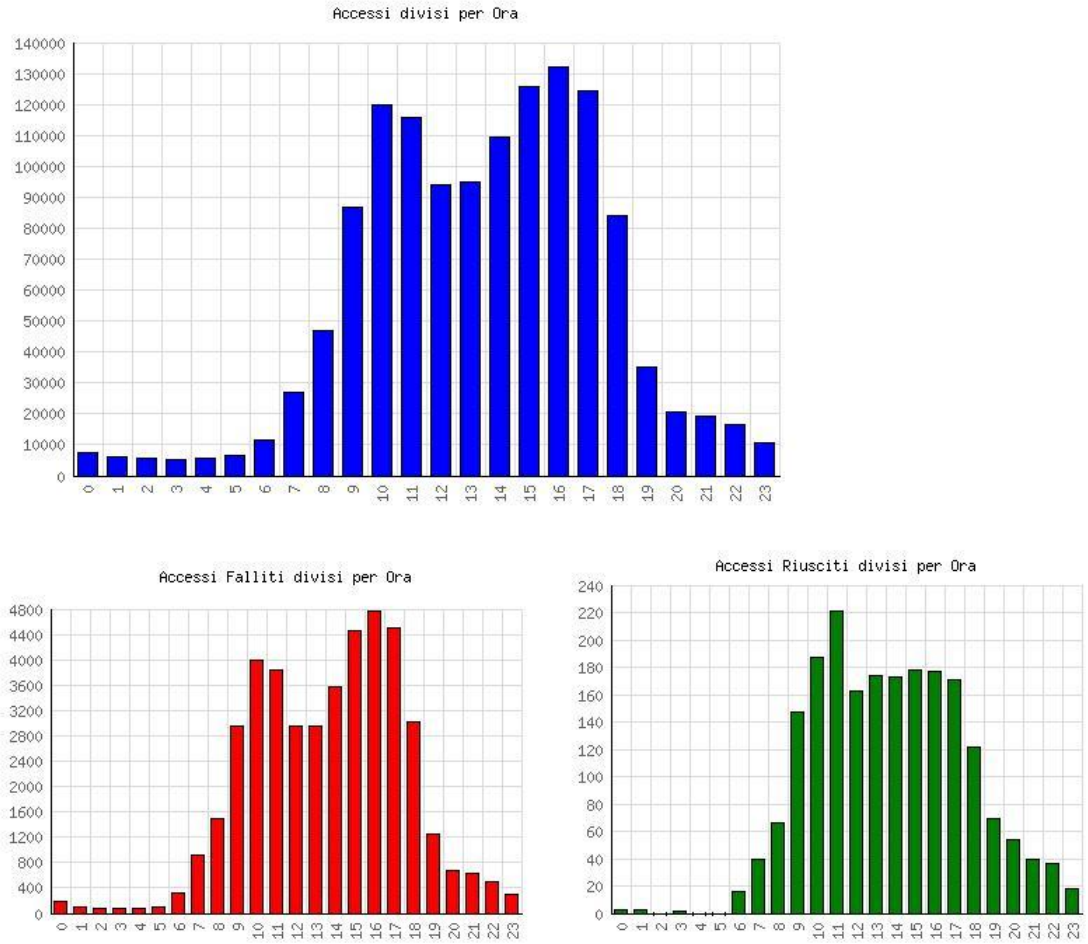


Figura A.10: Screenshot del modulo Bluetooth

A.4 Nuove Installazioni Auditing

La parte più complessa nel replicare la piattaforma di auditing non sta tanto nella lettura quanto nel raccoglimento dei dati.

A.4.1 Raccolta

Come descritto sopra è necessario configurare tutte le macchine una ad una, l'obiettivo di identificare la persona fisica che ha compiuto ogni accesso richiede necessariamente l'utilizzo di credenziali personali, sia che si acceda tramite LDAP che utilizzando sistemi e credenziali locali per accedere.

L'installazione dei pacchetti necessari per l'allacciamento all'LDAP e la loro successiva configurazione sono da eseguire seguendo la traccia indicata durante la trattazione dell'implementazione dell'Auditing, mentre l'installazione del pacchetto Auditd, il quale offre un livello di sicurezza maggiore rispetto all'utilizzo del solo Syslog, infatti questo pacchetto propone un sistema autobloccante per la propria configurazione e inoltre ogni login eseguita è interfacciata direttamente con il demone auditd, che, se non è attivo, impedisce ogni accesso alla macchina, mentre per modificarne la configurazione è necessario un riavvio completo della macchina.

Per le macchine Windows è necessario installare e configurare correttamente Snare su ogni macchina isolata e sull'eventuale PDC presente, se sono presenti Database vanno configurate le security di base, come descritto nel capitolo 8 e impostate le credenziali personali. In caso vi siano Database, va ricordato di impostare Snare con i codici associati ad essi, altrimenti gli eventi riguradanti i DB non verranno inoltrati.

Le configurazioni per la ricezione dei logs, invece, richiedono innanzitutto una macchina amministrata esternamente rispetto all'azienda in cui si opera, oppure che riceva manutenzione unicamente sotto supervisione di un responsabile, per evitare manomissioni. La configurazione del Server di raccolta richiede unicamente una corretta configurazione di eventuali firewall frapposti tra le sorgenti ed il server, Syslog configurato come descritto in precedenza ed il demone auditd in esecuzione, e configurato in maniera tale che non sia possibile interrompere syslog nè modificarne le impostazioni.

Fatto ciò è fondamentale adattare gli script per la gestione a lungo termine dei logs, le operazioni di gestione di maggior rilievo sono le seguenti:

- la compressione dei logfiles giornalieri

- il calcolo degli MD5 e salvataggio su file giornaliero/mensile
- la scrittura su disco e la chiusura del disco stesso
- l'invio delle e-mail di avviso, promemoria e quelle contenenti gli MD5

A.4.2 Log Analysis

Una volta configurato il sistema di invio, raccolta e gestione a lungo termine dei logs, è possibile procedere con l'installazione dei pacchetti utili al funzionamento della piattaforma di Analisi dei Logs.

I pacchetti in questione sono gli stessi del log Mining, qua, però, va ricordato di impostare la possibilità di effettuare upload e la dimensione massima di questi ultimi per Apache. Fatto ciò, fa comodo installare anche phpmyadmin, il quale permette di interagire con il proprio database Mysql presente sulla macchina (o altrove), importare il file table.sql contenente tutte le tabelle necessarie all'utilizzo dell'applicativo, l'user di default e qualche dato di test.

Arrivati a questo punto è sufficiente copiare la cartella audit all'interno della cartella che Apache rende disponibile per Internet : /var/www/.., configurare le impostazioni dell a piattaforma quali : i nomi delle cartelle utilzzate, le Regular expression se vengono cambiati i dati in ingresso, le credenziali per accedere al database creato poco fa ecc.., in pratica è sufficiente personalizzare il file Settings.ini per le proprie esigenze.

Appendice B

Tecnologie

In questo Appendice sono descritti i protocolli, e le loro implementazioni, che hanno permesso la realizzazione di questi sistemi di Auditing.

B.1 Syslog

Syslog (abbreviazione di System Log) è uno standard per l'invio di messaggi di log in una rete IP. Il termine "syslog" viene utilizzato per indicare sia l'effettivo protocollo Syslog, sia per l'applicazione o la libreria che si occupa della spedizione e della ricezione dei messaggi di log.

B.1.1 Il protocollo

Syslog è un protocollo di tipo client/server. Il *syslog sender* invia un piccolo (al massimo di taglia 1 KB o 1024 caratteri) messaggio testuale al *syslog receiver*. Quest'ultimo viene comunemente chiamato *"syslogd"*, *"syslog daemon"* o *"syslog server"*. I messaggi Syslog possono essere inviati sia via UDP sia via TCP. I dati vengono spediti in chiaro (*cleartext*); sebbene non faccia parte delle specifiche del protocollo stesso, è possibile utilizzare un *wrapper* in grado di fornire cifratura alla connessione tramite SSL/TLS. Per fare un esempio, un'applicazione Syslog viene spesso impiegata in simbiosi con *stunnel* (http://www.stunnel.org/).

Syslog viene tipicamente adottato per la gestione di sistemi di rete e per motivi di sicurezza ed affidabilità del sistema (*security auditing*). Il protocollo è supportato da un'ampia varietà di dispositivi di rete su numerosi tipi di piattaforme; per questo motivo, Syslog può essere sfruttato per integrare informazioni di log provenienti da differenti sistemi, convogliandole in un'unica repository centralizzata.

Syslog nacque nel 1980 come parte del progetto Sendmail, ma la sua flessibilità gli permise ben presto di applicarsi efficientemente anche all'interno di altri progetti software. Il software Syslog (o meglio il demone *syslogd*) è stato per molti anni lo standard *de facto* per effettuare logging, sia in locale che in remoto, su macchine Linux e in generale con sistema operativo Unix-based oltre che su diversi dispositivi di altro genere. Recentemente Syslog è diventato un protocollo ed è stato standardizzato dalla IETF (*Internet Engineering Task Force*), il cui *working group* omonimo ha prodotto nel 2001 il documento RFC 3164. Un secondo documento, RFC 3195, rilasciato nello stesso anno riguarda la consegna affidabile (*reliable delivery*) nel protocollo Syslog.

La porta assegnata dalla IANA (Internet Assigned Numbers Authority) al protocollo Syslog è la 514. Bisogna prestare attenzione in quanto la porta registrata è relativa al solo protocollo UDP, mentre la 514/TCP è allocata al protocollo *shell (cmd)*. Ad ogni modo, assicurandosi che la porta in questione non venga già impiegata da shell, nulla vieta all'istanza Syslog di utilizzare la 514 in TCP. La porta 601, riferita a *syslog-conn* (descritto nel RFC 3195), prevede l'utilizzo di entrambi i protocolli di trasporto. Infine, la porta 6514 di TCP è associata all'estensione *Syslog over TLS* (standardizzato in RFC 5425).

Ulteriori informazioni ai siti http://-www.syslog.cc/-ietf/-protocol.html e http://www.monitorware.com/common/en/articles/syslog-described.php .

B.1.2 Implementazioni

L'implementazione originaria, risalente ai tempi in cui Syslog non era ancora un protocollo, era nota come *syslogd* (http://linux.die.net/man/8/syslogd) ed era disponibile solamente per sistemi Unix-like (BSD e Linux). Versioni più recenti, anche queste open source, sono Rsyslog (http://www.rsyslog.com/) e Syslog-NG (http://www.balabit.com/network-security/syslog-ng/) : entrambe possono fare le veci del client così come del server.

Per quanto riguarda le macchine con sistema operativo Windows, sono presenti solamente implementazioni proprietarie (e quindi a pagamento), tra cui possiamo citare Kiwi Syslog Server (http://www.kiwisyslog.com/) e WinSyslog (http://www.winsyslog.com) .

In quanto a Syslog agents (client), sono numerosi gli apparati di rete che rispettano il protocollo: tra i maggiori produttori citiamo Cisco, Extreme Networks, Fujitsu, Huawei, IBM, NetGear, Symantec.

B.2 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. Fu sviluppato presso la Livingston Enterprises Inc. nel 1991 come protocollo per i server d'accesso ai servizi di connettività (NAS - Network Access Server), e venne successivamente a far parte della suite di standard della IETF. RADIUS è attualmente lo standard de-facto per l'autenticazione remota, prevalendo sia nei sistemi nuovi che in quelli già esistenti.

B.2.1 Il protocollo

Il protocollo RADIUS nella sua interezza e nelle sue varianti è definito in numerosi RFC; tra questi i principali sono RFC 2865 (Remote Authentication Dial In User Service) e RFC 2866 (RADIUS Accounting), rilasciati entrambi nel 2000.

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.

La figura B.1 mostra 4 tra le principali modalità di accesso alle risorse di rete con autenticazione basata su protocollo RADIUS. La quarta modalità, raffigurata in basso, è quella tipica nei casi di ISP che forniscano connettività wireless ai propri clienti registrati: questa soluzione fa uso di access point collegati direttamente ai NAS, i quali fungono da RADIUS client nella comunicazione con il server di autenticazione.

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). Gli schemi di autenticazione supportati sono PAP, CHAP e EAP. 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 pac-

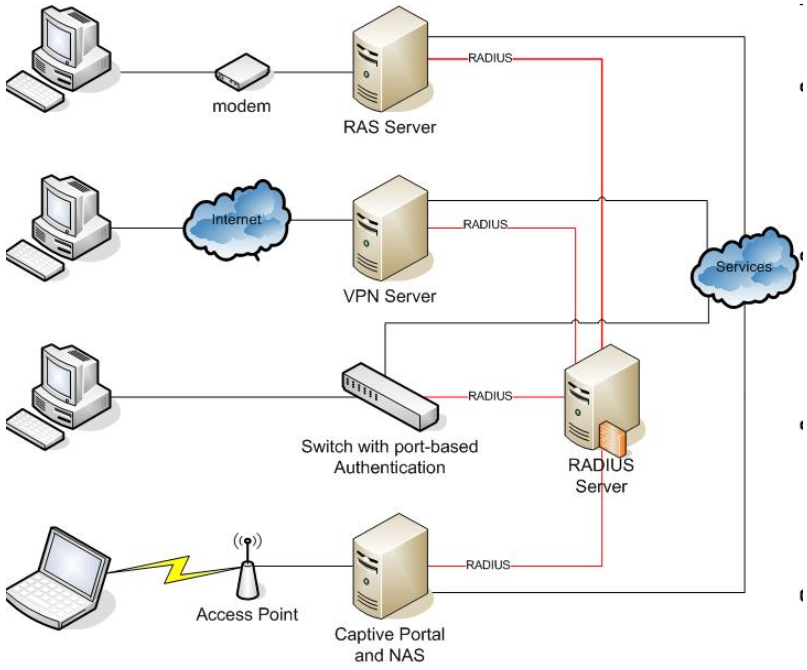


Figura B.1: Autenticazione in RADIUS

chetto 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.

La figura B.2 raffigura il flusso del processo con cui un utente si autentica nel protocollo RADIUS. Sono chiaramente distinguibili le 3 entità: utente, client NAS e server di autenticazione. Sottolineiamo che, tipicamente, una rete contiene un solo server di autenticazione (che gestisce il database RADIUS) e diversi client NAS, terminali a cui gli utenti si collegano direttamente.

La IANA ha assegnato le porte UDP 1812 a RADIUS Authentication e 1813 a RADIUS Accounting. Preme sottolineare che, precedentemente all'allocazione ufficiale da parte della IANA, venivano usate in maniera non ufficiale le porte 1645 and 1646 (per Authentication e Accounting, rispettivamente), che divennero così le porte impiegate di default da molte implementazioni Radius (sia client che server): per ragioni di backwards compatibility, in alcuni casi queste porte continuano tuttora ad essere adoperate.

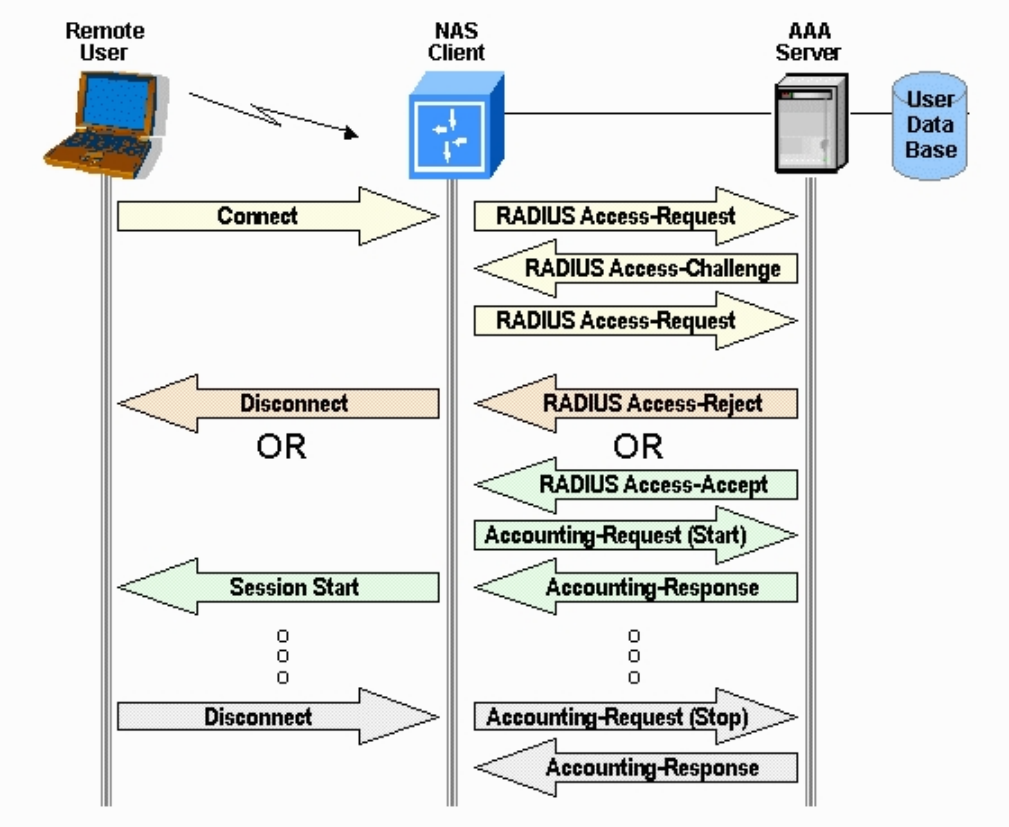


Figura B.2: Diagramma di flusso del processo di autenticazione in RADIUS

Protocolli di autenticazione concorrenti sono Kerberos ed il recente Diameter.

Ulteriori informazioni sul protocollo sono reperibili presso http://en.wikipedia.org/wiki/RADIUS .

B.2.2 FreeRadius

Le implementazioni del protocollo, relativamente al server RADIUS, sono numerose: per una lista completa si rimanda a http://en.wikipedia.org/wiki/List_of_RADIUS_servers .

La soluzione più popolare ed utilizzata nel mondo è certamente FreeRadius (http://freeradius.org/) . L'implementazione open source (licenza GPL) del server RADIUS viene distribuita con l'inclusione di varie features:

- supporto completo per gli attributi definiti nelle RFC 2865 e RFC 2866;
- implementazione del protocollo EAP, compresi i sotto-tipi EAP-MD5, EAP-SIM, EAP-TLS, EAP-TTLS, EAP-PEAP, e Cisco LEAP EAP;
- supporto per attributi vendor-specific per circa un centinaio di produttori;

- supporto per diversi RADIUS clients, tra cui ChilliSpot/CoovaChilli, JRadius, SecureW2 EAP, Xsupplicant, e altri;
- una development library (con licenza BSD) per l'implementazione di client RADIUS;
- il modulo PAM per l'autenticazione degli utenti e l'accounting;
- un modulo per l'integrazione con Apache;
- uno strumento di amministrazione, chiamato *dialupadmin* e programmato in PHP, è presente nella distribuzione ed è accessibile all'utente come applicazione web.

FreeRadius è modulare, altamente flessibile nella configurazione, ad elevato livello di performance e scalabilità. Il software è stato testato e si è dimostrato scalabile in sistemi dell'ordine dei milioni di utenti. Le piattaforme supportate sono tutte le Unix-based e Windows. La documentazione è consultabile all'indirizzo http://wiki.freeradius.org/Main_Page.

B.3 CoovaChilli

CoovaChilli è un software open-source per il controllo degli accessi ad una WLAN. Si base sul popolare ChilliSpot, progetto (ora abbandonato) di cui è erede: viene mantenuto attivamente da una web community, comprendente anche membri del progetto originario ChilliSpot.

L'applicativo è dotato di diverse funzionalità che lo rendono uno degli *access controller* e *captive portal* (anche noti come *walled-garden environment*) più utilizzati. I walled garden sono ambienti che controllano l'accesso degli utenti ai contenuti e ai servizi di rete: essi dirigono la navigazione all'interno di particolari aree, per permettere l'accesso ad un determinato insieme di risorse negando l'accesso ad altre risorse. Gli ISP possono stabilire che i propri utenti siano in grado di visitare alcune pagine web (*within the garden*) ma non altre (*outside the walls*).

Il captive portal offre il vantaggio di essere uno UAM (Universal Access Method), garantisce quindi la fruibilità da qualunque piattaforma, dispositivo o sistema operativo, e nel contempo non richiede nessun intervento lato utente.

Esso prevede infatti l'utilizzo di un comune web browser piuttosto che un client specifico per effettuare l'accesso alla rete.

Si appoggia a RADIUS per fornire l'accesso alla rete wireless, l'autenticazione al servizio e l'accounting delle risorse. Nel pacchetto software di CoovaChilli è presente, come parte integrante, il firmware CoovaAP (basato su OpenWRT[1]). CoovaAP è un'implementazione specifica e specializzata per hotspots, e svolge l'effettiva funzione di *access controller* per CoovaChilli.

La figura B.3 esemplifica il modello di rete adottato per sistemi basati su controllo degli accessi ChilliSpot/CoovaChilli e server di autenticazione RADIUS.

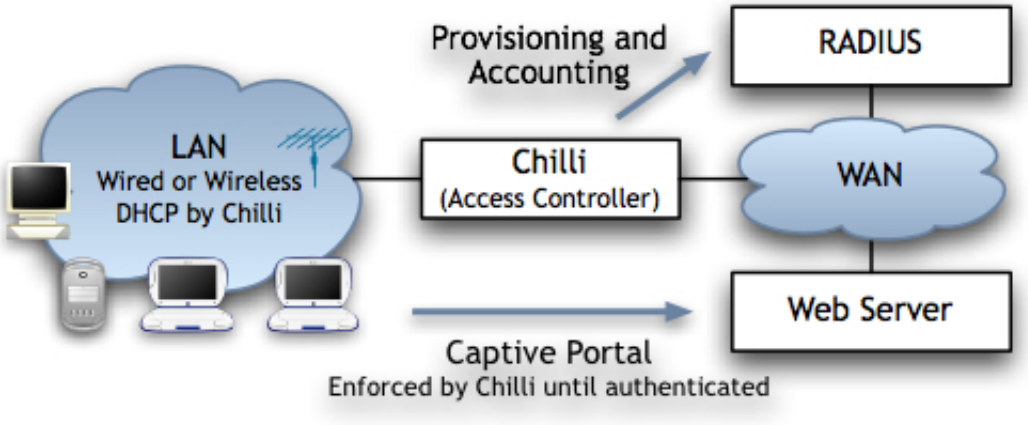


Figura B.3: Schema interazione tra CoovaChilli e RADIUS

Cenni sul funzionamento e sul processo CoovaChilli assume il controllo dell'interfaccia Ethernet interna (*eth1*), per mezzo di un modulo del kernel chiamato *vtun* (abbreviazione di Virtual Tunnel, sito ufficiale: http://vtun.sourceforge.net): questo modulo implementa e gestisce un'interfaccia virtuale, cui associa il nome *tun0*. Il modulo *vtun* è impiegato per redirigere i pacchetti IP dal kernel a user-mode, in maniera tale che CoovaChilli è in grado di funzionare senza moduli kernel oltre a quelli standard.

Sull'interfaccia virtuale *tun0*, CoovaChilli imposta un server DHCP. Un client che si connetta a tale interfaccia si vedrà respingere tutti i pacchetti, almeno fino a quando esso non si sia autenticato ed abbia ricevuto l'autorizzazione: questa operazione si realizza per mezzo della *login page* di CoovaChilli, la quale agisce da *supplicant* per l'autenticazione al servizio. Quando un client non autenticato

[1] OpenWRT è una distribuzione Linux per dispositivi embedded, licenziata GNU GPL e adottata da molti produttori di router e access point, tra cui Linksys, NetGear, D-Link e Asus

prova a connettersi ad una pagina web (sulle porte 80-http o 443-https), la richiesta viene intercettata da CoovaChilli e subisce un redirect ad uno script Perl chiamato *hotspotlogin.cgi* (servito da apache sulla 443-https).

hotspotlogin.cgi visualizza all'utente finale una pagina per l'inserimento dei parametri di autenticazione (tipicamente username e password). Questi valori vengono poi inoltrati al server FreeRadius, che ne effettua il matching con le informazioni presenti nel suo backend database[2], secondo i protocolli di autenticazione PAP o CHAP, eventualmente proteggendo la comunicazione con cifratura SSL/TLS. A questo punto, l'utente viene rifiutato oppure autenticato dal server Radius, che istruisce *hotspotlogin.cgi* a presentare al cliente, a seconda del caso, un messaggio di rifiuto oppure una pagina contenente tra le altre cose una notifica di successo ed un link per il logout dalla rete.

Recentemente, oltre alla versione basata sullo script CGI, l'implementazione del redirecting del captive portal può essere ottenuta anche per mezzo dell'interfaccia JSON (JavaScript Object Notation) che definisce un formato standard per lo scambio dei dati di autenticazione.

Ulteriori dettagli consultabili nella documentazione ufficiale del progetto: http://coova.org/wiki/index.php/CoovaChilli/Documentation.

B.4 Squid

Squid è un popolare software libero con funzionalità di proxy e web cache, rilasciato sotto la GNU General Public License. Ha una vasta varietà di usi, da quello di rendere più veloce un server web usando una cache per richieste ripetute, fornisce sia un servizio di cache per il web che per DNS e altri tipi di ricerche all'interno di reti con risorse condivise, e filtri sul traffico permesso. È stato primariamente sviluppato per piattaforme Unix-like.

Squid è in sviluppo da diversi anni ed è ormai considerato un'applicazione sicura e robusta. Supporta molti protocolli, ma è comunque primariamente un proxy HTTP e FTP. È inoltre disponibile supporto per TLS, SSL, Gopher e HTTPS.

Proxy Web La funzione di *caching* è un modo di salvare oggetti Internet richiesti (pagine web), è disponibile via HTTP, FTP e Gopher in un sistema più

[2]Il backend tipicamente è un DBMS (es. MySQL o PostgreSQL), ma può anche essere costituito da altri servizi, come LDAP, Kerberos, file *passwd* di Unix, Active Directory.

vicino al sito richiedente. Il browser può usare la cache di Squid locale come un proxy HTTP server, riducendo l'accesso ai server nonché il consumo di banda. Questo è funzionale ai service provider. L'introduzione di server proxy introduce comunque anche questioni relative alla privacy dal momento che tutte le richieste che vi transitano possono essere salvate, si possono includere informazioni relative al tempo esatto, il nome e la versione ed il sistema operativo del browser che richiede la pagina.

Sul programma client (nella maggior parte dei casi un browser) può avere specificato esplicitamente il server proxy che si vuole usare o può usare un proxy senza altre specifiche configurazioni, in questo caso si parla di *transparent proxy*, nel qual caso tutte le richieste HTTP sono interpretate da Squid e tutte le risposte sono salvate. L'ultima menzionata è tipicamente una configurazione aziendale (tutti i client sono sulla stessa LAN) questo spesso introduce i problemi di privacy menzionati precedentemente.

Squid possiede alcune funzioni che possono aiutare a rendere anonime le connessioni, per esempio disabilitando o cambiando dei campi specifici nell'intestazione delle richieste HTTP. Che questi campi siano impostati o meno dipende dalla configurazione del server Squid che funziona da proxy. Le persone che richiedono pagine attraverso una rete che usa Squid in modo trasparente generalmente non sono informate sul fatto che le informazioni sono memorizzate in un registro.

Per maggiori dettagli, si rimanda alla documentazione ufficiale:
www.squid-cache.org/.

B.5 LDAP

Due precursori di LDAP sono rappresentati dagli RFC rilasciati da IETF, Directory Assistance Service (RFC 1202) e DIXIE Protocol Specification (RFC 1249). Sono entrambi RFC informativi e non vennero proposti come standard. Il directory assistance service (DAS) definì un metodo per il quale un directory client può comunicare con un proxy su un host OSI che rilasciava richieste X.500 a nome del client. DIXIE è simile a DAS, ma offre una conversione più diretta del DAP. La prima versione di LDAP venne definita in X.500 Lightweight Access Protocol (RFC 1487), sostituito poi da Lightweight Directory Access Protocol (RFC 1777).

Più avanti LDAP raffinò le idee e i protocolli di DAS e DIXIE. Ha un'implementazione più neutrale e riduce la complessità del client. La maggior parte dei lavori in DIXIE e LDAP proviene dall'Università del Michigan, che offre una

documentazione delle implementazioni di LDAP e mantiene pagine Web e mailing list su LDAP. RFC 1777 definisce il protocollo LDAP stesso, insieme con: La rappresentazione in Stringhe e Sintassi degli Attributi Standard (RFC 1778), Rappresentazione in Stringa dei Distinguished Name (RFC 1779), Formato per l'URL LDAP (RFC 1959), Rappresentazione in stringa dei filtri di ricerca LDAP (RFC 1960) La versione 2 di LDAP ha ottenuto lo stato di standard bozza nel processo di standardizzazione IETF, un passo dall'essere uno standard. Oggi, tutte le implementazione dei directory server sono basati su LDAP versione 3.

Recentemente, il bisogno di unire le operazioni LDAP con XML nell'uso dei Web Services ha dato alla luce un nuovo linguaggio chiamato Directory Services Markup Language (DSML). La più recente versione è DSMLv2. DSML è un generico formato per importare/esportare tali informazioni. In DSML i dati della directory possono essere condivisi tra applicazioni che supportano tale formato senza esporre il protocollo 17 LDAP. XML offre un metodo efficace per presentare e trasferire i dati; i servizi di directory permettono di condividere e gestire i dati e sono così un prerequisito necessario per effettuare operazioni online. DSML è progettato per rendere il servizio di directory più dinamico impiegando XML. DSML è uno schema in XML per lavorare con le directory, ed è definito con un Document Content Description (DCD). Così DSML permette ai programmatori di XML di accedere alle directory LDAP senza avere a che fare con l'interfaccia LDAP o API per l'accesso alle directory, offrendo un modo uniforme per lavorare con directory multiple e differenti.

Ldap ha influenzato lo sviluppo di altri protocolli di rete, come il Service Provisioning Markup Language (SPML) e il Service Location Protocol. Directory LDAP.
Il termine di uso comune directory LDAP può essere fuorviante. Nessun tipo specifico di directory è una directory LDAP. Si potrebbe ragionevolmente usare il termine per descrivere qualsiasi directory accessibile tramite LDAP e che possa identificare gli oggetti contenuti tramite nomi X.500, ma Directory come OpenLDAP e i suoi predecessori sviluppati presso l'Università del Michigan, anche se progettati espressamente per l'accesso tramite LDAP piuttosto che come ponte verso X.500, come avveniva per i prodotti forniti da ISODE, non sono directory LDAP più di qualsiasi altra directory accessibile tramite protocollo LDAP.

L'informazione all'interno di una directory è organizzata in elementi chiamati entry.
Gli elementi di una directory LDAP presentano una struttura gerarchica che riflette confini politici, geografici o organizzativi. Nel modello X.500 originale, gli

elementi che rappresentano gli stati appaiono in cima all'albero, con sotto di essi gli elementi per gli stati federali o le organizzazioni nazionali (normalmente nelle installazioni di LDAP vengono usati i nomi del DNS per strutturare i livelli più alti della gerarchia). Più in basso potrebbero apparire elementi per rappresentare le divisioni all'interno di una singola organizzazione, singole persone, documenti, stampanti o qualsiasi altra cosa.

Nella struttura ad albero, ad ogni livello esiste un Relative distinguished name (RDN) che lo identifica (ad esempio ou=stage). L'unione di tutti i RDN, presi in successione dal nodo foglia fino alla radice, costituisce il distinguished name (DN), una stringa che rappresenta univocamente una entry nella directory. Un dn può essere ad esempio:

cn=Szabo Karoly,ou=stage,dc=telerete,dc=net

Ciascuna entry ha una serie di attributi, costituiti dall'associazione attributo-valore; per ogni attributo possono esserci più valori. Ognuno degli attributi dell'elemento è definito come membro di una classe di oggetti, raggruppati in uno schema. Ogni elemento nella directory è associato a una o più classi di oggetti, che definiscono se un attributo sia opzionale o meno, e che tipo di informazioni questo contenga. I nomi degli attributi solitamente sono scelti per essere facilmente memorizzabili, per esempio cn per common name, o mail per un indirizzo e-mail. I valori degli attributi dipendono dal tipo, e la maggioranza dei valori non binari sono memorizzati in LDAPv3 (LDAP versione 3) come stringhe UTF-8. Per esempio, un attributo di tipo mail potrebbe contenere il valore user@example.com, mentre un attributo jpegPhoto potrebbe contenere una fotografia nel formato (binario) JPEG.

Per maggiori dettagli, si rimanda alla documentazione ufficiale del sistema da me studiato:
www.openldap.org/.

B.6 Snare

Lo snare per Windows è un servizio compatibile di Windows NT, di Windows 2000, di Windows.xp e di Windows 2003 che si interagisce con il sottosistema di fondo di Windows Eventlog per facilitare il trasferimento a distanza e in tempo reale delle informazioni del ceppo di evento. I ceppi di evento dai ceppi di sicurezza, di applicazione e di sistema, così come il nuovo DNS, archiviano il servizio

della replica ed i ceppi attivi dell'indice sono sostenuti. I dati di ceppo sono convertiti in disposizione di testo e sono trasportati ad un assistente a distanza dello snare, o ad un assistente a distanza di Syslog con le regolazioni configurabili e dinamiche di priorità e della funzione.

Lo snare attualmente è usato dalle centinaia dei migliaia degli individui e dalle organizzazioni universalmente. Lo snare per Windows è usato da molte grandi organizzazioni finanziarie, di assicurazione, di Healthcare, della difesa, di aerospazio e di intelligenza per venire a contatto degli elementi dei requisiti locali e federali di sicurezza. Lo SNARE britannico di campione BS7799 per Windows è software libero (freeware), liberato sotto i termini della patente pubblica di GNU (GPL).

Sono supportati :

- gli Event Log dalla Security,
- dalle Applicazioni,
- i System logs,
- i nuovi DNS,
- File Replication Service,
- ed infine l'Active Directory

Le versioni supportate dell'agente si allacciano all'event log classico di Windows, I dati prelevati vengono convertiti in formato testuale e inoltrati verso un server Snare remoto, o verso un server Syslog remoto, con la possibilità di configurare facility dibnamicamente e impostazioni sulla priorità.

Per maggiori dettagli, si rimanda alla documentazione ufficiale:
http://www.intersectalliance.com/projects/SnareWindows/index.html.

B.7 MRTG

Il Multi Router Traffic Grapher (MRTG) è uno strumento per monitorare il carico di traffico sui collegamenti di rete. MRTG genera pagine HTML contenenti immagini PNG, che offrono una rappresentazione visiva LIVE di questo traffico. I suoi punti di forza sono:

- Portabilità: MRTG funziona sulla maggior parte delle piattaforme UNIX e Windows NT.

- Perl: MRTG è scritto in Perl e viene fornito con sorgente completo.
- Portable SNMP: MRTG Utilizza una applicazione SNMP estremamente portatile interamente scritto in Perl (grazie a Simon Leinen). Non c'è bisogno di installare alcun pacchetto esterno SNMP.
- Supporto SNMPv2c: MRTG in grado di leggere i nuovi contatori SNMPv2c 64bit.
- Affidabili Interface Identificazione: Router interfacce possono essere identificati tramite l'indirizzo IP, la descrizione e l'indirizzo Ethernet, oltre al numero normale interfaccia.
- File di log dimensioni costanti: I file di log MRTG's non crescono grazie all'utilizzo di un algoritmo di consolidamento dei dati.
- Configurazione automatica: MRTG viene fornito con una serie di strumenti di configurazione che semplificano la configurazione e l'installazione.
- Prestazione: Le routine critiche sono scritte in C.
- GIF Graphics: La grafica è generata direttamente in formato PNG utilizzando la libreria GD di Thomas Boutell.
- Personalizzazione: L'aspetto delle pagine web prodotte da MRTG è altamente configurabile.
- RRDtool: MRTG è dotato di ganci per l'utilizzo RRDtool. Se siete a corto di prestazioni questo può aiutare.

MRTG consiste in uno script Perl che utilizza SNMP per leggere i contatori del traffico del vostro router e un programma in C che registra i dati di traffico e crea grafici rappresentanti il traffico sulla connessione di rete monitorata. Questi grafici sono inseriti in pagine web che possono essere visualizzati da qualsiasi moderno Web-browser.

Oltre ad una vista dettagliata quotidiana, MRTG crea anche rappresentazioni visive del traffico visto nel corso degli ultimi sette giorni, le ultime cinque settimane e gli ultimi dodici mesi. Questo è possibile perché MRTG tiene un registro di tutti i dati estrapolati dai router. Questo registro è automaticamente consolidato in modo da non crescere nel tempo pur contenendo ancora tutti i dati rilevanti per tutto il traffico visto negli ultimi due anni. Questi compiti vengono svolti in maniera efficiente, pertanto, è possibile monitorare 200 o più collegamenti di rete.

MRTG non si limita al monitoraggio del traffico però, è possibile controllare tutte le variabili SNMP scelte. È possibile anche utilizzare un programma esterno per raccogliere i dati che devono essere controllati e rappresentati tramite MRTG. MRTG è utilizzato per monitorare variabili di rete quali: carico del sistema, Login Sessions, la disponibilità di modem e altro ancora. MRTG vi consente anche di accumulare due o più fonti di dati in un unico grafico.

Per maggiori dettagli, si rimanda alla documentazione ufficiale:
http://oss.oetiker.ch/mrtg/.

B.8 Nagios

Nagios è una nota applicazione open source per il monitoraggio di computer e risorse di rete. La sua funzione base è quella di controllare nodi, reti e servizi specificati, avvertendo quando questi non garantiscono il loro servizio o quando ritornano attivi.

Nagios è stato originariamente creato sotto il nome di Netsaint e mantenuto da Ethan Galstad. In origine Nagios è stato sviluppato per Linux, ma può funzionare correttamente anche su altre varianti di Unix; è rilasciato sotto la GNU General Public License Versione 2 pubblicata dalla Free Software Foundation.

Alcune caratteristiche:

- Monitoraggio di servizi di rete (SMTP, POP3, IMAP, HTTP, NNTP, ICMP, SNMP, FTP, SSH, ...);
- Monitoraggio delle risorse di sistema (carico del processore, uso della RAM, uso dell'hard disk, log di sistema sulla maggior parte dei sistemi operativi, ...);
- Monitoraggio remoto supportato attraverso SSH o SSL encrypted tunnels;
- Semplici plugin che permettono agli utenti di sviluppare facilmente nuovi controlli per i servizi in base alle proprie esigenze (usando Bash, C++, Perl, Ruby, Python, PHP, C#, ...);
- Controlli paralleli sui servizi;
- Capacità di definire gerarchie di nodi di rete usando nodi "parent", permettendo la distinzione tra nodi che sono down e nodi non raggiungibili (unreachable);

- Notifiche quando l'applicazione riscontra problemi o la loro risoluzione (via email, pager, SMS, o con altri sistemi per mezzo di plug-in aggiuntivi);
- Possibilità di definire "event handlers", ovvero azioni automatiche che vengono attivate all'apparire o alla risoluzione di un problema;
- Rotazione automatica dei file di log;
- Supporto per l'implementazione di monitoring ridondato;
- Interfaccia web opzionale per la visualizzazione dell'attuale stato, notifiche, storico dei problemi, file di log, etc.

Per maggiori dettagli, si rimanda alla documentazione ufficiale: http://www.nagios.org/documentation/.

B.9 Growisofs

dvd+rw-tools (conosciuto anche come growisofs) è una collezione di programmi open source popolari, che permettono la scrittura di DVD, e di recente anche di Blu-ray Disc, su Linux, FreeBSD, Windows e anche su Mac OS X v10.4.

Il pacchetto stesso richiede la dipendenza di un altro programma che viene utilizzato per creare immagini ISO9660 al volo. Funzione fornita da mkisofs (dal pacchetto cdrtools) o da genisoimage (dal pacchetto cdrkit). Il pacchetto è rilasciato sotto la GNU General Public License.

Per maggiori dettagli, si rimanda alla documentazione ufficiale: http://fy.chalmers.se/ appro/linux/DVD+RW/.

B.10 Software utilizzati durante lo stage

Oltre agli applicativi descritti in precedenza, ho utilizzato numerosi altri software:

- GZip (http://www.gzip.org/)
- OpenOffice (http://www.openoffice.org/)
- Microsoft Visio (http://office.microsoft.com/it-it/visio)

- MikTex (http://miktex.org/)
- PuTTY (http://www.putty.org/)
- TexnicCenter (http://www.texniccenter.org/)
- WinSCP (http://winscp.net/eng/docs/)
- VMWare (http://www.vmware.com/it/)
- Nagios (www.nagios.org/)
- PHPGraphLib (http://www.ebrueggeman.com/phpgraphlib/)
- Sendmail (http://www.sendmail.org/)
- Ssntp (linux.die.net/man/8/ssmtp)

Appendice C

Acronimi

AAA : Authentication Authorization Accounting
AD : Active Directory
ADSL : Asymmetric Digital Subscriber Line
API : Application Programming Interface
ARP : Address Resolution Protocol

BASH : Bourne Again SHell
BDC : Backup Domain Controller
BGP : Border Gateway Protocol
BSD : Berkeley Software Distribution

CGI : Common Gateway Interface
CHAP : Challenge-Handshake Authentication Prot.
CLI : Command Line Interface
CRM : Customer Relationship Management
CVS : Concurrent Version System

DBA : DataBase Administrator
DBMS : DataBase Management System
DHCP : Dynamic Host Configuration Prot.
DNS : Domain Name System
DW : Data Warehouse

EAP : Extensible Authentication Prot.
EDP : Extreme networks Discovery Prot.
ERP : Enterprise Resource Planning
ESB : Enterprise Service Bus
ETL : Extract, Transform, Load

FTP : File Transfer Protocol

GPL : (GNU) General Public License
GPRS : General Packet Radio Service
GPS : Global Positioning System
GUI : Graphic User Interface

HTML : HyperText Markup Language
HTTP : HyperText Transfer Protocol

IANA : Internet Assigned Numbers Authority
ICMP : Internet Control Message Protocol
ICT : Information and Communication Technology
IDS : Intrusion Detection System
IPS : Intrusion Prevention System
IEEE : Institute of Electrical and Electronics Engineers
IETF : Internet Engineering Task Force
IMAP : Internet Message Access Protocol
IP : Internet Protocol
IPSec : IP Security
ISO : International Standards Organization
ISP : Internet Service Provider

KDC : Kerberos Domain Controller

LAN : Local Area Network
LDAP : Lightweight Directory Access Protocol

MAC : Media Access Control

MAN : Metropolitan Area Network
MD5 : Message Digest algorithm 5
MIB : Management Information Base
MIME : Multipurpose Internet Mail Extensions
MRTG : Multi Router Traffic Grapher

NAS (1) : Network Attached Storage
NAS (2) : Network Access Server
NAT : Network Address Translation
NSS : Name Service Switch
NTP : Network Time Protocol

OID : Object IDentifier
OLAP : OnLine Analytical Processing
OSI : Open Systems Interconnection
OTRS : Open (Trouble) Ticket Request System

PAM : Pluggable Authentication Modules
PAP : Password Authentication Protocol
PCRE : Perl Compatible Regular Expressions
PDC : Primary Domain Controller
PHP : Hypertext Preprocessor
PLC : Programmable Logic Controller
PMI : Piccola e Media Impresa (in Italia)
PoE : Power over Ethernet
POP : Post Office Protocol
PPP : Point-to-Point Protocol

QoS : Quality of Service

RADIUS : Remote Authentication Dial-In User Service
RAID : Redundant Array of Inexpensive Disks
RDBMS : Relational DBMS
RFC : Request for Comments
RPC : Remote Procedure Calls

RTA : Round Trip Average
RTT : Round Trip Time

SAN : Storage Area Network
SCP : Secure Copy
SEC : Simple Event Correlator
SLA : Service Level Agreement
SMBFS : Server Message Block File System (Samba)
SMTP : Simple Mail Transfer Protocol
SNMP : Simple Network Management Protocol
SNMPTT : SNMP Trap Translator
SQL : Structured Query Language
SSH : Secure SHell
SSL : Secure Sockets Layer

TCP : Transmission Control Protocol
TLS : Transport Layer Security

UAM : Universal Access Method
UDP : User Datagram Protocol
UPS : Uninterruptible Power Supply
URL : Uniform (Universal) Resource Locator

VLAN : Virtual Local Area Network
VPN : Virtual Private Network

W3C : World Wide Web Consortium
WAN : Wide Area Network
WEP : Wired Equivalent Privacy
WINBIND : Windows Berkeley Internet Name Domain
WLAN : Wireless Local Area Network
WPA : Wi-Fi Protected Access

XML : eXtensible Markup Language

Bibliografia e sitografia

[1] Iptables - varie fonti

[2] Understanding and Deploying LDAP Directory Services - Howes, Smith, Good

[3] MySql Bible - Suehring

[4] http://en.wikipedia.org

[5] http://www.coova.org/

[6] http://www.tuttoirc.it/linux-squidproxy.php

[7] http://www.chillispot.info/

[8] http://www.cyberciti.biz/

[9] http://it.php.net/manual/en/index.php (PHP)

[10] http://dev.mysql.com/doc/ (MySQL)

[11] http://www.w3.org/TR/html401/ (HTML)

[12] http://tldp.org/LDP/abs/html/ (Bash scripting)

[13] "The syslog-ng Administrator Guide", 1.1.4 Edition, Published July 23, 2008

[14] "Manuale d'uso: Syslog-NG", Facoltà di Scienze matematiche, fisiche e naturali, Università degli Studi di Salerno, Corso di Sicurezza su Reti

[15] http://www.rsyslog.com/doc (Rsyslog)

[16] http://www.clavister.com/manuals/ver8x/manual/

[17] http://www.garanteprivacy.it/

[18] http://www.camera.it/

[19] http://www.samba.org

[20] http://www.linuxquestions.org/

[21] http://www.comptechdoc.org/os/linux/manual4/smbconf.html

[22] http://www.acmeconsulting.it/Squid-Book/HTML/

[23] http://wiki.squid-cache.org/

[24] http://openskill.info/

[25] www.oracle.com

[26] http://msdn.microsoft.com/it-it/library/cc280386.aspx

[27] http://technet.microsoft.com/en-us/library/dd277388.aspx

[28] http://technet.microsoft.com/en-us/library/bb742436.aspx

[29] http://www.digicert.com/ssl-certificate-installation-apache.htm

[30] Elisabetta Braggion - Personale Telerete NordEst

[31] MAN pages

[32] http://www.openldap.org/

[33] http://www.tcpipguide.com/free/ (TCP/IP Guide)

[34] http://www.ntp.org/documentation.html (NTP)

[35] http://www.parlamento.it/leggi/decreti/05144d.htm (decreto legge 144)

[36] http://www.senato.it/parlam/leggi/05155l.htm (legge 155)

[37] http://www.linfo.org/command_index.html (Linux main commands)

[38] http://forums.gentoo.org/ (Gentoo Linux distribution)

[39] http://www.cs.rit.edu/ cslab/vi.html (VI and VIM Editors)

[40] http://www.technoids.org/myfirstsendmailcf.html

[41] http://www.sistemistiindipendenti.org/pdf/dansguardian.pdf

[42] http://www.shelldorado.com/

[43] http://wiki.apache.org/httpd/

[44] "The Not So Short Introduction to LATEX 2", by Tobias Oetiker, Hubert Partl, Irene Hyna and Elisabeth Schlegl, Version 4.26, September 25, 2008

Ringraziamenti

In ogni ringraziamento che si rispetti manca sempre qualcuno, quindi probabilmente questo non farà eccezione.

Innanzitutto volevo ringraziare la mia famiglia, per avermi aiutato e sostenuto in questi anni di Università, durante questi anni, nonostante tutto avete sempre creduto in me e questo mi ha aiutato a portare a termine questo primo obiettivo; mi avete sempre dato tutto ciò di cui avevo bisogno, e per questo vi sono grato, vivendo con persone che non hanno sempre avuto questa fortuna sto notando le opportunità che mi state dando per il mio futuro; spero di non essere costretto a fermare qua i miei studi Padovani e di riuscire a proseguire nella specialstica, anche se sarà dura proverò a farlo mantenendomi da solo, in modo da non dover più pesare su di voi.

Un grosso ringraziamento va senza dubbio al professor Filira, per avermi dato questa opportunità, durante la quale ho esplorato il mondo del lavoro e chiarito i miei obiettivi per il futuro, inoltre lo ringrazio per tutti i consigli che mi ha dato sul modo corretto di procedere, la pazienza ed i consigli, per avermi insegnato, durante questo stage, ad applicare ciò che ho appreso in questi anni di Università e per l'avermi fatto documentare sempre tutto, cosa che spesso mi ha salvato dai problemi, causati da macchine, da persone, avvenimenti..ma soprattutto da me. (P.S.: Nonostante tutto confermo che nell'Exchange non ero mai entrato).

Nell'area tecnica devo ringraziare buona parte del personale per i vari consigli tecnici che mi hanno dato, per avermi introdotto nel mondo del lavoro in modo divertente seppur professionale, in particolare ringrazio Giuseppe per avermi aiutato durante questi sei mesi, Alberto, Luigi e Francesco per i consigli tecnici e la professionalità che mi hanno trasmesso. Da non dimenticare Daniele, per essere stato stagista assieme a me e per l'avermi dato la possibilità di dimostrare che il mio lavoro è applicabile in più campi. Un ringraziamento a Giuliano per gli strobili da conifera e per l'iperpiressia, a Barbara per i consigli sulle eterne malattie che hanno accompagnato questo stage e Riccardo, per aver reso le ultime ore di ogni giorno in ufficio più divertenti.

Marija, con la quale convivo da ormai tre anni, durante i quali mi ha temprato e mi ha fatto capire quanto io sia fortunato, la ringrazio anche per l'avermi ingrassato a forza di torte, per il suo carattere totalmente anomalo, e quindi

unico, per avermi insegnato che esprimersi, spesso, è molto meglio rispetto al far finta che vada tutto bene, e quindi per aver eliminato almeno in parte quella timidezza che mi ha sempre bloccato e soprattutto per avermi ridato la voglia di studiare e di inseguire i miei progetti.

Per evitare di dimenticare qualcuno ringrazio in un sol colpo tutti i miei amici di Verona, di Padova, i miei ex compagni di corso (soprattutto dei primi anni), di sicuro tutti i miei coinquilini, attuali e degli anni passati, e tutti i miei compagni pallavolisti/e è anche grazie a tutti voi che sono come sono.

I miei compagni pallavolisti, grazie ai quali ho mantenuto sempre la passione per lo sport

Grazie a tutti quelli che saranno con me in questa giornata, sperando che adesso siate clementi...