Applications Internet : Services de base

(Version provisoire, Fevrier 96)

PULV - ISTM

Enseignement d'Informatique

Module Systèmes Distribués de Traitement de l'Information

Année 1996-97

Patrick.Duval@devinci.fr

Plan du chapitre

Bibliographie

Beaucoup d'ouvrages de référence ont été originellement publiés en Anglais (une part d'entre eux est aujourd'hui disponible en version française).

Internet, Réseaux, Introduction

- Douglas Comer; The Internet Book; Prentice Hall; 1994 (Le B-A-BA de l'Internet)
- Douglas Comer; Internetworking with TCP/IP, Vol I, Principles, Protocols, and Architecture; Prentice Hall; 1995 (la Bible, précis et pédagogique)

Programmation TCP/IP, Applications Client-Serveur

- Douglas Comer & D.Stevens; Internetworking with TCP/IP, Vol III, Client Server Programming & Applications, BSD Socket Version; Prentice Hall; 1992 (une référence)
- W.R. Stevens; TCP/IP Illustrated; Prentice Hall; 1994 (une autre référence)
- J.M. Rifflet; Communications sous Unix; Ediscience International; 1995 (une référence, française cette fois)

Courrier Electronique (Email)

- F.M. Avolio, P.A. Vixie; Sendmail theory and practice; Digital Press:; 1995 (L'essentiel sur l'email et sendmail: technologie, usages, administration, problèmes, solutions. Ouvrage de référence très complet et dense).
- Brian Costales, E.Allman, N.Rickert; Sendmail; Nutshell Handbook, O'Reilly; 1994 (Tout sur l'administration de sendmail. Un ouvrage de référence très abordable et complet pour comprendre et gérer l'outil de base du système global de courrier électronique de l'Internet).
- Marshall T. Rose; The Internet Message; Nutshell Handbook, O'Reilly; 1994 (Présentation synthétique de toutes les composants de l'email, y compris les "user agents" courants, et les principaux standards complémentaires à SMTP: MIME, POP, IMAP, PEM, X400, etc...)

Service de Noms (Domain Name Service)

- Paul Albitz, Cricket Liu ; DNS and Bind; Nutshell Handbook, O'Reilly; 1994 (Tous les details sur le DNS et son administration, ses liens avec l'email, le NIS, etc...)

Administration Réseau

- Quaterman & Carl-Mitchell; The internet Connection; Addison Wesley; 1994 (Un excellent ouvrage introductif à la pratique de l'administration de réseaux et serveurs internets)

Objectifs

Comprendre les fonctions, applications et concepts de base du Courrier Electronique (Email) et du Service de Nommage (DNS) de l'Internet.

Introduction aux services complémentaires de communication de groupe et de contrôle à distance (News, Telnet, R*...)

Comprendre plus précisement:

- Comment utiliser efficacement ces outils Internet dans un contexte professionnel pour gérer l'Information et la Communication d'entreprise
- Les principes de base de la gestion de ces services

pour être mieux à même:

- De développer des applications basées sur ces ressources
- De s'adapter à l'évolution technologique et des standards.

Domain Name Service: des Noms pour l'Internet

Historique du DNS

- Fin des années 60: le réseau Arpanet
- Années 70: quelques centaines de "hosts", répertoriés dans une base centralisée du NIC (Network Information Center, fichier "hosts.txt")
- 1980: sortie de TCP/IP, diffusion dans Unix Berkeley; sources gratuites pour les Universités et la Recherche: bond en avant du nombre de machines interconnectées, Arpanet devient l'épine dorsale d'un grand réseau TCP/IP, en déploiement rapide.
- La gestion centralisée par le NIC des ressources réseau (adresses IP des hosts et serveurs Email) devient problématique face à la multiplication des machines interconnectées.
- 1984: naissance du Domain Name Service (DNS, RFC 882, 883, aujourd'hui RFC 1034, 1035), diffusion rapide avec Unix Berkeley (implémentation **BIND**), application à l'Email

---

Principes de Base du DNS

- Système de nommage abstrait, applicable pour tous les types de Ressources de l'Internet (en premier lieu les "internets", leurs "hosts", leurs "serveurs de mail", etc...).
- Exemples de "noms de domaines":
  - `harward.edu.  java.sun.com.  linux.org.  mail.istm.devinci.fr.`
- Attributs ("Resource Records"):
  - Informations Typées Attachées à ces "noms de domaine"
  - Peuvent fournir des adresses IP de machines, des noms de serveurs de mail, des descripteurs de domaine...
- Le DNS constitue :
  - une Base de Données Distribuée, Répliquée et d'Administration Décentralisée

    - Un Système Client/Serveur ouvert (RFCs publiques: 1033, 1034, 1035...) et "World-Wide" (Standard Internet).
- **Les noms sont totalement indépendants des numéros de réseaux où sont localisées les ressources**:
    - un même domaine peut contenir plusieurs réseaux IP de numéros distincts
    - un même réseau IP peut abriter plusieurs domaines distincts
- Il définit un Arbre de Nommage global (système hiérarchique)
- Concepts de base:
    - Domaines Racines, Sous-Domaines,
    - Zones d'Administration, Délégation,
    - Serveurs de Noms

Serveurs, Délégation, Zones d'Autorité

Les Serveurs de noms :

- Sont responsables des réponses aux requêtes concernant les noms situés dans leur zone d'administration (en général un domaine):
    - Sont responsables du routage de ces requêtes vers les serveurs de leurs sous-domaines,
    - Ou peuvent avoir connaissance directe des attributs ("RR") des éléments de certains de leurs sous-domaines
- Délégation : un serveur de domaine peut déléguer à un autre serveur la responsabilité d'un de ses sous-domaines (gestion des noms et attributs, réponses aux requêtes)
- La Zones d'Autorité d'un serveur est son domaine diminué des sous-domaines dont il a délégué l'autorité à un serveur tiers.
- Zone d'Autorité d'un serveur

Résolutions de Noms

Elles utilisent le modèle Client-Serveur:

- un client ("résolveur") émet une requête vers un serveur ("serveur de noms")
- Le client est soit:
    - une application qui utilise une simple bibliothèque DNS (libc.a, libresolv.a). Il en est ainsi des clients ftp, telnet, web, qui acceptent de se connecter à des machines désignées par nom.
    - un utilitaire dns (nslookup(1)).
- Le client émet sa requête vers un Serveur DNS pour lequel il a été configuré (e.g. `/etc/resolv.conf` sur Unix, tableau de bord TCP/IP sur Mac ou Windows)
- Cette requête par défaut :
    - utilise UDP/IP (TCP/IP est aussi possible, sur option de l'utilisateur)
    - est faite en mode dit "récursif" : le serveur doit effectuer tout le travail et retourner la réponse finale (un mode "iteratif" est aussi possible, sur option utilisateur: le serveur retourne l'information s'il la possède, ou sinon une référence de serveur "plus capable" de répondre que lui-même).

- Schémas de Resolutions
- Liste des ressources d’un domaine

Caches, Réplications

- Caches:
  - les serveurs quand ils travaillent en mode récursif mettent en "cache" les attributs ("RR") attachés aux noms qu’ils voient passer
  - Lors d’une requête ultérieure portant sur le même nom, ils répondront directement au client en retournant les valeurs d’attributs qu’ils ont "cachées".
- Réponses "autoritatives"
  - une réponse DNS sera dite autoritative s’il elle provient directement d’un serveur de de la zone d’autorité dans laquelle se trouve le nom
  - elle est dite non-autoritative si elle provient du cache d’un serveur hors de de cette zone d’autorité
- Pour assurer la cohérence des réplicats: chaque attribut d’un nom DNS a une "Durée De Vie" (TimeTo Live) fournie par son serveur, au delà de laquelle les copies en cache doivent e^tre invalidées (typiquement quelques heures à quelques jours).
- Pour assurer la disponibilité du service, tout domaine doit posséder au moins un serveur Primaire et un serveur Secondaire (qui fonctionne en copie mirroir du Primaire)
- Le serveur Primaire accède à une Base de Donnée locale décrivant les ressources de son domaine. Voir l’ouvrage de Paul Albitz, Cricket Liu sur le DNS pour des détails sur les fichiers de configuration :
  - /usr/local/named/named.domain,
  - /usr/local/named/named.in-addr.net,
  - /usr/local/named/named.in-addr.local

  n.b. les noms de ces fichiers et les emplacements dans le système de fichiers sont indicatifs, ils sont donnés par le fichier de configuration du serveur DNS /etc/named.boot (man named(8)).
- Le serveur secondaire prend copie de cette base aupres du serveur primaire via le réseau et la stocke localement pour repondre aux requêtes des clients.
- La synchronisation primaire/secondaire est ensuite périodique (typiquement toutes les quelques heures)
- Un serveur DNS peut etre configuré en Cache-Seul: il n’a pas alors de base locale, mais juste un cache. Il interroge à chaque requête son serveur Primaire, sauf si son cache lui permet de répondre directement.

Types de requêtes, types d’attributs

Un requête DNS peut demander des attributs ("RR") d’un type donné, ou la liste de tous les attributs disponibles sur un nom..

Ces attributs sont décrits par la base named.domain du serveur primaire du domaine (base maintenue par l’administrateur du domaine).

Les types d'attributs standards sont:

- SOA: Source Of Authority pour le domaine (nom du serveur, email de son administrateur, TTL par défaut pour ses éléments...)
- NS: nom d'un Name Server pour le nom de domaine spécifié
- A: Addresses IP de la ressource (Name to Address mapping, plusieurs si plusieurs connexions réseau pour une machine)
- MX: Noms de Mail Exchangers capables d'acheminer du mail vers cette ressource (plusieurs si plusieurs serveurs de mail pour ce nom de domaine)
- CNAME: Nom Canonique (= alias DNS)
- TXT: "Textual information" attachéé à la ressource (sémantique variable)
- HINFO: "Host Information" attachéé à la ressource (pour permettre l'utilisation de protocoles adaptatifs à différentes plateformes)
- PTR: Address to Name Mapping (usuellement dans les domaines in-addr-arpa)
- RFC 1183
  - RT: Nom de Routeurs capables d'acheminer le traffic IP vers cette ressource (en général ressource hors Internet, p.ex. accessible via ISDN, ou X25)
  - AFSDB: nom d'un serveur AFS pour une cellule AFS donnée.
  - RP: "Responsible Person" pour un domaine ou une resqsource Ressource (email)
- Voir exemples d'interrogation par nslookup: ex 1, ex 2, ex 3

Résolution Inverse d'Adresses

Une question importante pour la gestion du mail par "sendmail" est:

- Quel est le nom DNS de la machine qui me connecte pour m'envoyer un courrier ?
- Soit: quelles ressources DNS sont-elles attachées d'une adresse IP donnée ?

Solution:

- Chaque Domaine gère un sous-domaine du domaine "in-addr.arpa" dont l'arbre des noms est constitué de tous les numéros de **réseaux IP**. Par exemple, devinci.fr. gère les sous-domaines associés à ses réseaux IP (classes C):
  - 160.107.193.in-addr.arpa
  - 161.107.193.in-addr.arpa
- On dirige une requête de type PTR sur le numéro du host dans le domaine de `in-addr.arpa` associé au réseau du host (voir l'exemple utilisant ex 1)

Serveurs de Mail, attributs MX

- Mail Exchangers: serveurs de mail, capables de router du courrier electronique.
- attributs MX: les noms des serveurs de mail des utilisateurs de cette machine ou de ce domaine.
- Tout domaine et toute machine fournissant des adresses email doit avoir au moins un attribut MX
- Algorithme de choix du serveur de mail: le serveur courant supprime toutes les entrées de priorité

inférieure à la sienne pour éviter des boucles de routage, et s'adresse au serveur de plus haute preference MX, ou à defaut au Host lui-même.
- Voir pour plus de détails le chapitre consacré à SMTP.

Gestion d'un Domaine

- Outils: `h2n` (host-to-names: génère les fichiers de configuration du dns), `check-soa` (vérifie une configuration, voir l'ouvrage sur le DNS de Paul Albitz, Cricket Liu pour plus de détails)
- Création d'un Domaine: aspect Logique, Nom de Domaine, Serveurs publics,
- Aspect Physique, Enregistrement des numéros IP des réseaux IP
- Lancement : demon named.
- DNS, Email, et Firewalls: un serveur de mail accessible au travers d'un firewall permet l'échange de mails entre machines derrière le firewall et l'Internet. Souvent on ne permet pas de lister un domaine à partir de l'extérieur, si ce domaine contient des informations critiques
- Pour la gestion de la sécurité, suivre les avis du CERT(http://www.cert.org)

Email, SMTP, sendmail

Historique, fonctionnalités

- En 1979/80, TCP/IP commence à diffuser surt l'Arpanet. Le courrier électronique est acheminé au départ via ftp.
- 1980 voit la publication du protocole SMTP (RFC 821, 822, puis ultérieurement RFC 1123).
- En 1982 sendmail et TCP/IP sont diffusés gratuitement en source avec UNIX-Berkeley aux universités et laboratoires des USA.
- Aujourd'hui la quasi-unanimité des grands sites Internet tourne sendmail V8 (version 8) pour l'acheminement de leur courrier électronique.
- Sendmail constitue l'outil d'acheminement du système World-Wide de Courrier Electronique de l'Internet (l'"email").
- L'Email reste de nos jours l'outil fondamental de la communication électronique, il permet:
  - un service universel de communication entre personnes et groupes tres rapide et économique, de couverture mondiale
  - une gestion flexible et décentralisée de listes de distribution, d'adresses de renvoi,
  - des traitements automatiques des courriers: filtrage et routage par le CONTENU, réponses automatiques, gestion de souscriptions et désabonnements à des listes, interrogations de serveurs, transferts de fichiers...
  - le transport de documents de formats multimédia...

Architecture: user agents, sendmail, mailers

- Sendmail est un logiciel très largement configurable et programmable spécialisé dans les tâches d'acheminement du courrier électronique
- Pour Fournir un service Email complet il faut le compléter par :
  - des agents utilisateurs (ou mailers utilisateurs)
    - Mail, xmh, exmh, netscape (mozilla), pine, emacs (sur Unix)
    - eudora (Mac, PCs)
  - des agents de livraison (delivery agents)
    - /bin/mail, popd
- On obtient le schéma global suivant: schéma
- Les messageries proprietaires non-smtp nécessitent des passerelles (MSMAIL)

Introduction: Adresses, En-Tête, Corps, Enveloppe

- Le format des messages et des adresses email (ou smtp), sont spécifiés dans la RFC822
- Message : la partie utilisateur se compose de:
  - en-tête utilisateurs (from: to: cc: bcc: subject: reply-to: In-Reply-To: Precedence:) et transport (Return-Path: Received: Date:)
  - corps : données utilisateurs, pas de sémantique pour le système de mail, sauf markup MIME dans les messages "multipart"
- Une partie complémentaire est ajoutée lors du transport:
  - enveloppe (contient les destinataires effectifs de cette copie, pour le sendmail distant)
- Les Adresses sont locales si elles sont de la forme:
  - user
  - exemples:
    - tom
    - Jean-Paul.Martin
- Les Adresses sont non-locales si elles sont de la forme:
  - user@domain
  - exemples:
    - vh@mail.academie.fr
    - victor.hugo@academie.fr
- Toute adresse email peut comprendre un zone de commentaires. Deux formes standard:
  - Victor Hugo <vh@mail.academie.fr>
  - vh@mail.academie.fr (Victor Hugo)
- Exemples de messages rfc822

Aliases, .forward

- A la réception d'un message par sendmail (émission de mail par un process local, ou réception d'un mail via smtp), les noms de destinations locales sont comparées aux entrées du fichier système Aliases pour réécriture
- Les entrées d'Aliases sont de la forme:

```
# Commentaire, sur une ligne complète
```

```
# Alias simple : vers une adresse email
addr : addr1

# Alias : vers un fichier d'archivage
addr : /archive/path/name

# Alias : vers un processeur de mail (programme)
addr : " | /prog/path/name args..."

# Alias : vers une liste d'éléments des types précédents

addr : addr1, addr2, addr3

# Alias : vers une liste donnée par un fichier d'inclusion

addr : :include: /liste/path/name

# Alias : fin d'expansion (ne pas continuer à expanser addr1):

addr: \addr1
```

- En fin de processus d'expansion, si sendmail aboutit à un nom d'utilisateur local, il continue l'expansion par l'exploitation du fichier `.forward` de cet utilisateur (sauf adresses \user).
- fichiers `.forward` ou d'inclusion: contiennent une liste d'adresses (mêmes possibilités que dans les parties droites du fichier Aliases)

```
$ cat .forward
addr1
\user2
/archive/path/name
" | /prog/path/name args"
$
```

- **NB**: les fichiers d'aliases des agents utilisateurs (.mailrc) ne sont pas connus de sendmail, ils sont appliqués par l'UA pour la génération des chmps destination soumis à sendmail.
- **NB**: le fichier Aliases est compilé par sendmail, apres sa modification, il faut demander à sendmail de le recompiler (commande newaliases, ou sendmail -bi). Par contre les listes d'inclusion ne sont pas compilées, les modifications sont prises en compte sans autre traitement.

Applications du mécanisme d'Aliases

- mailing listes
- listes de distribution locales
- exploders
- programme vacation
- archives
- processeurs de mail : repondeurs (vacation), processeurs de mail (procmail), serveurs (ftpmail), gestionnaires de listes (majordomo, listprocessor)

Routage smtp

- Lors de chaque soumission de mail, sendmail examine les adresses de destination inscrites en enveloppe, effectue leur réécriture (règles: voir sendmail.cf et utilitaire ease) et leur expansion par Aliases.
- Il obtient alors une (ou des) destination
  - locale: transmise à un agent de livraison (/bin/mail, uucp, gateway x400...)
  - fichier: append
  - processeur: execution
  - distante: routage à nouveau via smtp, apres interrogation du DNS
- Les messages à router sont gardés en copie dans la queue ("`/var/spool/mqueue`") jusqu'à acheminement correct vers la destination d'enveloppe.
- Les messages à destination des utilisateurs locaux sont déposés dans la zone de "spool" locale par un agent de livraison (e.g. /bin/mail)
  - /var/spool/mail/user
- Les UA vont chercher le mail dans cette zone de spool
  - utilitaire "movemail"
  - service POP

Routage du mail et DNS

Les attribut MX du nom du domaine destinataire indiquent le serveur de mail associé à cette destination.

- sendmail élimine de la liste tous les mail exchangers de priorite inferieure ou egale à la sienne propre (indice MX superieurs au sien)
- il route le message vers le serveur de plus haute priorité restant (indice MX minimal), ou a défaut, vers le host lui-même
- si cette machine ne répond pas, sendmail essaie le serveur suivant, ou garde le mail pour une tentative ultérieure.

Service Post Office Protocol (POP3)

Ce protocole (standard Internet) permet à des machines incapables de tourner un serveur smtp d'accéder a une zone de spool utilisateur située sur un serveur de mail, via le réseau

- micros, stations
- réseau à faible latence recommandé: réseau local, ou accès PPP pas trop lent (sinon prévoir le saut des "gros messages" (images en MIME, attachements volumineux)
- effectue une identification du client par mot de passe.

Extensions MIME

- Multi-purpose Internet Mail Extensions
- Codages 8bits (iso-latin par exemple)
- Formats de données typées non textuelles: Images, Graphique, Vocal
- Ecodages vers l'ascii:
  - Quoted-printable (texte 8bits)
  - Base64 (binaire quelconque, remplace le classique uuencode)
- Multi-part : permet d'envoyer une suite de segments de types variés, et donc d'implémenter des "attachements".
- Exemples simples de messages MIME

Patrick.Duval@devinci.fr