OPTIMISER SON UTILISATION D'UNIX

BERNARD TATIN

BERNARD.TATIN@OUTLOOK.FR

RÉSUMÉ. Ce document vient des tréfonds de l'espace temps. Il a débuté il y a bien plus de trois ans de cela, repris de manière plus systématique et se trouve fortement complété aujourd'hui.

La première partie rappelle (rapidement) l'histoire et les concepts principaux des *SHELLS*. La deuxième partie est très orientée sur la recherche de *qui a piraté ma machine* mais peut être d'une grande utilité pour les débutants. La troisième partie, quant à elle, se focalise sur les scripts. Une quatrième partie donnera des notions des outils indispensables pour utiliser correctement son système **UNIX**.

Ce document et ses sources en ETEX sont disponibles sur GitHub.

Première partie	-a. l'histoire et les concepts	2
- a.1. Une his	toire d' Unix	2
- a.2. Les sн	ELLS	3
- a.2.1. Le fonctionnement		5
- a.2.2. Quelques shells célèbres		7
- a.2.2.1.	sh, le Bourne shell	7
- a.2.2.2.	csh, le C shell	7
<i>- a.</i> 2 <i>.</i> 2 <i>.</i> 3 <i>.</i>	tcsh ou le csh interactif	7
<i>- a.</i> 2 <i>.</i> 2 <i>.</i> 4 <i>.</i>	ksh, le Korn shell	7
- a.2.2.5.	zsh, le Z shell	7
- a.2.2.6.	bash, Bourne Again shell	8
Deuxième partie	e - b. la configuration et la ligne de commande	8
- b.1. La configuration		8
- b.1.1. Le shell personnel		8
- b.1.2. Configurer le prompt		9
- b.2. La ligne de commande		12
- b.2.1. Les boucles		12
- b.2.1.1.	La boucle for	12

Date: Novembre 2013/Novembre 2016.

- b.2.1.2. La boucle while	13
- b.2.2. Surprises avec stat, find et xargs	13
- b.2.2.1. stat	13
- b.2.2.2. Réfléchissons un peu	14
- b.2.2.3. Tous les fichiers du monde	15
- b.2.2.4. Application pratique	16
- b.2.3. Les surprises de sudo	18
- b.2.4. POSIX et GNU	20
Troisième partie -c. les scripts et les exemples	20
- c.1. Les scripts SHELL	20
- c.1.1. Structure des scripts	21
- c.1.2. Choisir son SHELL	22
- c.1.3. Les paramètres des scripts	22
- c.1.4. Tests et boucles	22
- c.1.5. Conditions, valeurs de retour des programmes	23
- c.1.6. Redirections et tubes (ou pipes)	23
- c.2. Exemples de manipulation de texte	24
- c.2.1. Des stats	24
- c.2.2. Peut-on faire mieux?	26
- c.2.2.1. Les options	26
- c.2.2.2. Avec bash	31

Table des matières

Première partie - a. l'histoire et les concepts

- A.1. UNE HISTOIRE D'**UNIX**

Voici une (rapide) histoire d'**UNIX**, choisie parmi d'autres, parmi celles qui évoluent avec le temps autant parce que des personnages hauts en couleur et ayant réussi à voler la vedette à de plus modestes collègues se font effacer eux-même par de plus brillants qu'eux, soit parce que, vieillissant ils se laissent aller à des confidences inattendues.

En nous basant sur Brève histoire d'**UNIX**, on rappelle que *AT&T* travaillait à la fin des années 60, sur un système d'exploitation **MULTICS** qui devait révolutionner l'histoire de l'informatique. Si révolution il y eut, ce fut dans les esprits : de nombreux concepts de ce système ont influencés ses successeurs, dont **UNIX**. Ken Thompson et Dennis

Ritchie des fameux *Bell Labs* et qui travaillaient (sans grande conviction, semble-t-il) sur **MULTICS**, décidèrent de lancer leur propre projet d'OS :

baptisé initialement UNICS (UNiplexed Information and Computing Service) jeu de mot avec "eunuchs' (eunuque) pour "un **MULTICS** emasculé", par clin d'œil au projet **MULTICS**, qu'ils jugeaient beaucoup trop compliqué. Le nom fut ensuite modifié en **UNIX**¹.

L'essor d'**UNIX** est très fortement lié à un langage de programmation, le C. À l'origine, le premier **UNIX** était écrit en assembleur, puis Ken Thompson crée un nouveau langage, le B. En 1971, Dennis Ritchie écrit à son tour un nouveau langage, fondé sur le B, le C. Dès 1973, presque tout **UNIX** est réécrit en C. Ceci fait probablement d'**UNIX** le premier système au monde écrit dans un langage portable, c'est-à-dire autre chose que de l'assembleur².

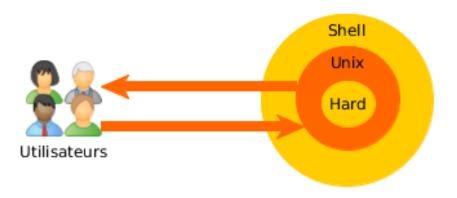
Ce que j'ai surtout retenu de tout cela, c'est qu'**UNIX** a banalisé autant l'utilisation des stations de travail connectées en réseau que le concept de *shell*, des système de fichiers hiérarchisés, des périphériques considérés comme de simples fichiers, concepts repris (et certainement améliorés) à **MULTICS** comme à d'autres. Pour moi, la plus grande invention d'**UNIX**, c'est le langage C qui permet l'écriture des systèmes d'exploitations et des logiciels d'une manière très portable. N'oublions pas qu'aujourd'hui encore, C (mais pas C++) est un des langages les plus portable, même s'il commence à être concurrencé par Java par exemple.

- A.2. LES SHELLS

Un *SHELL* est une *coquille*, pour reprendre la traduction littérale, autour du système d'exploitation. Voici un magnifique diagramme (d'après ce que l'on trouve sur le WEB comme dans d'anciens ouvrages) donnant une idée du concept :

^{1.} cfl'article Multics de Wikipedia

^{2.} *cf* Brève histoire d'**UNIX**



shell **UNIX** Source: le WEB, ouvrages divers

Entre mes débuts dans le monde de l'informatique et aujourd'hui, le concept de SHELL a quelque peu évolué. Certains qualifient l'explorateur de Windows comme un SHELL. Ont-ils raison? Certainement si l'on se réfère à l'image précédente : nos commandes (clique, clique et reclique) envoyée au shell graphique sont transmises au noyau qui nous renvoie, par l'intermédiaire du shell graphique, de belles images. Il faut avouer que l'explorateur Windows est le premier contact que l'utilisateur a avec sa machine. Et sur l'article interface système de Wikipedia, on trouve cette définition :

Une interface système (SHELL en anglais) est une couche logicielle qui fournit l'interface utilisateur d'un système d'exploitation. Il correspond à la couche la plus externe de ce dernier.

Ce même article cite les:

shells graphiques fournissant une interface graphique pour l'utilisateur (GUI, pour Graphical User Interface)

Dans le monde UNIX, le concept de SHELL reste plus modeste, même si Midnight Commander (mc) est parfois considéré comme un SHELL:



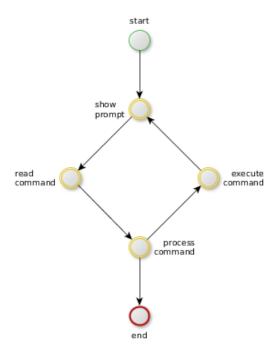
mc dans une session Cygwin Source: mon PC

Pour nous et dans tout ce qui suit, nous considérons comme SHELL:

un interpréteur de commandes destiné aux systèmes d'exploitation **UNIX** et de type **UNIX** qui permet d'accéder aux fonctionnalités internes du système d'exploitation. Il se présente sous la forme d'une interface en ligne de commande accessible depuis la console ou un terminal. L'utilisateur lance des commandes sous forme d'une entrée texte exécutée ensuite par le SHELL. Dans les différents systèmes **WINDOWS**, le programme analogue est **COMMAND. COM OU CMO. EXE.**

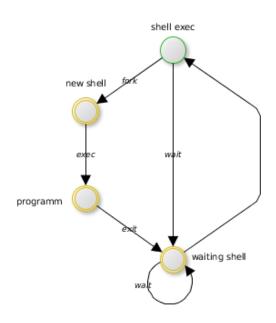
cf SHELL **UNIX** sur Wikipedia

- a.2.1. **Le fonctionnement.** Le fonctionnement général est assez simple, surtout si l'on ne tient pas compte de la gestion des erreurs comme dans le graphique suivant qui peut être appliqué à tout bon interpréteur. Seuls les détails de **process command** et **execute command** vont réellement changer.



SHELL : fonctionnement général Source: créé avec yEd

L'exécution d'un programme suit l'algorithme :



SHELL: exécution d'un programme Source: créé avec yEd

À noter que la commande **exec** se comporte différemment : elle correspond à l'appel système **exec** .

- a.2.2. Quelques shells célèbres.

- a.2.2.1. sh, le Bourne SHELL. L'ancêtre, toujours vivant et avec lequel sont écrits une grande majorité des scripts actuels. Son intérêt essentiel est justement l'écriture de scripts. Pour l'interaction, il est absolument nul mais bien utile parfois pour dépanner.
- a.2.2.2. CSh, le C SHELL. Il se voulait le remplaçant glorieux de l'ancêtre sh avec une syntaxe considérée plus lisible car proche du C. Il est de plus en plus abandonné y compris par ses admirateurs les plus fervents, vieillissants dans la solitude la plus complète. Essayez d'écrire un script en csh d'un peu d'envergure sans faire de copié/collé! Il n'y a en effet pas de possibilité de créer des fonctions et, ce qui gêne peut-être encore plus les administrateurs système, il n'y a pas de gestion d'exception. Cependant, il fût certainement le premier à proposer l'historique des commandes.

A noter qu'il fût crée par Bill Joy, l'un des fondateurs historiques de la société Sun Microsystems.

-a.2.2.3. tosh ou le osh interactif. Le pendant interactif du précédent. Il lui reste des afficionados qui aiment bien sa gestion de l'historique et de la ligne de commande. Il est une extension de osh, i.e.tout ce qui peut-être fait par osh est fait par tosh. Sur de nombreux systèmes (Mac OS X entre autre), ces deux shells pointent sur le même exécutable (avec un lien symbolique).

En séquence nostalgie, je me souviens que c'est ce SHELL interactif que j'utilisais sur mon premier **UNIX**, en 87/88.

- a.2.2.4. ksh, le Korn shell. Initialement écrit pour **Unix** par David Korn au début des années 80, ce shell a été repris par Microsoft pour Windows. Compatible avec sh, il propose de nombreuses avancées comme beaucoup de fonctionnalités de tcsh, des fonctions, des exceptions, des manipulations très évoluées de chaînes de caractères,
- a.2.2.5. zsh, le Z shell. C'est mon préféré pour l'interactivité, la complétion et bien d'autres choses encore dont il est capable depuis sa création ou presque. Comme ksh, il est compilable en bytecode et propose des bibliothèques thématiques comme la couleur, les sockets, la gestion des dates...

- a.2.2.6. bash, Bourne Again shell. C'est le descendant le plus direct de sh. C'est
certainement le SHELL le plus répandu dans le monde Linux aujourd'hui.

Lors de ma découverte de Linux, je l'ai vite abandonné car il était très en retard pour la complétion en ligne de commande par rapport à d'autres, y compris t csh qui commençait pourtant à vieillir un peu. Il a fallu beaucoup d'années (pratiquement 10) pour qu'il en vienne à peu près au niveau de zsh.

Aujourd'hui, c'est le *SHELL* par défaut de nombreuses distributions Linux et il commence à devenir très utilisé comme *SHELL* de script par défaut.

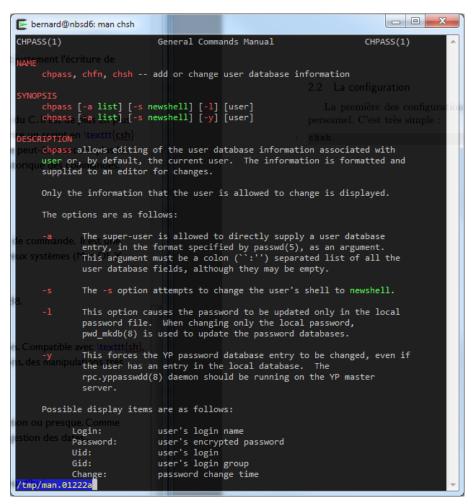
Deuxième partie -b. la configuration et la ligne de commande

- B.1. LA CONFIGURATION

- *b*.1.1. **Le shell personnel.** La première des configuration est le choix de son shell par défaut sur son compte personnel. C'est très simple :

1 chsh

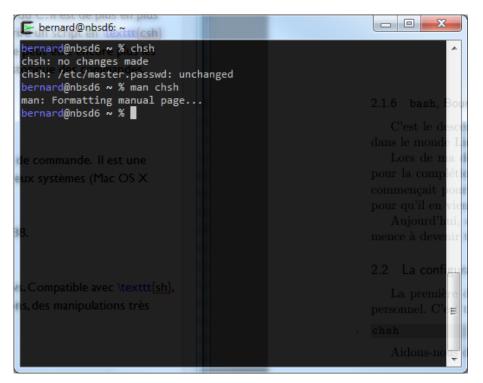
Aidons-nous du manuel (sous **NETBSD**):



man chsh sous NETBSD

Source: ma machine virtuelle

- *b*.1.2. **Configurer le prompt.** Sur ma machine virtuelle **NETBSD**, j'obtiens quelque chose comme ceci :



un prompt sous **NETBSD**, avec zsh Source: ma machine virtuelle

Le prompt, ce sont les caractères colorés que l'on voit en début de chaque lignes de commande. Ce prompt m'a aidé, voire sauvé plusieurs fois. Celui-ci m'affiche le nom de l'utilisateur courant en bleu, de la machine en blanc et du répertoire courant en blanc et gras. Lorsque j'ai des sessions sur plusieurs machines, je vois tout de suite où je me trouve avec son nom. Ensuite, lorsque je me déplace de répertoires en répertoires, je n'ai pas besoin de faire d'éternels *pwd* pour savoir où je me trouve. En plus, lorsque je trouve dans un dépôt SVN, j'ai un affichage me donnant les indications sur le répertoire de travail (on ne peut pas le faire sous **Cygwin**):

```
bernard@nbsd6: ~/tmp/AlimPIC12.X-2 010
     AlimPIC12.X-2/AlimPIC12.X/nbproject/Package-TIME 5 5 LEDS.bash
    AlimPIC12.X-2/AlimPIC12.X/nbproject/Makefile-local-TIME_5_5_LEDS
    AlimPIC12.X-2/AlimPIC12.X/nbproject/Makefile-TIME_5_5.mk
AlimPIC12.X-2/AlimPIC12.X/nbproject/Makefile-variables.mk
    AlimPIC12.X-2/AlimPIC12.X/nbproject/Package-default.bash
    AlimPIC12.X-2/AlimPIC12.X/nbproject/configurations.xml
    AlimPIC12.X-2/AlimPIC12.X/nbproject/Makefile-local-default.mk
    AlimPIC12.X-2/AlimPIC12.X/nbproject/Makefile-TIME_5_5_LEDS.mk
    AlimPIC12.X-2/AlimPIC12.X/nbproject/project.properties
    AlimPIC12.X-2/AlimPIC12.X/nbproject/Makefile-genesis.properties
    AlimPIC12.X-2/AlimPIC12.X/nbproject/Makefile-default.mk
     AlimPIC12.X-2/AlimPIC12.X/nbproject/project.xml
    AlimPIC12.X-2/AlimPIC12.X/nbproject/private
    AlimPIC12.X-2/AlimPIC12.X/nbproject/private/private.properties
    AlimPIC12.X-2/AlimPIC12.X/nbproject/private/configurations.xml
    AlimPIC12.X-2/AlimPIC12.X/nbproject/private/private.xml
    AlimPIC12.X-2/AlimPIC12.X/nbproject/Package-TIME 5 5.bash
    AlimPIC12.X-2/AlimPIC12.X/funclist
    AlimPIC12.X-2/AlimPIC12.X/configuration_bits.c
     AlimPIC12.X-2/AlimPIC12.X/main.h
    AlimPIC12.X-2/AlimPIC12.X/system.h
     AlimPIC12.X-2/AlimPIC12.X/Makefile
Checked out revision 280.
pernard@nbsd6 ~/tmp % cd AlimPIC12.X-2
pernard@nbsd6 ~/tmp/AlimPIC12.X-2 (svn)-[AlimPIC12.X-2:280] %
```

dans un répertoire de travail **SVN**, avec zsh Source: ma machine

Pour finir, le nom de l'utilisateur change de couleur lorsque je suis en **root** :



en root avec zsh
Source: ma machine

Tous les shells interactifs de ma connaissance ont au moins un fichier de configuration exécuté au lancement: avec zsh, c'est . zhrc, avec bash, c'est . bashrc et avec csh

, c'est . cshrc . Aussi loin que mes souvenirs remontent, on personnalise le prompt avec la variable PS1 et ce, même pour le MS/DOS.

Voici un hexdump de mon **PS1**:

- B.2. LA LIGNE DE COMMANDE

Pour de multiples raisons déjà plus ou moins évoquées plus haut, j'ai choisi de travailler avec zsh comme shell par défaut. C'est ce que nous allons faire ici, autant sous **FreeBSD** que sous **Linux**, tout simplement pour prouver que l'utilisation du shell est assez indépendante du système sous-jacent. Mais comme je sais que certains systèmes viennent avec bash ou tcsh, sans possibilité de modification, je les évoqueraient donc, en particulier tcsh qui est utilisé très souvent, avec csh, pour l'administration. Ce n'est que fortuitement que j'examinerais ksh, autant par manque d'habitude que parce que je ne l'ai jamais rencontré.

- b.2.1. **Les boucles.** Il y a **while** et **for**.

- b.2.1.1. La boucle **for**. On commence par celle-ci car elle en a dérouté plus d'un. Nous avons, avec zsh et bash, deux syntaxes essentielles. La première parcourt un ensemble de données :

```
for file_name in *.txt

do

cat $file_name

done
```

Il y a la boucle plus classique pour les spécialistes de Java :

```
1 for ((i=5; i< 8; i++))
2 do
3     echo $i
4 done</pre>
```

Avec tcsh, nous aurons:

```
foreach file_name (*.txt)

cat $file_name

end

foreach i ('seq 5 1 8')

echo $i

end
```

- b.2.1.2. La boucle **while**. Elle permet de boucles infinies comme celle-ci avec zsh, bash et ksh:

```
while true; do date ``+%T''; sleep 1; done
```

Avec tcsh, nous écrirons en deux lignes:

```
while (1); date ``+%T''; sleep 1;
end
```

- b.2.2. Surprises avec stat, find et xargs.
- b.2.2.1. **stat**. La commande **stat** permet de connaître bon nombre de détails à propos d'un fichier comme ici :

```
bernard@debian7 ~ % stat *
install:
device 70
inode 1837064

mode 16877
nlink 3
uid 1000
gid 1000
rdev 7337007
size 512
atime 1383862259
mtime 1383085660
```

```
ctime 1383085660
  blksize 16384
15 blocks 4
  link
17
  userstart.tar.gz:
  device 70
  inode
          1837156
21 mode 33188
22 nlink
 uid 1000
24 gid 1000
25 rdev 7371584
26 size 3498854
27 atime 1383855566
28 mtime
          1383855552
29 ctime 1383855552
30 blksize 16384
31 blocks 6880
32 link
```

On obtient, sous zsh, un résultat totalement identique sous **NETBSD** et sous **LINUX**. Si l'on fait un **which stat**, nous obtenons, sur les deux systèmes, le message **stat**: **shell built-in command**. C'est ce qui me plait sous zsh, les commandes non standard comme **stat** sont remplacées par des fonctions dont le résultat ne réserve pas de surprise. Si je veux une sortie plus agréable et n'afficher que la date de dernière modification (cf The zsh/stat module pour de plus amples explications):

```
bernard@debian7 ~ % stat -F "%Y-%m-%d %T" +mtime -n *
install 2013-10-29 23:27:40

userstart.tar.gz 2013-11-07 21:19:12

bernard@debian7 ~ %

bernard@NBSD-64bits ~ % stat -F "%Y-%m-%d %T" +mtime -n *
install 2013-11-14 09:52:56

userstart.tar.gz 2013-11-08 00:54:06

bernard@NBSD-64bits ~ %
```

- b.2.2.2. Réfléchissons un peu. Grâce à zsh, nous avons une méthode extrêmement portable entre Unix pour afficher des données détaillées des fichiers. Pour l'exemple, prenons le **stat** d'origine:

```
1 bernard@debian7 ~ % /usr/bin/stat --printf="%n %z\n" *
2 install 2013-10-29 23:27:40.000000000 +0100
3 userstart.tar.gz 2013-11-07 21:19:12.0000000000 +0100
4 bernard@debian7 ~ %
1 bernard@NBSD-64bits ~ % /usr/bin/stat -t "%Y-%m-%d %T" -f "%Sc %N" *
2 2013-11-14 09:52:56 install
3 2013-11-08 00:54:06 userstart.tar.gz
4 bernard@NBSD-64bits ~ %
```

- b.2.2.3. Tous les fichiers du monde. Si je veux faire la même chose que précédemment, mais sur tous les fichiers de la machine, on peut tomber sur ce message d'erreur :

```
bernard@debian7 ~ % stat -F "%Y-%m-%d %T" +ctime -n $(find / -name "*
   ")
zsh: liste d'arguments trop longue: stat
bernard@debian7 ~ %
```

C'est là que xargs entre en jeu, mais avec un nouveau problème :

La commande xargs va chercher non pas la fonction de zsh mais le binaire qui se trouve sur le PATH de la machine. On doit donc faire :

```
bernard@debian7 ~ % find . -name "*" | xargs stat --printf="%n %z\n"

...

./.w3m 2013-11-07 23:07:03.000000000 +0100

./.w3m/configuration 2013-11-07 23:03:48.000000000 +0100

./.w3m/history 2013-11-07 23:06:29.000000000 +0100

./.w3m/cookie 2013-11-07 23:06:29.000000000 +0100

./.viminfo 2013-11-07 23:07:03.000000000 +0100

bernard@debian7 ~ %
```

Sous **NETBSD**:

- *b*.2.2.4. *Application pratique*. Sur le serveur, qui est sous **LINUX**, faisons la même chose ou presque, on place la date en premier et c'est la surprise du jour :

```
[bigserver] (688) ~ % find /etc -name "*" | xargs stat --printf="%z
%n\n" | sort

find: "/etc/ssl/private": Permission non accordee

stat: option invalide -- 'o'

Pour en savoir davantage, faites: stat --help .

[bigserver] (689) ~ %
```

En rajoutant l'option -print0 à find, l'option -0 à xargs, nous obtenons le bon résultat:

En fait, les noms de fichier sous **UNIX** peuvent contenir beaucoup de caractères étranges en dehors de /. **xargs** prend le caractère **LF** comme fin d'enregistrement de la part de son entrée standard. Si jamais un fichier contient ce caractère, plus rien ne va. Les nouvelles options permettent à **find** d'utiliser **0x00** comme séparateur d'enregistrement et à **xargs** de bien l'interpréter.

Il y a aussi une autre explication, depuis bien longtemps les outils **GNU** fonctionnent comme ceci et ce n'est que très récemment que le couple **find**/**xargs** fonctionne ainsi.

Après toutes ces considérations, on constate que le 14 Novembre 2013, un peu après 1 heure du matin, quelqu'un a modifié les fichiers /etc/php5/conf.d/mysqli.ini et /etc/php5/conf.d/mcrypt.ini, tout simplement pour remplacer les commentaires de type shell par des commentaires de type fichier ini.

Après une attaque du serveur, il est intéressant de faire le même exercice sur les répertoires vitaux comme /bin. Pour éviter des listings trop important, on limite la sortie à l'année 2013 et on fait une jolie boucle :

```
[bigserver] (694) ~ % for d in /bin /sbin /lib /lib32 /usr/bin /usr/
    sbin /usr/lib /usr/lib32; do

find $d -name "*" -print0 | xargs -0 stat --printf="%z %n\n" |
    egrep "^2013"

done | sort
```

Nous obtenons un listing fort long, correspondant aux mises à jour faites le 8 et le 12 Novembre. Maintenant que nous savons que l'attaque a eu lieu avant le 8 Novembre, on sélectionne plus sévèrement :

Cependant, rien ne prouve que nous n'avons pas eu de désordres un peu avant ou un peu pendant. Comme le gros des fichiers est dans /usr/lib, éliminons le de la liste:

```
[bigserver] (695) ~ % for d in /bin /sbin /lib /lib32 /usr/bin /usr/
    sbin; do

find $d -name "*" -print0 | xargs -0 stat --printf="%z %n\n" |
    egrep "^2013"

done | sort
```

```
4 ...
5 [bigserver] (696) ~ %
```

Nous n'avons des modifications qu'entre le 8 et le 12 Novembre.

Plus fort encore, afficher les fichiers modifiés ce jour :

On est débordé par l'affichage des fichiers système de Linux. Pour palier à cet inconvénient, on demande à **find** d'abandonner les répertoire **/sys** et **/proc** :

- *b*.2.3. **Les surprises de** *sudo* **.** Reprenons l'exemple précédent en redirigeant la sortie standard vers / *dev/nul1* :

On aura une sortie comme celle-ci:

```
find: "/var/lib/postgresql/9.1/main": Permission non accordee

find: "/var/lib/sudo": Permission non accordee

find: "/var/cache/ldconfig": Permission non accordee

find: "/var/log/exim4": Permission non accordee

find: "/var/log/apache2": Permission non accordee

...
```

Pour éliminer les find: ... Permission non accordee, on utilise sudo:

```
sudo find / \( -path /proc -o -path /sys \) -prune -o -name "*" -
print0 | xargs -0 stat --printf="%z %n\n" | egrep "^$(date '+%Y-%
m-%d')" > /dev/null
```

C'est pire:

```
stat: impossible d'evaluer ' /root/.aptitude/config ': Permission non
    accordee

stat: impossible d'evaluer ' /root/.viminfo ': Permission non
    accordee

stat: impossible d'evaluer ' /root/.bash_history ': Permission non
    accordee

...
```

Nous avons demandé à **sudo** de traité **find** et avec le *pipe*, nous demandons à **xargs** de traiter les lignes de sorties avec **stat**. Ce dernier récupère un nom de fichier et le traite comme tel mais comme il n'est pas lancé avec **sudo**, nous avons ces erreurs. Essayons ceci:

C'est pas mieux, autant sous LINUX que sous NETBSD:

```
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
sudo: unable to execute /usr/bin/stat: Argument list too long
```

Il faut bien l'avouer, je ne sais pas quoi dire de plus ici - sinon noter un TODO: comprendre ce qui ce passe. Ce qui est bien avec **UNIX**, c'est qu'il y a toujours un moyen de s'en sortir. On remarquera quelques différences entre les mondes **LINUX** et **BSD**, en particulier avec **man sh**, où le premier nous renvoie sur bash alors que le second traite bien directement de sh. Dans tous les cas, **sudo sh -c** ''...' est notre amie et nous obtenons avec **NETBSD**:

Et sous **LINUX**:

- b.2.4. **POSIX et GNU.** Profitons d'un moment de calme pour remarquer que de nombreuses commandes se comportent de manière très standard entre différents systèmes, y compris parfois, sous **MS/DOS**. Tout cela vient de **POSIX** ou de **GNU**.

Les guerres de religions qui opposent parfois violemment les mondes **BSD** et **Linux**, les supporters de **Vi** et **Emacs**, ... finissent par être absorbées avec le temps et seuls quelques irréductibles les raniment, souvent plus pour s'exposer aux yeux (blasés maintenant) du petit monde concerné. Seule reste l'opposition farouche entre tenants du libre et leurs opposants.

Ici, nous avons utilisé **find** de la même manière sous **LINUX** et sous **NETBSD** ce qui n'a pas été toujours le cas, de même, sh se comporte de manière identique à quelques octets près sur les deux systèmes, ce qui n'était pas forcément vrai il y a quelques années. Pour revenir à **find**, nous avions un paquet de compatibilité **GNU** disponible sur plusieurs **BSD** qui reprenait le **find** que nous connaissons maintenant et l'on pouvait différencier **gfind** de **bsdfind** ³.

Troisième partie -c. les scripts et les exemples

- C.1. LES SCRIPTS SHELL

La magie des *SHELLS* est infinie, ils nous permettent en effet de créer des programmes complets, complexes... parfois aux limites du lisible. On les appelle *scripts* pour les opposer aux applications généralement créées à partir de langages compilés mais cela ne

^{3.} A vérifier dans les détails.

devrait rien changer au fait qu'ils doivent être conçus avec un soin égal à celui apporté aux autres langages comme *C/C++, Java...*

Dans tout ce qui suit, il ne faut pas perdre de vue que le *SHELL*est une *coquille* entourant le noyau d'**UNIX**. Certains aspects des *SHELLS*ne font que recouvrir des appels systèmes.

- c.1.1. **Structure des scripts.** Ce qui est décrit ici est valable autant pour des langages interprétés comme l'horrible *Perl* ⁴, le sublime *Scheme* ⁵, le célèbre *Python* que pour n'importe quel *shell*.

La première ligne d'un script est le *shebang*. Cette ligne est très importante car elle indique de manière sûre quel interpréteur il doit utiliser pour exécuter le corps du script. Voici quelques exemples :

```
sh: :#!/bin/sh
bash: :#!/bin/sh
Perl: :#!/usr/bin/env perl
Python 2.7: :#!/usr/bin/env python2.7
Python: :#!/usr/bin/env python
```

awk::#!/bin/awk -f

Les deux caractères #! sont considérés comme un nombre magique par le système d'exploitation qui comprend immédiatement qu'il doit utiliser le script dont le nom et les arguments suivent les deux caractères.

Dans un SHELLinteractif, l'exécution d'un script suit l'algorithme suivant :

```
fork ();
if (child) {
   open(script);
   switch(magic_number) {
      case 0x7f'ELF':
        exec_binaire();
      break;
      case '\#!':
      load_shell(first_line);
      exec_binaire(shellname, args);
      break;
```

^{4.} je ne suis pas objectif, mais quand même...

^{5.} là, je me sens plus objectif... ou presque.

- c.1.2. **Choisir son SHELL.** Par tradition autant que par prudence, on écrit ses scripts SHELLpour le SHELLd'origine, soit Sh. Par prudence car on est certain qu'il sera présent sur la machine même si elle démarre en mode dégradé. Cependant, beaucoup de scripts sont applicatifs et ne pourront pas fonctionner en mode dégradé. Autant se servir d'un SHELLplus complet comme bash.
- -c.1.3. Les paramètres des scripts. Les paramètres, leur nombre et leur taille n'ont de limites que de l'ordre de la dizaine de Ko. Il faut donc pouvoir y accéder. Le paramètre \$0 est le nom du script parfois avec le nom du répertoire. Les neufs suivants sont nommés \$1, ..., \$9. Pour accéder aux autres il faut ruser un peu avec l'instruction shift.
- c.1.4. **Tests et boucles.** Les tests se font avec *if* de cette manière :

```
if condition

then

if condition

file

file

if condition

then

if condition

i
```

Dans le même ordre d'idée, nous avons le while:

```
while condition
do
do
done
```

La construction des *condition* est tout un art, d'autant plus qu'en lieu et place du *if* nous pouvons écrire :

```
ou bien:

condition && condition_true && ...

ou bien:

condition || condition_false
```

Nous avons aussi une boucle for:

```
1 for index in ensemble
2 do
3  ...
4 done
```

La détermination de **ensemble** est assez naturelle comme par exemple avec **\$ (1s** *. java) . Mais il faut être prudent : selon les SHELLSles résultats peuvent différer.

-c.1.5. **Conditions, valeurs de retour des programmes.** Tout les programmes sous **UNIX** s'achèvent par un **return EXIT_CODE** ou bien un **exit (EXIT_CODE)** bien senti. La valeur **EXIT_CODE** est renvoyée au programme appelant, notre **SHELL**. On peut le récupérer depuis la variable **\$#?** puis étudier le cas :

```
myprogram arg1 arg2 ...
   case \$\#? in
     0)
3
     its-okayyy
       1)
           bad_parameterzzz
           cant-open-filezzz
10
11
12
     *)
13
           unknow-error
14
           ;;
15
```

UNIX considère que la valeur de retour *0* est signe que tout va bien et que tout autre valeur exprime une condition d'erreur. On peut donc utiliser cette propriété ainsi :

```
myprogram arg1 arg2 ... || onerror '`Error code \$\#?''
```

- c.1.6. **Redirections et tubes (ou pipes).** Le premier piège dans l'utilisation des tubes dans un scripts est simple : pour chaque tube, on crée un nouveau processus. Ainsi le script suivant ne renvoie pas le résultat escompté :

```
1 #!/bin/sh
2
3 compteur=0
4
```

```
1s -l /bin | while read line; do
6    compteur=$(( compteur + 1 ))
7  done
8
9  printf "Il y a %d fichiers dans /bin\n" ${compteur}

1  $ ./badl.sh
2  Il y a 0 fichiers dans /bin
```

La variable compteur fait partie de l'environnement de **bad1** . **sh** . Lorsque la boucle **while** se lance, elle est dans un nouveau contexte et sa modification se perd à la fin de la boucle. On corrige de cette manière :

```
#!/bin/sh

compteur=0

ls -l /bin > /tmp/ls-l.tmp

while read line; do

compteur=$((compteur + 1))

done < /tmp/ls-l.tmp

printf "Il y a %d fichiers dans /bin\n" ${compteur}

./not-so-badl.sh

Il y a 173 fichiers dans /bin</pre>
```

- C.2. EXEMPLES DE MANIPULATION DE TEXTE

- c.2.1. **Des stats.** Voici un extrait d'un fichier /var/log/messages :

Nous voulons déterminer les moments les plus actifs de ce fichier avec une granularité de une heure. La manipulation est simple :

```
afficher le fichier: cat file-name,
découper le ficher: cut -d':'-f1,
trier le fichier: sort,
compter les occurences: uniq-c,
trier en décroissant: sort-n.
```

Ce qui nous donne la commande :

```
cat $file-name |
cut -d ':' -f 1 |
sort |
uniq -c |
sort -n
```

On obtient rapidement un script (**stat1.sh**) à partir de cette ligne de commande:

```
#!/bin/sh
   scriptname="$(basename $0)"
  dohelp() {
       cat << DOHELP
   ${scriptname} [-h|--help] : this text
  ${scriptname} file file\ldots : stats
   DOHELP
       exit 0
10
11
12
  [ $# -eq 0 ] && dohelp
13
   case $1 in
14
15
       -h | --help)
            dohelp
16
17
            ;;
       *)
18
           cat "$@" | \
19
20
```

On peut tester:

```
1 $ ./statl.sh messages-1 /var/log/messages /var/log/messages.1

2 ...

3 152 Nov 4 10

4 155 Oct 27 11

5 156 Oct 28 09

6 164 Oct 28 17

7 186 Oct 25 14

8 213 Oct 28 11

9 216 Nov 3 10

10 260 Oct 28 10

11 636 Oct 26 15

12 774 Oct 28 07

13 1770 Nov 3 09

14 3844 Oct 28 14

15 26201 Oct 28 15
```

- c.2.2. **Peut-on faire mieux?** Bien sûr! On peut avoir d'autres options que la simple aide, on peut aussi gérer correctement les erreurs, les *signaux* **UNIX**...
- -c.2.2.1. Les options. Depuis longtemps il existe une norme **POSIX** permettant de gérer les options de la ligne de commande. Malheureusement, il fut une époque où la norme avait beaucoup de variantes ce qui m'a poussé à faire ma propre gestion de ces paramètres.

Voici ma méthode, facile à mémoriser mais pas parfaite et un peu lourde :

l'aide: créer une fonction **dohelp** comme dans l'exemple précédent; le nom **dohelp** permet d'éviter un clash avec une éventuelle commande **help**.

s'assurer de l'existence de paramètres: il suffit de faire le test [\$# -eq 0] et exécuter le code nécessaire.

vider la liste des paramètres: une boucle while [\$# -ne 0] fait l'affaire.

Voici un exemple plus parlant (script **stat2.sh**):

```
#!/bin/sh
2
   scriptname="$(basename $0)"
3
4
   dohelp() {
5
       cat << DOHELP
   \{scriptname\} [-h|--help] : this text
   ${scriptname} [options] file file...: stats
   options:
       -s|--size N: number of most important hours, default 5
10
       -b|--byhour : for each hours
11
12
   exit 0
13
14
15
   size=5
16
  byhour="cut -d ':' -f 1"
17
  is byhour=0
   after=""
19
20
   set_size() {
21
       [ "$1" -lt 1 ] && onerror 3 "size must be > 1"
22
       size=$1
23
24
   }
   set_byhour() {
25
       byhour="${byhour} | tr -s ' ' | cut -d ' ' -f 3"
26
       after=" | sort -k 2"
27
     is_byhour=1
28
       set_size 24
29
30
   doit() {
31
32
       end=1
       cmd="cat $0 | ${byhour} | sort | uniq -c | sort -nr | head -n ${
33
           size} ${after}"
       case ${is_byhour} in
34
          0)
35
               printf "%-7.7s %-6.6s %-2.2s\n" "occurences" "date" "hour
36
               printf "%-17.17s\n" "---
37
38
39
           1)
```

```
printf "%-7.7s %-2.2s\n" "occurences" "hour"
               printf "%-10.10s\n" "-----
41
42
       esac
43
       eval "${cmd}"
   [ $# -eq 0 ] && dohelp
  while [ $end -eq 0 ]
51
52
       case $1 in
53
           -h | --help)
               dohelp
55
               ;;
          -s | --size)
56
               shift
57
               [ $# -eq 0 ] && onerror 2 "$1 needs a parameter"
58
               set_size "$1"
59
               shift
60
61
           -b | --byhour)
62
               shift
63
               set_byhour
           *)
66
               doit "$@"
               ;;
       esac
  done
```

Et maintenant avec le **getopts** 6:

```
#!/bin/sh

criptname=$(basename $0)

dohelp() {
    cat << DOHELP
    ${scriptname} [-h] : this text</pre>
```

^{6.} soyez prudents avec les (très) anciennes versions de Red Hat

```
8 ${scriptname} [options] file file\ldots : stats
  options:
       -s N : number of most important hours, default 5
10
       -b : for each hours
11
   DOHELP
12
       exit 0
13
14
   }
   onerror() {
15
       local exit_code=$1
16
       shift
17
       local error msq="$@"
18
19
       echo "ERROR: $error_msg" 1>&2
20
       exit "$exit_code"
21
22
23
24
  byhour="cut -d ':' -f 1"
25
  is_byhour=0
26
   after=""
27
28
   set_size() {
29
       [ "$1" -lt 1 ] && onerror 3 "size must be > 1"
30
     size=$1
31
32
   set_byhour() {
33
       byhour="${byhour} | tr -s ' ' | cut -d ' ' -f 3"
34
       after=" | sort -k 2"
35
       is_byhour=1
36
       set_size 24
37
38
39
   doit() {
       cmd="cat $0 | ${byhour} | sort | uniq -c | sort -nr | head -n ${
40
           size} ${after}"
       case ${is_byhour} in
41
          0)
42
               printf "%-7.7s %-6.6s %-2.2s\n" "occurences" "date" "hour
43
                printf "%-17.17s\n" "---
44
45
           1)
46
```

```
printf "%-7.7s %-2.2s\n" "occurences" "hour"
47
                 printf "%-10.10s\n" "-
48
49
50
        esac
        eval "${cmd}"
51
53
   # [ $# -eq 0 ] && dohelp
56
   [ $# -eq 0 ] && echo "you need arguments" && dohelp
57
   while getopts "s:bh" opt
58
59
        case $opt in
60
61
62
                 dohelp
63
            s)
64
                 set_size "$OPTARG"
65
66
            b)
67
                 set_byhour
68
69
70
                 onerror 2 "$OPTARG needs a parameter"
71
72
            \?)
73
74
                 onerror 7 "option $OPTARG is unknown"
76
        esac
77
   done
78
   shift $((OPTIND-1))
   doit "$@"
```

En fait **getopts** ne sait traiter que les options courtes et classiques d'**Unix**. Les options longues à la mode **Linux** ne sont pas supportées. L'avantage de **getopts** est son mode de fonctionnement assez simple. Son inconvénient principal est d'être très spécifique à bash même si **POSIX** le soutient, ce qui fait qu'il n'est pas forcément disponible partout.

Pour avoir les *options longues*, il faut utiliser l'outil **GNU** *getopt* (sans le *s* de fin) ⁷. Je reste donc sur ma méthode qui n'est finalement ni meilleure ni pire.

- c.2.2.2. Avec **bash**. On peut profiter des avantages de bash (boucles, tableaux, ...) comme dans ce script (qui ne fonctionne **pas** avec sh):

```
#!/usr/bin/env bash
   scriptname="$(basename $0)"
3
4
   dohelp() {
       cat << DOHELP
6
   \{scriptname\} [-h|--help] : this text
   ${scriptname} [options] file file\ldots : stats
9
   options:
       -s \mid --size N: number of most important hours, default 5
10
       -b|--byhour : for each hours
11
   DOHELP
12
       exit 0
13
14
15
  size=5
  byhour="cut -d ':' -f 1"
17
  is_byhour=0
  after=""
  hours=
21
22
23
       [ "$1" -lt 1 ] && onerror 3 "size must be > 1"
       size=$1
24
25
   set_byhour() {
26
27
       local i
       byhour="${byhour} | tr -s ' ' | cut -d ' ' -f 3"
28
       after=" | sort -k 2"
29
       is_byhour=1
30
       set_size 24
31
       for ((i=0; i<24; i++)) {</pre>
32
            hours[i]=0
33
34
```

^{7.} voir cette discussion sur StackOverflow

```
doit() {
       local ftmp=/tmp/tmp.file
       end=1
38
       cmd="cat $0 | ${byhour} | sort | uniq -c | sort -nr | head -n ${
         size} ${after}"
       case ${is_byhour} in
         0)
               printf "%-7.7s %-6.6s %-2.2s\n" "occurences" "date" "hour
               printf "%-17.17s\n" "-----
43
       1)
45
               printf "%-7.7s %-2.2s\n" "occurences" "hour"
46
               printf "%-10.10s\n" "------
47
48
               ;;
       esac
49
       eval "${cmd}" > ${ftmp}
50
       while IFS=' ' read count index; do
51
           local h=${index#0}
52
          hours[${h}]=${count}
53
       done < $ftmp</pre>
54
55
       for ((i=0; i<24; i++)) {
        printf "%7d %02d\n" ${hours[i]} $i
     }
57
  [ $# -eq 0 ] && dohelp
  end=0
  while [ $end -eq 0 ]
      case $1 in
           -h | --help)
66
               dohelp
67
68
              ;;
          -s | --size)
70
               [ $# -eq 0 ] && onerror 2 "$1 needs a parameter"
71
               set_size "$1"
72
               shift
73
```

Conrairement aux précédents, si une tranche horaire n'est pas représentée dans les fichiers logs passés en paramètres, elle sera tout de même affichée avec la valeur 0.