DansGuardian

Ce tutoriel développe la mise en place d'un serveur proxy filtrant les URLs/domaines, ainsi que le contenu des pages (les mots composants les phrases, les extensions de fichiers, ...).

DansGuardian est **plus rapide** que squidGuard (voir le site).
Il est **utilisé derrière** un serveur proxy tel que Squid par exemple.

Installation des packages

Mandrake	debian
DansGuardian-2.7.7.8-2mdk perl-Mail-Sender-0.8.10-2mdk	dansguardian_2.8.0.4-2_i386.deb

Configuration de DansGuardian

Editez le fichier **/etc/dansguardian/dansguardian.conf**

```
# DansGuardian config file for version 2.8.0
# **NOTE** as of version 2.7.5 most of the list files are now in
dansguardianf1.conf
```

debian

```
# Comment this line out once you have modified this file to suit your needs
#UNCONFIGURED
```

```
# Web Access Denied Reporting (does not affect logging)
#
# -1 = log, but do not block - Stealth mode
# 0 = just say 'Access Denied'
# 1 = report why but not what denied phrase
# 2 = report fully
# 3 = use HTML template file (accessdeniedaddress ignored)
#
reportinglevel = 3

# Language dir where languages are stored for internationalisation
# Le répertoire "languages" contient pour chaque langues un modèle HTML
# qui sera affiché quand une page est déclarée interdite.
# Il est facile d'adapter ce modèle template.html (voici le mien)
# Vous devez mettre l'option reportinglevel = 3
#
languagedir = '/etc/dansguardian/languages'
```

```
# language to use from languagedir.
language = 'french'

# Logging Settings
#
# 0 = none 1 = just denied 2 = all text based 3 = all requests
loglevel = 2

# Log Exception Hits
# Log if an exception (user, ip, URL, phrase) is matched and so
# the page gets let through. Can be useful for diagnosing
# why a site gets through the filter. on | off
logexceptionhits = on

# Log File Format
# 1 = DansGuardian format 2 = CSV-style format
# 3 = Squid Log File Format 4 = Tab delimited
logfileformat = 1

# Log file location
#
# Defines the log directory and filename.
loglocation = '/var/log/dansguardian/access.log'

# Network Settings
#
# the IP that DansGuardian listens on.
# If left blank DansGuardian will listen on all Ips.
# That would include all NICs, loopback, modem, etc.
# Normally you would have your firewall protecting this,
# but if you want you can limit it to only 1 IP. Yes only one.
filterip =

# Indiquez le port qui sera écouté par le daemon DansGuardian.
# C'est ce numéro de port que vous allez configurer sur le
# navigateur des clients. il doit être supérieur à 1024.
filterport = 8080

# Indiquez l'IP de votre serveur proxy (Squid par exemple)
# Si DansGuardian se trouve sur la même machine que Squid
# mettez quand même votre IP à la place de la boucle locale.
proxyip = 192.168.1.4

# Indiquez le port d'écoute de votre serveur proxy(3128 sur Squid)
proxyport = 3128

# accessdeniedaddress is the address of your web server to which
# the cgi dansguardian reporting script was copied
# Do NOT change from the default if you are not using the cgi.
#
accessdeniedaddress = 'http://srv4.dmz.alex.fr/cgi-bin/dansguardian.pl'
```

```
# Non standard delimiter (only used with accessdeniedaddress)
# Default is enabled but to go back to the original standard mode dissable
it.
nonstandarddelimiter = on

# Banned image replacement
# Images that are banned due to domain/url/etc reasons including
# those in the adverts blacklists can be replaced by an image.
# This will, for example, hide images from advert sites and
# remove broken image icons from banned domains.
# 0 = off
# 1 = on (default)
usecustombannedimage = 1
custombannedimagefile = '/etc/dansguardian/transparent1x1.gif'

# Filter groups options
# filtergroups sets the number of filter groups. A filter group is a set of
# content filtering options you can apply to a group of users.  The value
# must be 1 or more. DansGuardian will automatically look for
# dansguardianfN.conf where N is the filter group.  To assign users to groups
# use the filtergroupslist option.  All users default to filter group 1.
# You must have some sort of authentication to be able to map users
# to a group.  The more filter groups the more copies of the lists will be in
# RAM so use as few as possible.
filtergroups = 1
filtergroupslist = '/etc/dansguardian/filtergroupslist'

# Authentication files location
bannediplist = '/etc/dansguardian/bannediplist'
exceptioniplist = '/etc/dansguardian/exceptioniplist'
banneduserlist = '/etc/dansguardian/banneduserlist'
exceptionuserlist = '/etc/dansguardian/exceptionuserlist'

# Show weighted phrases found
# If enabled then the phrases found that made up the total which excedes
# the naughtyness limit will be logged and, if the reporting level is
# high enough, reported. on | off
showweightedfound = on

# Weighted phrase mode
# There are 3 possible modes of operation:
# 0 = off = do not use the weighted phrase feature.
# 1 = on, normal = normal weighted phrase operation.
# 2 = on, singular = each weighted phrase found only counts once on a page.
#
weightedphrasemode = 2

# Positive result caching for text URLs
# Caches good pages so they don't need to be scanned again
# 0 = off (recommended for ISPs with users with disimilar browsing)
# 1000 = recommended for most users
# 5000 = suggested max upper limit
urlcachenumber = 1000
```

```
#
# Age before they are stale and should be ignored in seconds
# 0 = never
# 900 = recommended = 15 mins
urlcacheage = 900

# Smart and Raw phrase content filtering options
# Smart is where the multiple spaces and HTML are removed before
# phrase filtering. Raw is where the raw HTML including meta tags
# are phrase filtered. CPU usage can be effectively halved by
# using setting 0 or 1
# 0 = raw only
# 1 = smart only
# 2 = both (default)
phrasefiltermode = 2

# Lower casing options
# When a document is scanned the uppercase letters are converted
# to lower case in order to compare them with the phrases.
# However this can break Big5 and other 16-bit texts.
# If needed preserve the case. As of version 2.7.0 accented
# characters are supported.
# 0 = force lower case (default)
# 1 = do not change case
preservecase = 0

# Hex decoding options
# When a document is scanned it can optionally convert %XX to chars.
# If you find documents are getting past the phrase filtering
# due to encoding then enable. However this can break Big5
# and other 16-bit texts.
# 0 = disabled (default)
# 1 = enabled
hexdecodecontent = 0

# Force Quick Search rather than DFA search algorithm
# The current DFA implementation is not totally 16-bit character
# compatible but is used by default as it handles large phrase
# lists much faster. If you wish to use a large number of 16-bit
# character phrases then enable this option.
# 0 = off (default)
# 1 = on (Big5 compatible)
forcequicksearch = 0

# Reverse lookups for banned site and URLs.
# If set to on, DansGuardian will look up the forward DNS for an
# IP URL address and search for both in the banned site and URL
# lists. This would prevent a user from simply entering the IP
# for a banned address. It will reduce searching speed somewhat
# so unless you have a local caching DNS server, leave it off
# and use the Blanket IP Block option in the bannedsitelist file instead.
reverseaddresslookups = off

# Reverse lookups for banned and exception IP lists.
```

```
# If set to on, DansGuardian will look up the forward DNS for the IP
# of the connecting computer. This means you can put in hostnames
# in the exceptioniplist and bannediplist. It will reduce
# searching speed somewhat so unless you have a local DNS server,
# leave it off.
reverseclientiplookups = off

# Build bannedsitelist and bannedurllist cache files.
# This will compare the date stamp of the list file with the date
# stamp of the cache file and will recreate as needed.
# If a bsl or bul .processed file exists, then that will be used
# instead. It will increase process start speed by 300%.
# On slow computers this will be significant.
# Fast computers do not need this option. on | off
createlistcachefiles = on

# POST protection (web upload and forms)
# does not block forms without any file upload, i.e. this is
# just for blocking or limiting uploads measured in kibibytes
# after MIME encoding and header bumph
# use 0 for a complete block
# use higher (e.g. 512 = 512Kbytes) for limiting
# use -1 for no blocking
#maxuploadsize = 512
#maxuploadsize = 0
maxuploadsize = -1

# Max content filter page size
# Sometimes web servers label binary files as text which can
# be very large which causes a huge drain on memory and cpu
# resources. To counter this, you can limit the size of the
# document to be filtered and get it to just pass it straight
# through. This setting also applies to content regular
# expression modification.
# The size is in Kibibytes - eg 2048 = 2Mb
# use 0 for no limit
maxcontentfiltersize = 256

# Username identification methods (used in logging)
# You can have as many methods as you want and not just one.
# The first one will be used then if no username is found,
# the next will be used.
# * proxyauth is for when basic proxy authentication is used
# (no good for transparent proxying).
# * ntlm is for when the proxy supports the MS NTLM authentication
# protocol. (Only works with IE5.5 sp1 and later).*NOT IMPLEMENTED
# * ident is for when the others don't work. It will contact the
# computer that the connection came from and try to connect to
# an identd server and query it for the user owner of the connection.
usernameidmethodproxyauth = on
usernameidmethodntlm = off # **NOT IMPLEMENTED**
usernameidmethodident = off

# Premptive banning - this means that if you have proxy auth
```

```
# enabled and a user accesses a site banned by URL for example
# they will be denied straight away without a request
# for their user and pass. This has the effect of requiring
# the user to visit a clean site first before it knows who they
# are and thus maybe an admin user. This is how DansGuardian
# has always worked but in some situations it is less than
# ideal. So you can optionally disable it. Default is on.
# As a side effect this makes AD image replacement work better
# as the mime type is know.
preemptivebanning = on

# Misc settings
#
# if on it adds an X-Forwarded-For: <clientip> to the HTTP request
# header. This may help solve some problem sites that need to
# know the source ip. on | off
forwardedfor = off

# if on it uses the X-Forwarded-For: <clientip> to determine the
# client IP. This is for when you have squid between the clients
# and DansGuardian. Warning - headers are easily spoofed. on | off
usexforwardedfor = off

# if on it logs some debug info regarding forking and accepting
# which can usually be ignored. These are logged by syslog.
# It is safe to leave it on or off
logconnectionhandlingerrors = on

# Fork pool options
#
# sets the maximum number of processes to sporn to handle the
# incomming connections. Max value usually 250 depending on OS.
# On large sites you might want to try 180.
maxchildren = 120

# sets the minimum number of processes to sporn to handle the
# incomming connections. On large sites you might want to try 32.
minchildren = 8

# sets the minimum number of processes to be kept ready to handle
# connections. On large sites you might want to try 8.
minsparechildren = 4

# sets the minimum number of processes to sporn when it runs out
# On large sites you might want to try 10.
preforkchildren = 6

# sets the maximum number of processes to have doing nothing.
# When this many are spare it will cull some of them.
# On large sites you might want to try 64.
maxsparechildren = 32

# sets the maximum age of a child process before it croaks it.
# This is the number of connections they handle before exiting.
```

```
# On large sites you might want to try 10000.
maxagechildren = 500

# Process options
# (Change these only if you really know what you are doing).
# These options allow you to run multiple instances of
# DansGuardian on a single machine. Remember to edit the
# log file path above also if that is your intention.

# IPC filename
#
# Defines IPC server directory and filename used to communicate with the log
process.
ipcfilename = '/tmp/.dguardianipc'

# URL list IPC filename
#
# Defines URL list IPC server directory and filename used to communicate with
# the URL cache process.
urlipcfilename = '/tmp/.dguardianurlipc'

# PID filename
#
# Defines process id directory and filename.
pidfilename = '/var/run/dansguardian.pid'

# Disable daemoning
# If enabled the process will not fork into the background.
# It is not usually advantageous to do this.
# on|off ( defaults to off )
nodaemon = off

# Disable logging process
# on|off ( defaults to off )
nologger = off

# Daemon runas user and group
# This is the user that DansGuardian runs as.
# Normally the user/group nobody.
# Uncomment to use. Defaults to the user set at compile time.
# daemonuser = 'nobody'
# daemongroup = 'nogroup'

# Soft restart
# When on this disables the forced killing off all processes
# in the process group. on|off ( defaults to off )
softrestart = off
```

Editez le fichier **/etc/dansguardian/dansguardianf1.conf**

```
# DansGuardian filter group config file for version 2.8.0

# Content filtering files location
bannedphraselist = '/etc/dansguardian/bannedphraselist'
weightedphraselist = '/etc/dansguardian/weightedphraselist'
exceptionphraselist = '/etc/dansguardian/exceptionphraselist'
bannedsitelist = '/etc/dansguardian/bannedsitelist'
greysitelist = '/etc/dansguardian/greysitelist'
exceptionsitelist = '/etc/dansguardian/exceptionsitelist'
bannedurllist = '/etc/dansguardian/bannedurllist'
greyurllist = '/etc/dansguardian/greyurllist'
exceptionurllist = '/etc/dansguardian/exceptionurllist'
bannedregexpurllist = '/etc/dansguardian/bannedregexpurllist'
bannedextensionlist = '/etc/dansguardian/bannedextensionlist'
bannedmimetypelist = '/etc/dansguardian/bannedmimetypelist'
picsfile = '/etc/dansguardian/pics'
contentregexplist = '/etc/dansguardian/contentregexplist'

# Naughtyness limit
# Nombre de points maximum alloués aux utilisateurs par défaut.
# Selon les mots rencontrés dans une page web, un nombre de points
# est attribué. L'addition de ces points donne le niveau d'accès :
# 50 pour les jeunes enfants, 100 pour des enfants,
# 160 pour des adolescents.
naughtynesslimit = 160

# Temporary Denied Page Bypass
# It provides a link on the denied page to bypass the ban for a few minutes.
# To be secure it uses a random hashed secret generated at daemon startup.
# You define the number of seconds the bypass will function for before the
# deny will appear again.
# 300 = 5 minutes
# 0 = disable ( defaults to 0 )
bypass = 0

# Temporary Denied Page Bypass Secret Key
# Rather than generating a random key you can specify one. It must be more
than 8 chars.
# '' = generate a random one (recommended and default)
# 'Mary had a little lamb.' = an example
# '76b42abc1cd0fdcaf6e943dcbc93b826' = an example
bypasskey = ''
```

En plus de l'annalyse des mots dans les pages Web, vous pouvez ajouter une liste des Sites blacklistés.
Téléchargez gratuitement sur le Site http://urlblacklist.com/ la derrière version de l'archive et décompressez la dans le répertoire **/etc/dansguardian/**
Vous pouvez utiliser le script de mise à jour de l'archive (UpdateBL) donné sur le même Site et le lancer avec cron.

Editez les fichiers **/etc/dansguardian/bannedsitelist** et **/etc/dansguardian/bannedurllist** pour décommenter les lignes "`.Include</etc/dansguardian/blacklists/...`"

Il y a des sites normaux qui sont blacklistés par erreur comme http://www.free.fr/ car dans la liste, des sites hébergés chez Free sont non recommendable. Il faudra ajouter dans le fichier **/etc/dansguardian/exceptionsitelist** : free.fr par exemple.

Vous pouvez utiliser DansGuardian pour filtrer le contenu des pages uniquement ou lui confier toutes vos règles d'accès vers les sites, extensions, ...

DansGuardian permet de conserver une authentification depuis Squid (OpenLADP par exemple) ou simplement derrière un proxy transparent.

Document mis à jour : 13/06/05