Analyse de l'écosystème des antivirus et malwares

par Stéphane Aubry @ AnimaMachina.com - août 2015

L'année dernière j'ai passé une longue période sur un travail d'analyse de l'écosystème des antivirus et de celui des malwares (logiciels malveillants, « virus » informatiques). Cet article est le compte-rendu de cette analyse.

Avertissement préalable et motivations

Le site VirusTotal qui est maintenant une sous société de Google, est la source principale des données que j'utilise pour mon analyse. Il déconseille justement fortement d'utiliser leur site pour un travail d'analyse comparative des antivirus (cf. : « BAD IDEA: VirusTotal for antivirus/URL scannertesting »). Même si les raisons invoquées, principalement techniques, sont réelles, et donc à prendre en considération, chacun comprendra qu'une analyse comparative des antivirus implique bien évidemment aussi de nombreux enjeux commerciaux, surtout pour une régie publicitaire comme Google qui peut avoir certains des fabricants d'antivirus comme annonceurs. Ayant pris en considérations ces éléments, chaque lecteur de cet article, saura prendre le recul nécessaire à la vue des résultats et analyses présentés.

En outre, mon désir principal n'était pas une simple comparaison des « performances » des antivirus. Une de mes curiosités m'ayant poussé à ce travail, était surtout de comparer les comportements de ces antivirus, de voir les manières de détecter les malwares, de les nommer, de les classer, de ressortir des similitudes ou différences entre eux à ces niveaux.

Une de mes autres motivations était aussi d'analyser les malwares dans leur ensemble, éventuellement par famille, non pas face à tel ou tel antivirus, mais face à l'écosystème des antivirus dans leur globalité. L'évolution des malwares au cours du temps peut être vue dans sa globalité, comme un véritable écosystème. En défense face à cette menace, dans le même ordre, les antivirus, impliquant arrangements et compétitions entre leurs fabricants, peut aussi être vu comme un écosystème évolutif, qui progresse et se modifie en fonction de la menace qu'il essaye de détecter et de défendre.

Méthodologie

Afin de récupérer un échantillon de fichiers malwares suffisamment grand pour que l'analyse statistique ait le plus de cohérence possible, j'ai utilisé l'outil Maltrieve. Cet outil se connecte à un ensemble de listes autour de la sécurité et récupère tous les fichiers, postés par les utilisateurs de ces sources, contenant de nouveaux malwares et fichiers suspects. Pour analyser les résultats de détections avec le plus grand nombre possible d'antivirus, j'ai utilisé le site VirusTotal. Pour automatiser l'upload des fichiers à analyser chez VirusTotal, je me suis servi de plusieurs outils, dont principalement Uirusu.

Pour analyser de manière plus pointue une partie de la structure des malwares exécutables (malwares trouvés classiquement sous Windows), j'ai aussi utilisé l'outil Pyew. Il permet un nombre de fonctions intéressantes pour analyser le code de l'exécutable et sa structure.

La première étape de l'utilisation de l'outil de récupération des malwares m'a parmi de récupérer **930 fichiers suspects**. Parmi ces fichiers, se trouvaient des vrais malwares, des suspicions qui n'en étaient peut-être pas, des pages web incluant des virus embarqués (malwares en code JavaScript ou autre). Mais ces fichiers comprenaient aussi des pages web d'erreurs indiquant des liens non-disponibles, donc des fichiers non malveillants. J'ai uploadé tous ces fichiers chez VirusTotal, et récupéré les données, une première fois incluant les détections des **49 antivirus** paramétrés avec les

définitions de virus des **13-14 janvier 2014**. J'ai effectué un second lot d'upload 10 mois plus tard, afin d'analyser les modifications des détections. Ce second lot de détections a été effectué par les **54 antivirus** de VirusTotal, certains ayant été supprimés et d'autres ajoutés parmi la liste de janvier. Ce second lot de détections a impliqué les définitions de virus des **31 octobre et 1er novembre 2014**.

Parmi tous ces fichiers envoyés pour analyse, il n'était pas possible de ressortir les vrais malwares de manière certaine, ni après analyse par VirusTotal. En effet des malwares « zero days », pouvaient potentiellement être inclus, c'est-à-dire des malwares n'étant détecté par aucun antivirus. Dans ces fichiers pouvaient aussi se trouver des « faux positifs », c'est à dire des fichiers totalement sains, faussement détectés comme malwares par un ou plusieurs antivirus. Ces fichiers « faussant » les données, celles-ci permettaient quand même de créer des statistiques, et d'analyser les résultats de manière intéressante. Ces biais sont donc à prendre en considération dans la lecture des taux de détections.

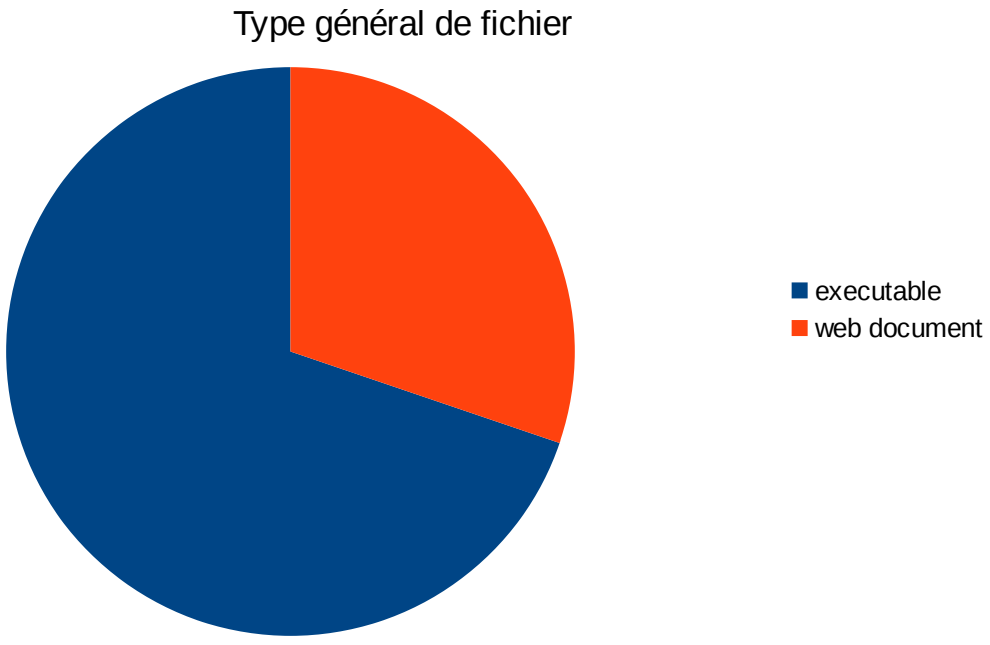
Types de fichiers

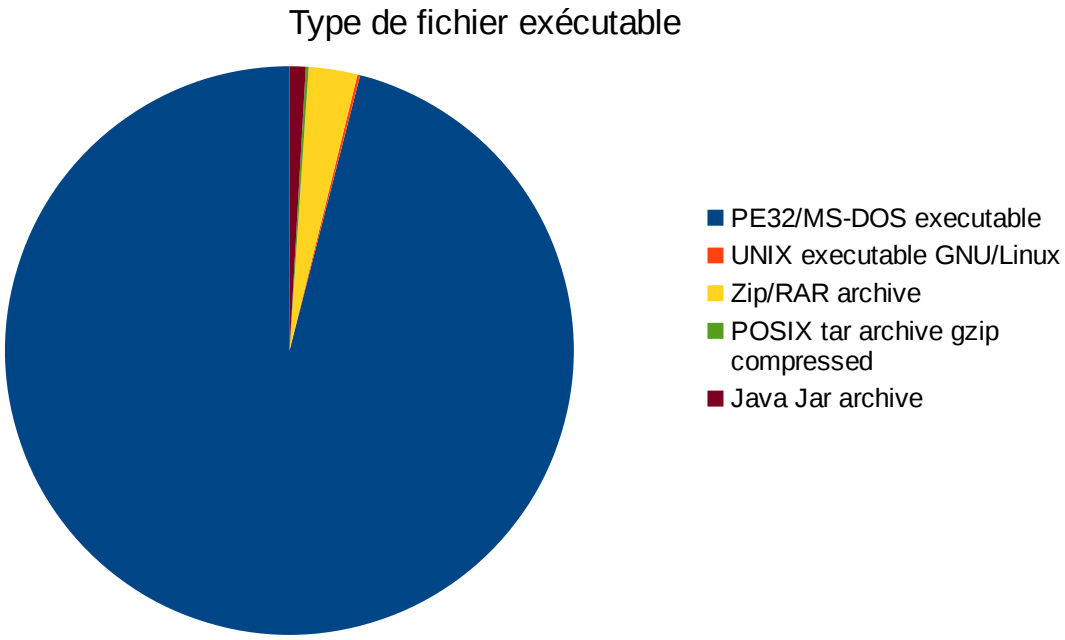
La première étape que j'ai menée était d'identifier les types de fichiers récupérés depuis les listes de malwares. J'ai regroupé les différents types trouvés en 2 grandes catégories :

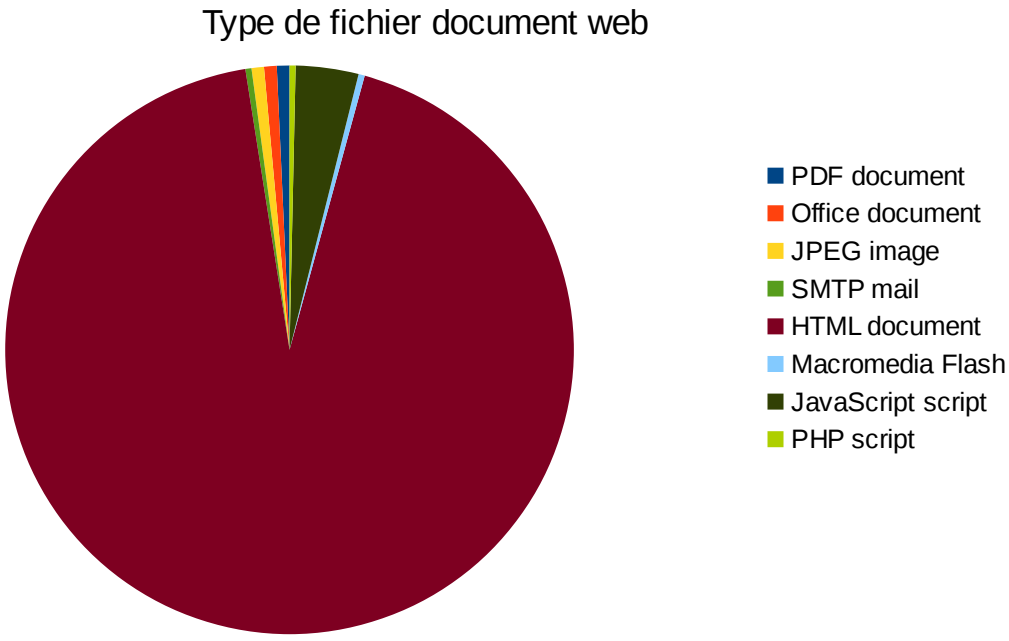
– les **exécutables** :
Principalement exécutables Windows, mais aussi Linux, archives de programmes, programme java

– les **documents web** :
Y sont classés principalement des fichiers de pages web en HTML, mais aussi des scripts utilisés dans les pages web comme du JavaScript, Flash, fichier de style CSS, image Jpeg, fichier mail.
Les types de fichiers dans cette catégorie ne sont pas tous des fichiers « purement » internet. J'y ai regroupé aussi les documents de bureautique (documents PDF, documents Office). Comme leur nombre est petit, et qu'ils sont plus proches de documents que de programmes ils peuvent être classés dans cette catégorie.

Voici donc sous forme de graphiques, les proportions des divers types de fichiers récupérés parmi l'échantillon des 930 fichiers à analyser :







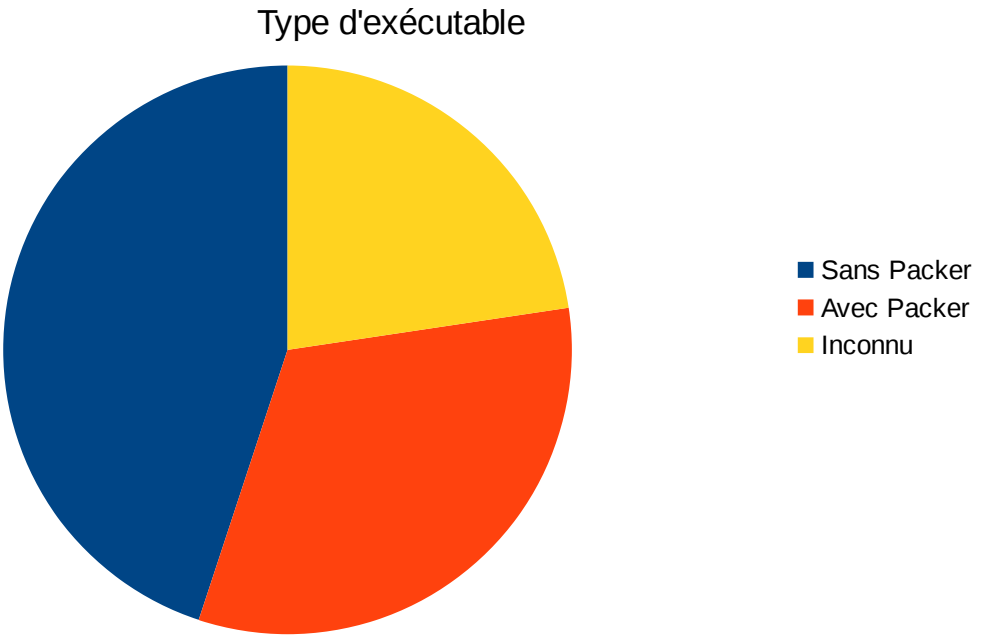
70 % des fichiers sont donc des exécutables, des types de fichiers « classiques » pour les malwares. La majorité de ces exécutables sont des exécutables « standard » Windows. À part les exécutables on y trouve aussi **30 % de documents web**, pages web infectées, documents bureautiques infectés, mais aussi des pages d'erreurs retournées par l'outil de récupération Maltrieve. Dans les types de fichiers on y trouve même des images, où certaines vulnérabilités des routines d'affichage des images ont permis de créer des malwares sous forme d'images.

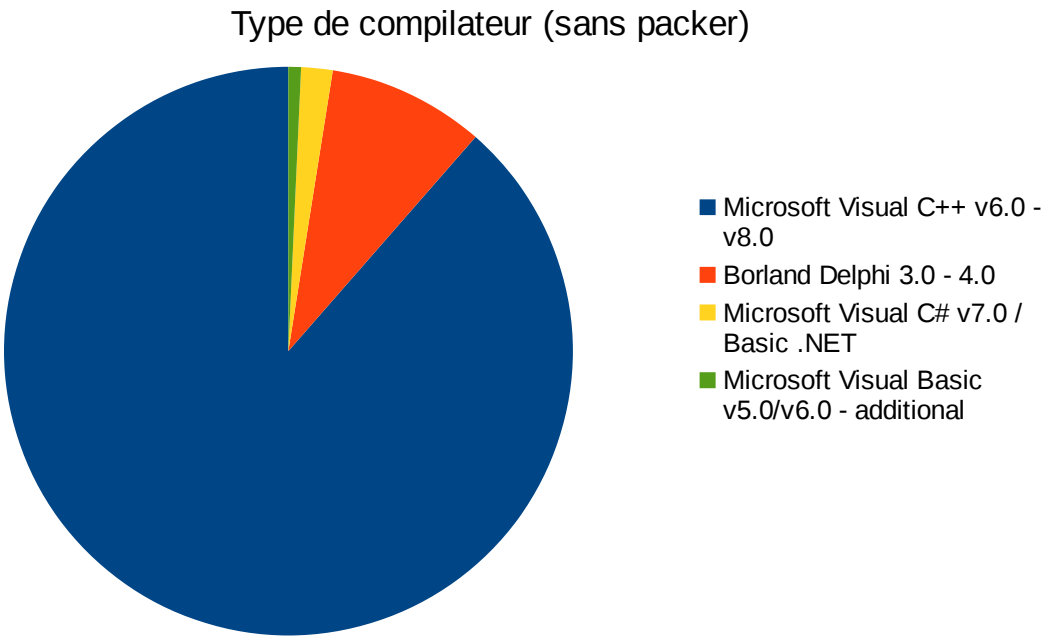
Exécutables : packers et compilateurs

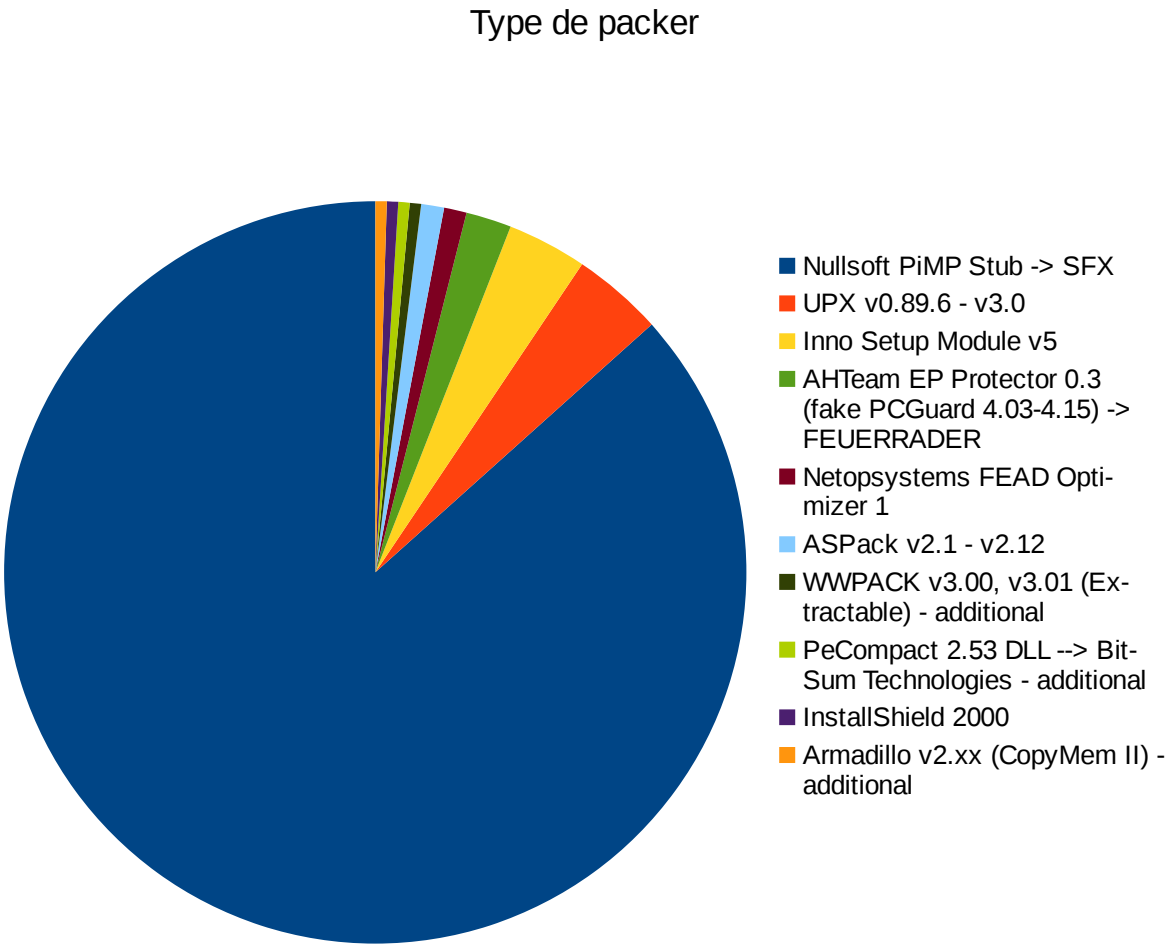
La majorité des fichiers étant des exécutables, je voulais analyser plus en profondeur la structure de ces fichiers. Pour cela j'ai utilité l'outil Pyew qui permet de faire de l'analyse de code assembleur des malwares Windows. J'ai notamment utilisé une fonction intéressante de l'outil qui permet

d'identifier le type de packer utilisé sur l'exécutable. Les packers sont utilisés par les créateurs de malwares (entre-autre), pour compresser l'exécutable, mais aussi pour ajouter des fonctions d'obfuscation, de polymorphie, de cryptographie et autres méthodes permettant de rendre toute analyse de l'exécutable et la création de définitions par les antivirus plus complexe. Pour les exécutables n'incluant pas de packer, l'outil Pyew permet aussi d'identifier le type de compilateur utilisé par le créateur du malware pour créer l'exécutable depuis le code source de programmation.

Voici sous forme graphique, la proportion des exécutables avec packers, les types de compilateurs identités et les types de packers :





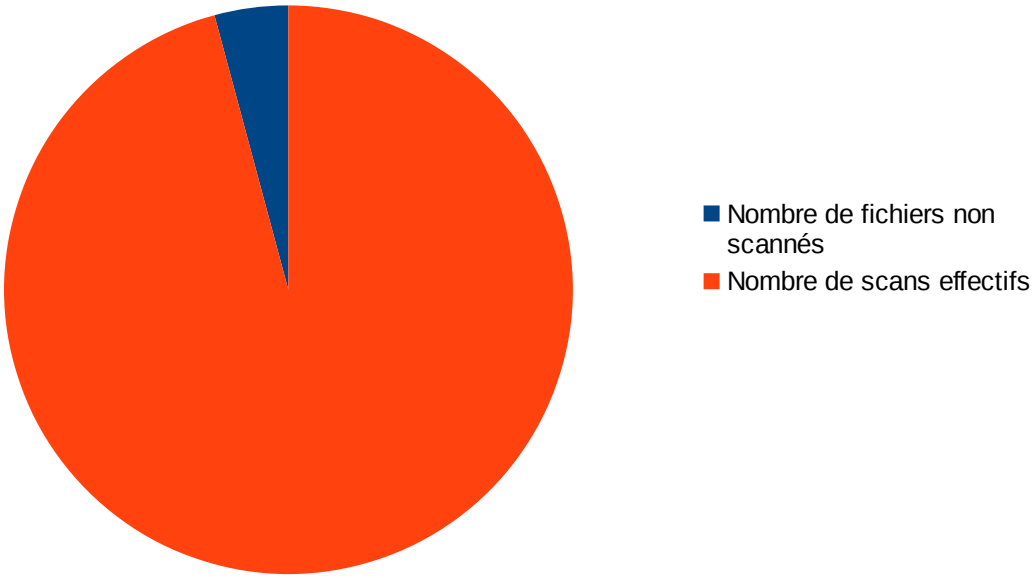


Parmi les types identifiés par Pyew, environ la moitié utilisent des packers. Le packer le plus utilisé est NSIS (Nullsoft Scriptable Install System), un outil open source permettant de créer des exécutables autoinstallables pour Windows. SFX signifie archive auto-extractible. Pour les exécutables ne possédant pas de packer, le compilateur utilisé de manière la plus importante est le compilateur C++ de Microsoft. On retrouve aussi dans une proportion moindre, le compilateur C# de Microsoft ayant créé des malwares « .NET » (dotNet).

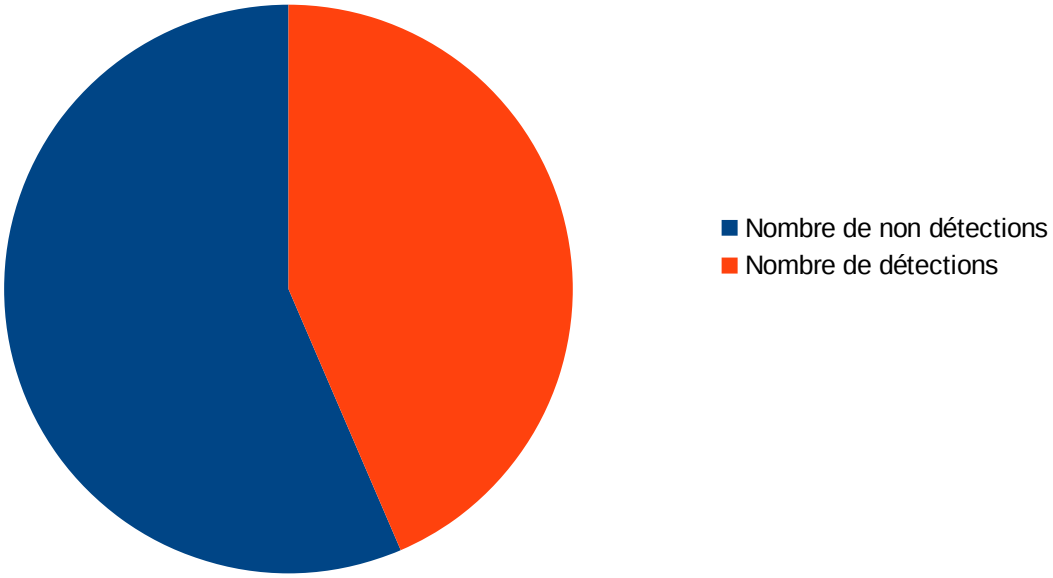
Détections des antivirus

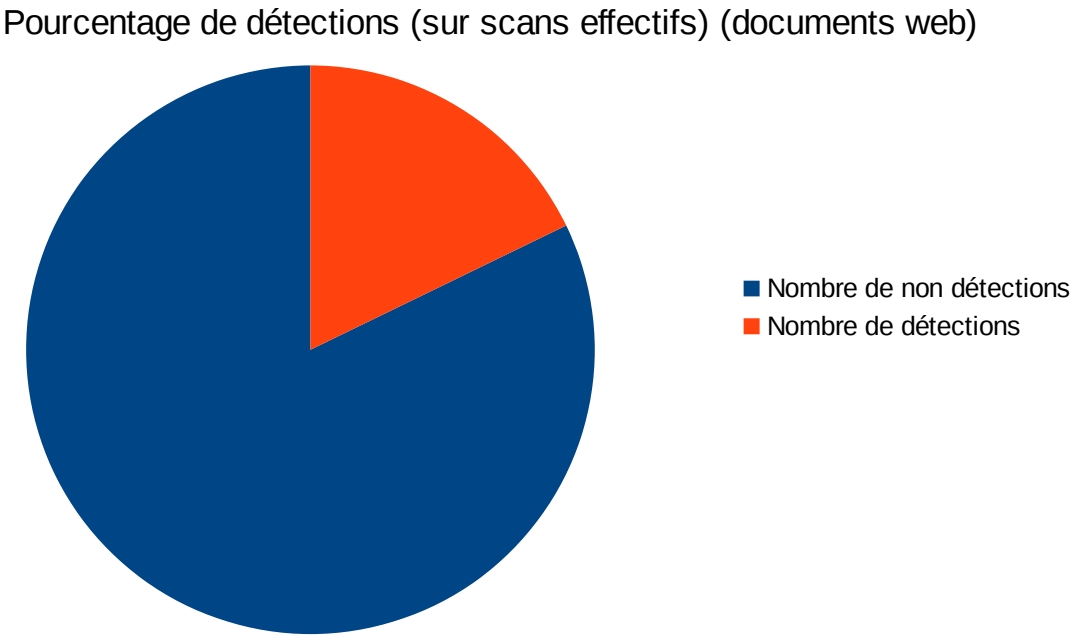
Après récupération des données de détections effectuées par VirusTotal, le premier travail d'analyse que j'ai effectué sur les données de janvier 2014, a été de regarder la proportion de scans effectifs, par rapport aux fichiers non scannés (problème d'un antivirus ne répondant pas après un certain délai). J'ai ensuite ressorti les proportions de détections (malwares identifiés) / non détection (fichier semble-t-il « sain », ou « zero-day ») pour les catégories « exécutables » et « documents web ».

Pourcentage de scans



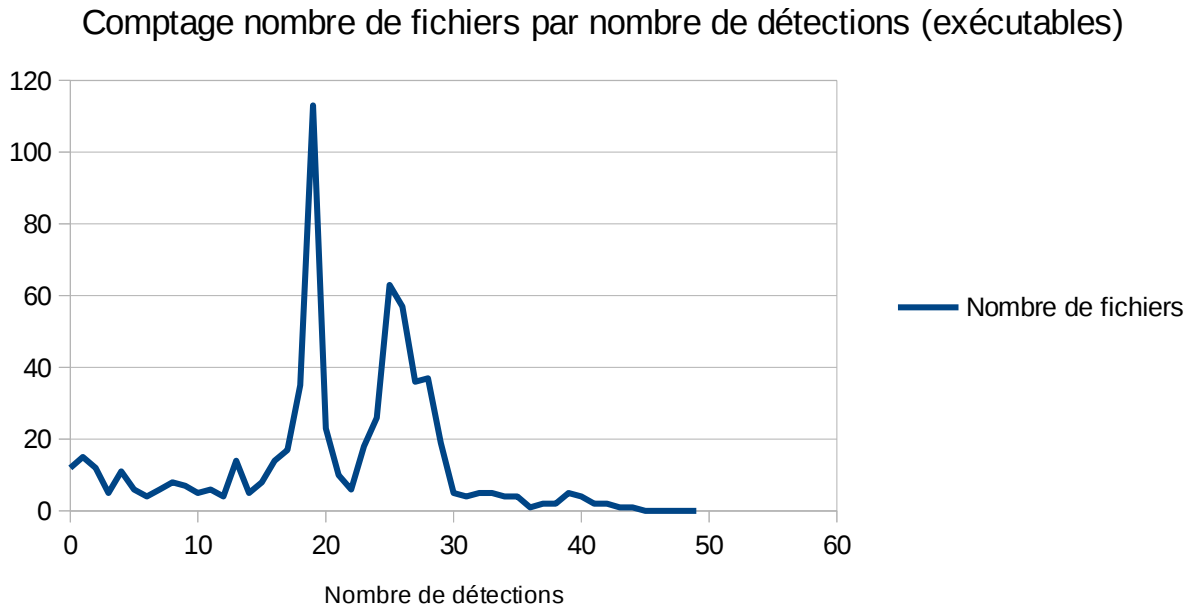
Pourcentage de détections (sur scans effectifs) (exécutables)





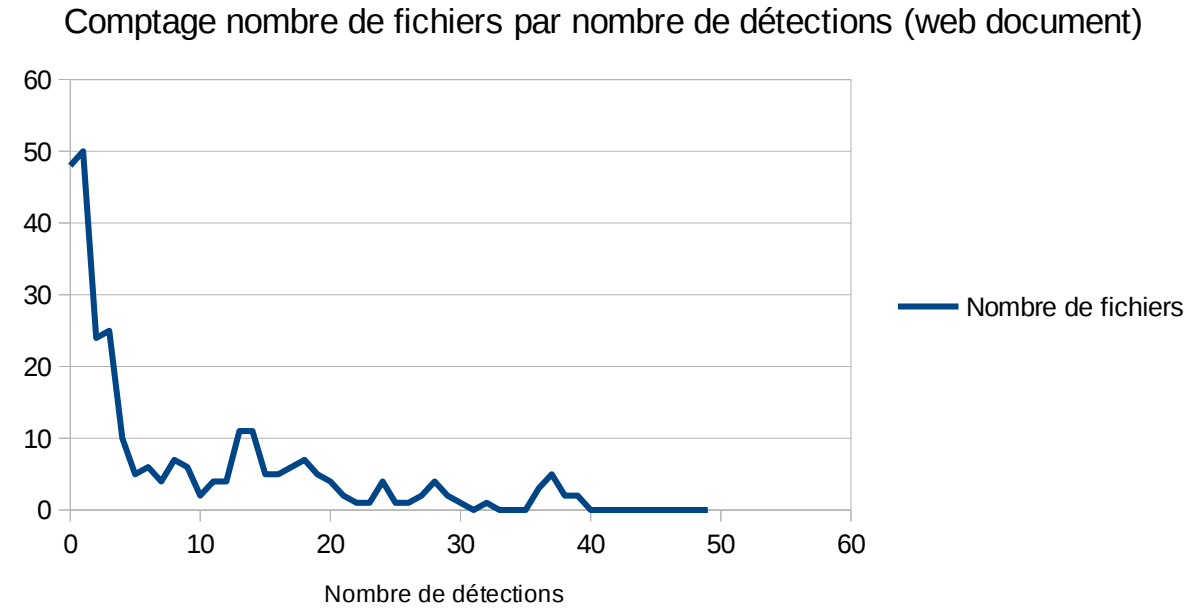
La proportion de scans effectifs est quasiment la même pour les exécutables ou les documents web. Les **détections** effectives de malwares pour les **exécutables sont d'environ 45 %**. À ce stade il n'est pas possible de ressortir les non-détections des antivirus dues à une « incompétence » de l'antivirus, des non-détections qui indiquent réellement un fichier « sain ». Pour les **documents web** les **détections effectives sont d'environ 18 %**. Ce taux plus bas est aussi dû à un nombre de pages web d'erreurs récupérées par l'outil Maltrieve, donc saines. Mais ce nombre de détections web dévoile aussi une plus grande difficulté pour les antivirus d'identifier des pages web malveillantes que d'identifier des exécutables malveillants. Le code JavaScript directement interprété permet une importante variabilité de manière plus souple et plus rapide avec un grand nombre de méthodes possibles pour cacher un fonctionnement, ou créer du code qui en exécute un autre.

J'ai aussi calculé et créé un graphique d'un simple décompte du nombre de fichiers par nombre de détections des 49 antivirus (le nombre de fichiers détectés par 1 antivirus, puis le nombre de fichiers détectés par 2 antivirus, etc.). Ceci permet de ressortir un premier graphique de « distribution » des détections, avec des « pics » des événements qui ressortent. J'ai créé un graphique pour les fichiers exécutables et un pour les documents web.



Concernant les deux « Pics » qui ressortent, voici les noms des malwares dominant dans ces zones :

Position :	**Nom du malware dominant :**
Pic 18-19	InstallRex
Pic 24-28	DomainIQ



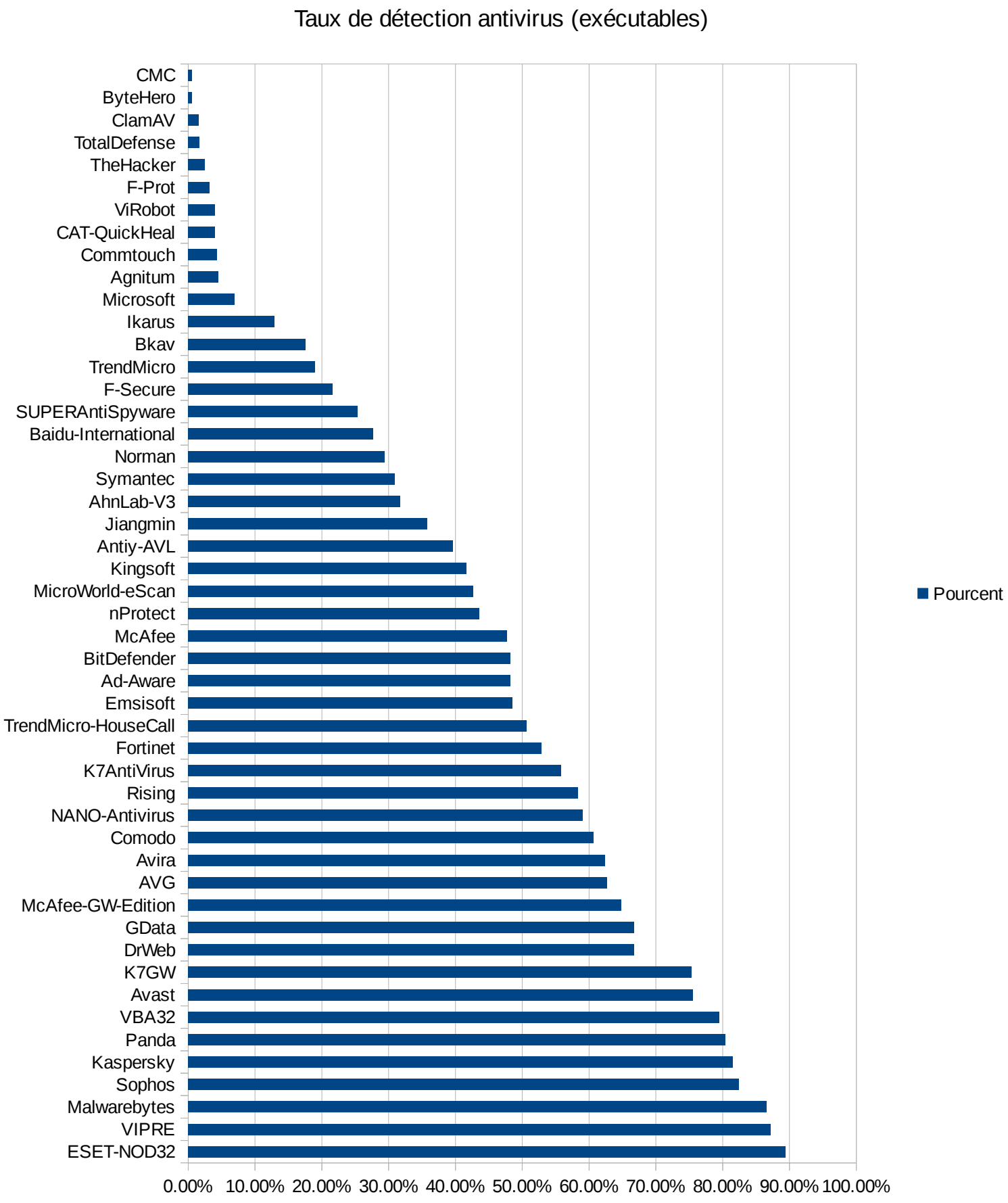
Concernant le « Pic » du début de graphique, voici les noms des malwares dominant dans ces zones :

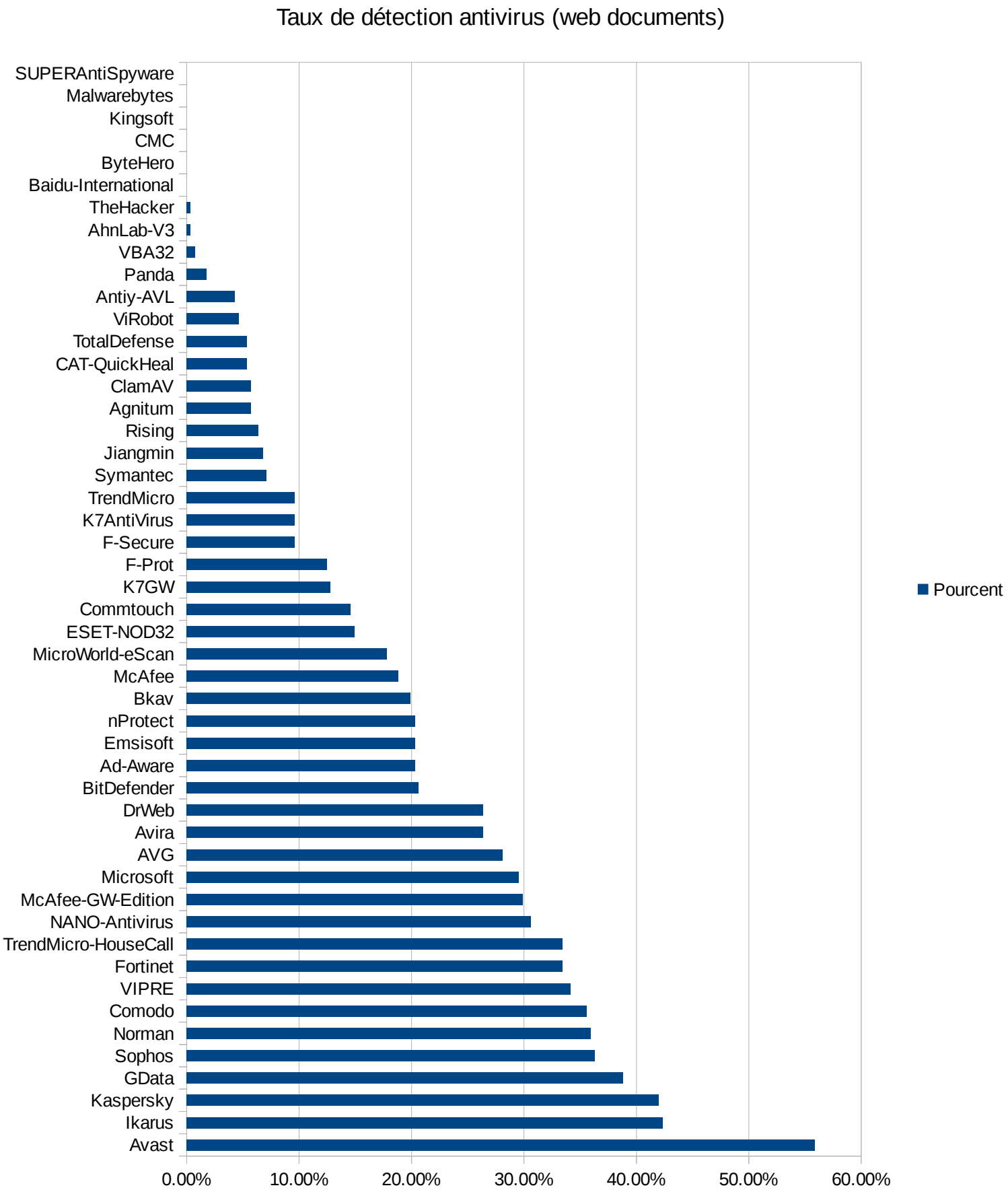
Position :	**Nom du malware dominant :**
Pic 2-3	Includer
Pic 1	HfsIframe

Taux de détections et différences des antivirus

Le travail suivant que j'ai effectué a été de séparer la détection de chaque antivirus par fichier, afin de pouvoir faire des comparaisons entre antivirus.

Les deux premiers graphiques sont un classement par taux de détections des antivirus, une première fois pour les fichiers exécutables et une seconde pour les documents web.

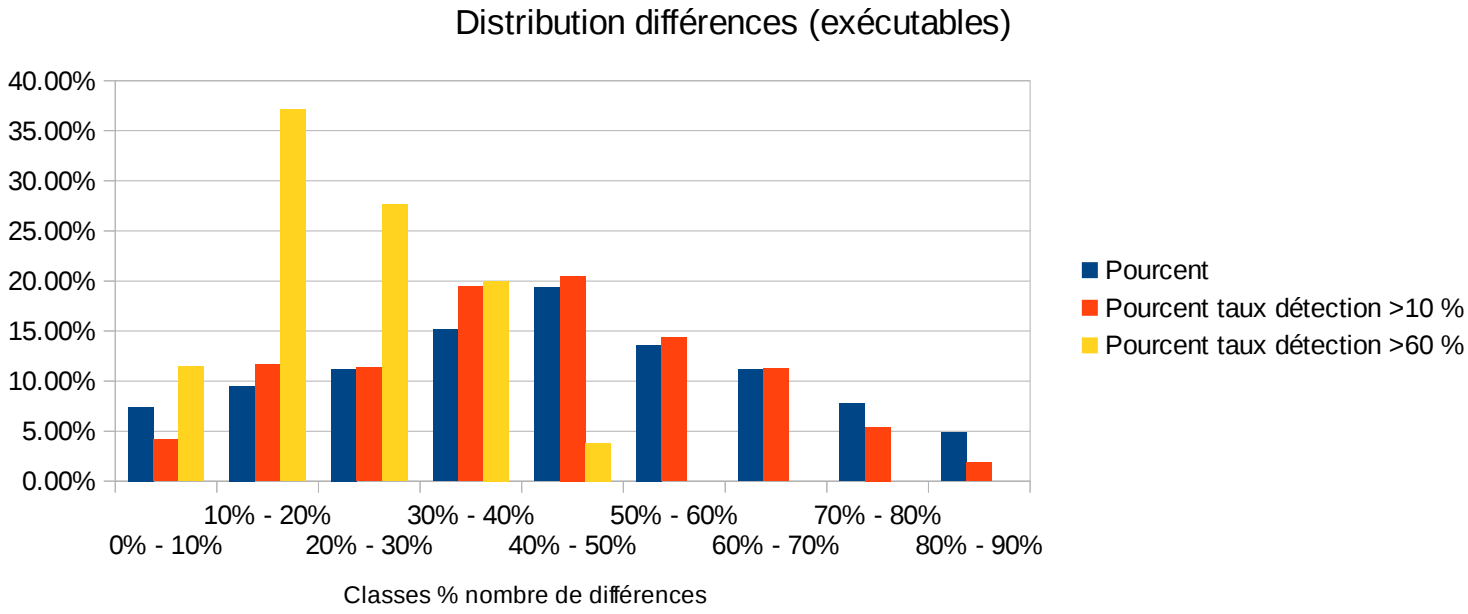

Taux de détection antivirus (exécutables)
CMC
ByteHero
ClamAV
TotalDefense
TheHacker
F-Prot
ViRobot
CAT-QuickHeal
Commtouch
Agnitum
Microsoft
Ikarus
Bkav
TrendMicro
F-Secure
SUPERAntiSpyware
Baidu-International
Norman
Symantec
AhnLab-V3
Jiangmin
Antiy-AVL
Kingsoft
MicroWorld-eScan
nProtect
McAfee
BitDefender
Ad-Aware
Emsisoft
TrendMicro-HouseCall
Fortinet
K7AntiVirus
Rising
NANO-Antivirus
Comodo
Avira
AVG
McAfee-GW-Edition
GData
DrWeb
K7GW
Avast
VBA32
Panda
Kaspersky
Sophos
Malwarebytes
VIPRE
ESET-NOD32
0.00% 10.00% 20.00% 30.00% 40.00% 50.00% 60.00% 70.00% 80.00% 90.00% 100.00%
Pourcent



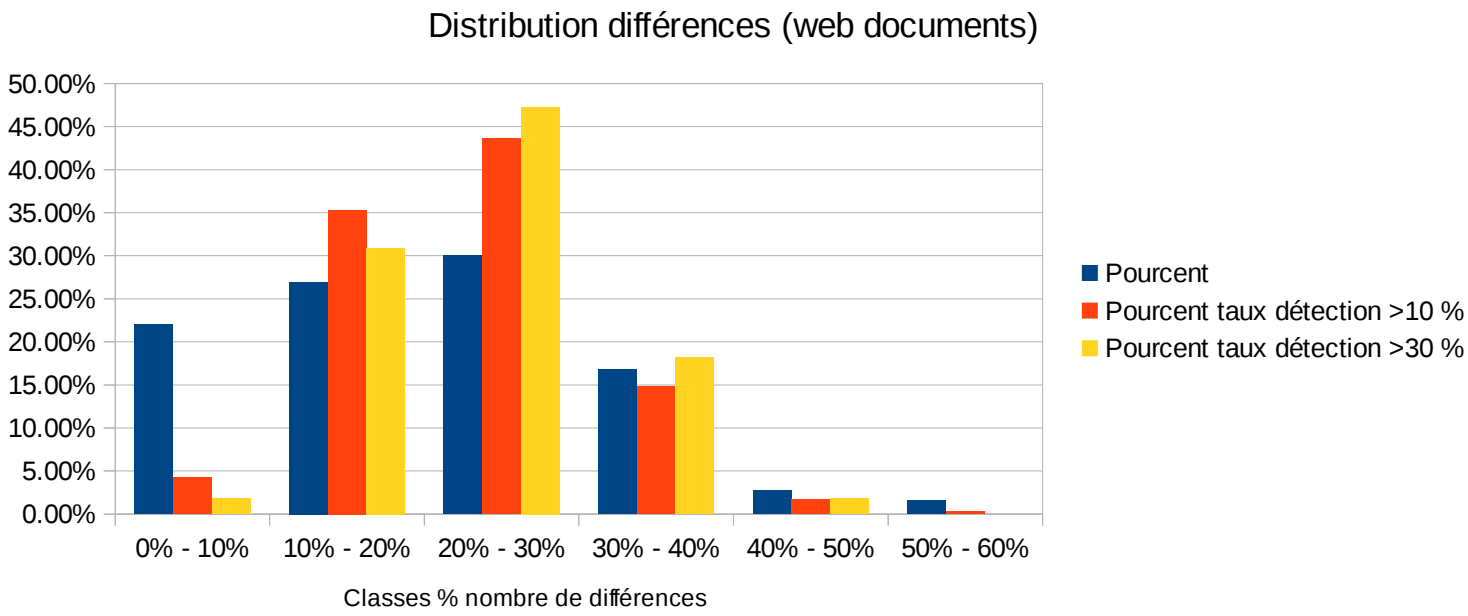
Avertissement : dans les deux graphiques précédents, ce classement peut être faussé par le fait que ces détections incluent les « faux-positifs » qui ne peuvent pas être identifiés de manière certaine et

écartés. De plus dans le graphique concernant les documents web il faut prendre en considération le fait que le nombre de détections des « faux-positifs » est sans doute aussi plus élevé que pour les exécutables.

Le travail suivant que j'ai mené a été un comparatif des différences de détections de chaque antivirus pris deux à deux (une différence est une détection seul de l'un ou l'autre des deux antivirus comparés). Les différences ont été comptées pour chacune de ces combinaisons d'antivirus, soit pour les 49 antivirus, **1176 comparaisons** de détections sur les 930 fichiers. Ensuite ces différences ont été regroupées par classes (les antivirus ayant 0-10 % de nombre de détections différentes, ayant 10 %-20 % de nombre de détections différentes, ainsi de suite). Ceci a permis de créer un graphique de la distribution des différences de détections de ces antivirus. J'ai aussi créé des sous graphiques en ne gardant que les antivirus ayant un taux de détections minimum (par exemple 10 %, pour écarter les antivirus ayant des taux très bas), puis les antivirus ayant les taux de détections les meilleurs (par exemple 60 % pour les exécutables). Voici le premier graphique pour les fichiers exécutables, et le second pour les documents web.



La **courbe bleue** montre la distribution des différences de détections de tous les antivirus. On voit que **la majorité des antivirus ont entre 30 % et 60 % de différences entre eux**, ce qui est important. Si on prend la courbe rouge qui écarte les antivirus les plus médiocres (taux de détection d'au moins 10%), elle n'est pas très différente. Par contre la **courbe jaune** qui ne regroupe que les **meilleurs antivirus** (taux de détection d'au moins 60%), montre que ceux-ci ont **un sommet de différence bien moins important entre eux (10 %-30 % de différences entre eux)**. Les meilleurs antivirus ont aussi un faible taux de différence de détections entre eux. Les antivirus les meilleurs se comportent de manière similaire entre eux (pour les détections des exécutables).



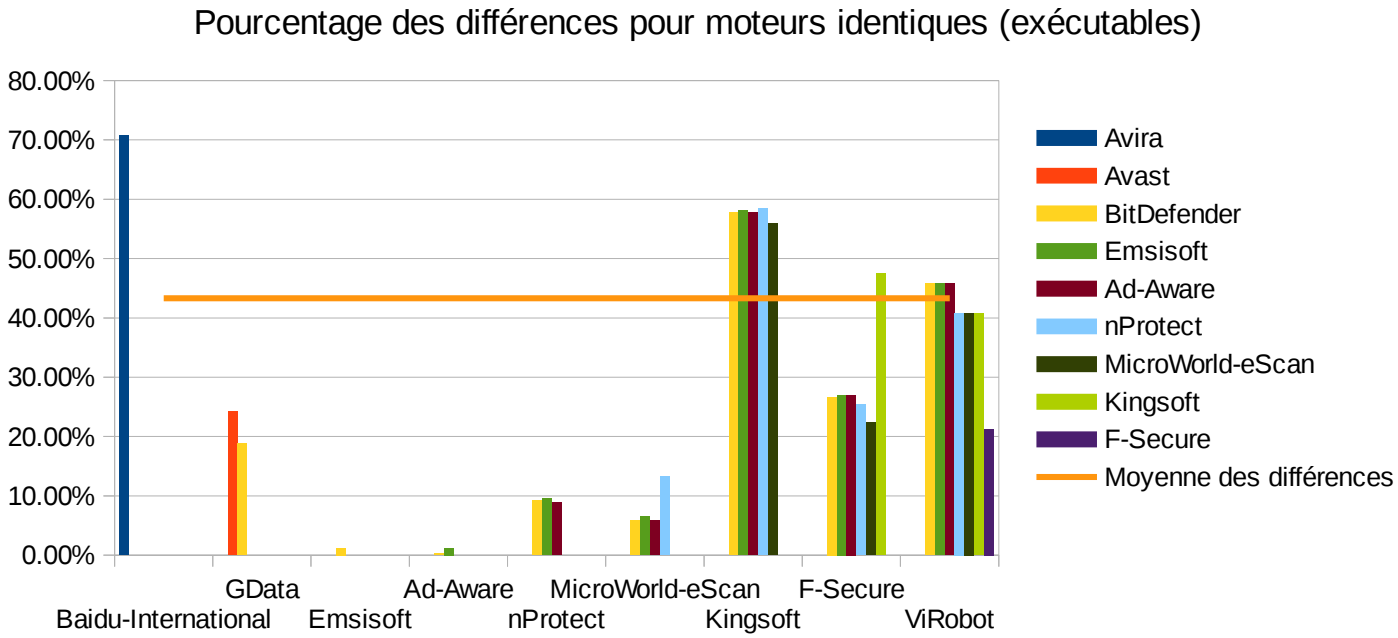
Pour les documents web les graphiques montrent des éléments différents. Les sommets du graphique pour tous les antivirus est le même que ceux qui écartent les plus médiocres ou les meilleurs. On voit donc que **la détection des documents web est plus complexe, et les meilleurs n'ont pas moins de différences entre eux**.

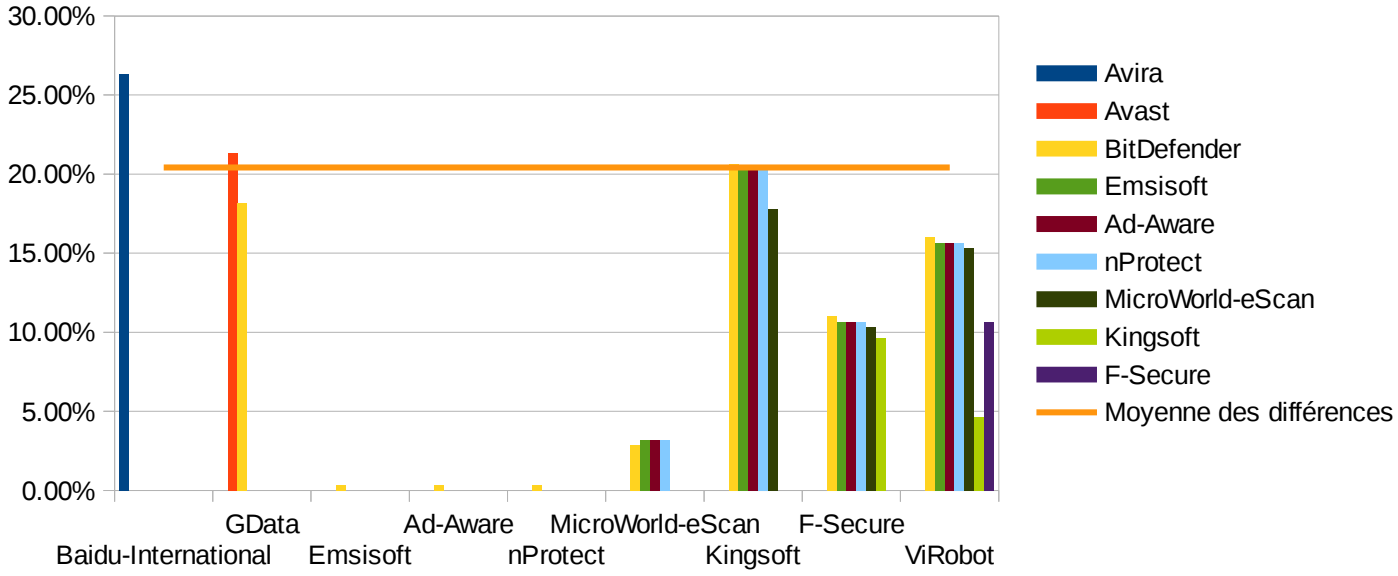
Moteurs d'antivirus communs

J'ai ensuite créé des statistiques comparatives pour les antivirus ayant un moteur de détection d'un fabricant externe. La page des antivirus de AV-Comparatives (cf. : http://www.av-comparatives.org/av-vendors/), et d'autres sources sur les sites des fabricants eux-mêmes, donnent le nom de l'autre fabricant ayant créé leur moteur. Avoir un moteur identique entre deux antivirus ne veut pas dire que les définitions sont les mêmes, mais que le moteur de traitement est identique. Les fabricants externes des moteurs sont résumés dans la liste suivante :

Antivirus :	**Moteur fabriquant externe :**
Baidu-International	Avira
Gdata	BitDefender, Avast (double moteur externe)
Emsisoft	BitDefender
Ad-Aware	BitDefender
nProtect	BitDefender
MicroWorld-eScan	BitDefender
Kingsoft	BitDefender
F-Secure	BitDefender
ViRobot	BitDefender

Ici j'ai retenu dans les données précédentes des différences des détections, que celles qui concernent les fabricants avec un moteur identique. Le premier graphique montre ces différences pour les fichiers exécutables, et le second pour les documents web.





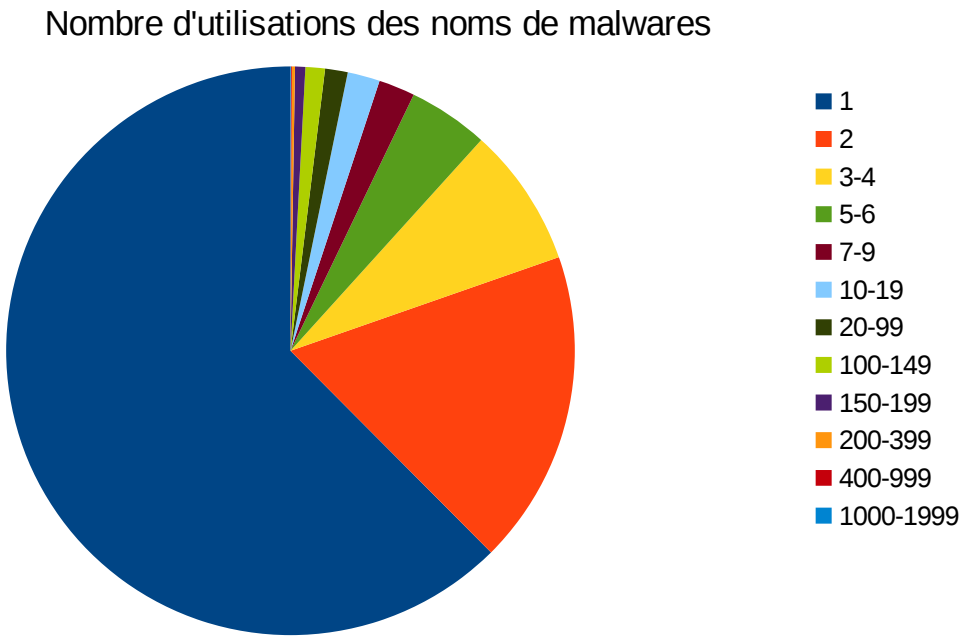
On voit qu'avoir un moteur identique, ne diminue pas forcément le pourcentage de différence de détections entre ces antivirus par rapport à la différence moyenne des autres antivirus. Donc l'argument d'**un moteur identique semble avoir peu d'influence sur la qualité de l'antivirus**. Il ressort surtout la qualité des définitions de chaque fabricant.

Une deuxième partie des analyses a été concentrée sur les malwares eux-mêmes et non plus seulement sur les antivirus.

Noms des malwares

Le premier travail que j'ai mené sur les malwares a été de recenser tous les noms complets des malwares détectés par tous les antivirus. Un malware est identifié selon une nomenclature incluant le nom de famille (le nom usuel), la version (pour les malwares ayant de nombreuses versions), le type de malware, la plateforme ou le langage de programmation impliqué. Pour un exemple de convention de nommage utilisée par Microsoft : https://www.microsoft.com/security/portal/mmpc/shared/malwarenaming.aspx. Malgré le fait qu'il existe une convention de nommage qui se veut commune, CARO, chaque fabricant d'antivirus s'en inspire seulement et adopte souvent sa propre convention de nommage, ce qui rend les comparaisons encore plus compliquées.

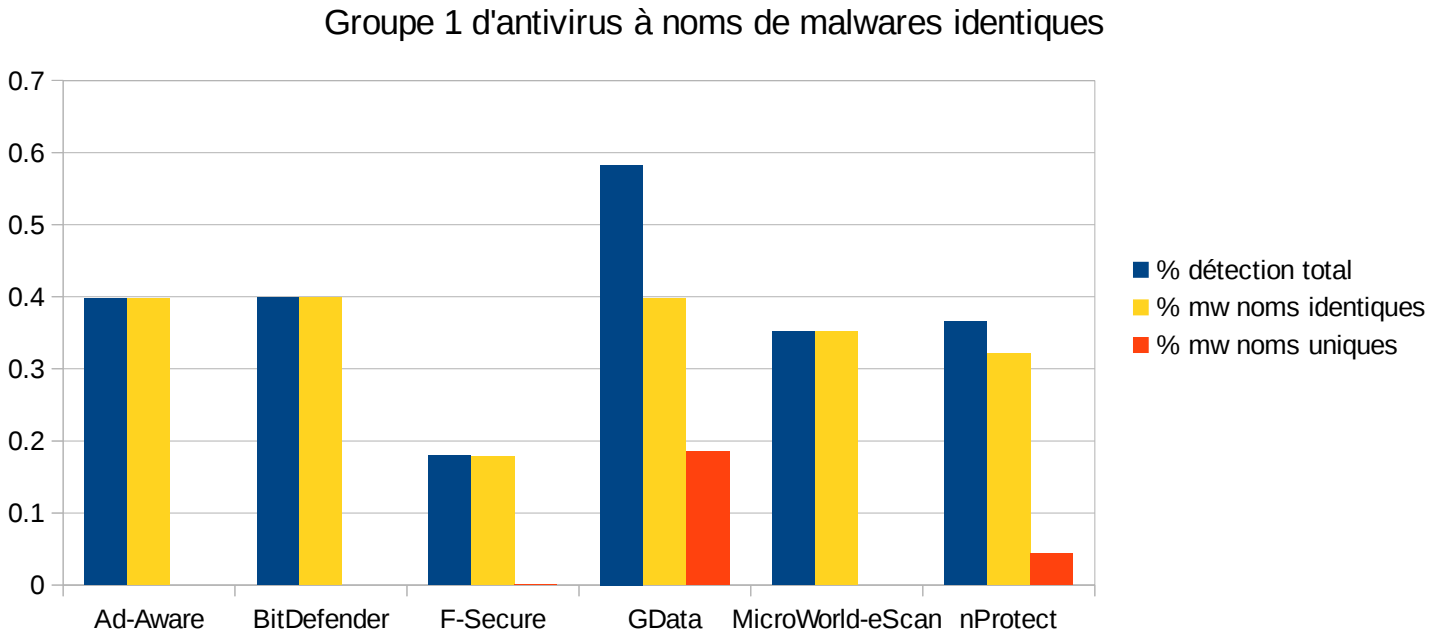
Le « nom complet » d'un malware, est donc le nom long incluant l'ensemble des mots clés utilisés pour l'identifier de manière unique. Le travail a été de compter le nombre d'utilisation de ces noms (lors d'une détection). J'ai regroupé ces nombres d'utilisations par classes, pour pouvoir créer un graphique de distribution du pourcentage d'utilisation des noms complets de malwares. Les différentes proportions de ces utilisations est résumé dans le graphique suivant.

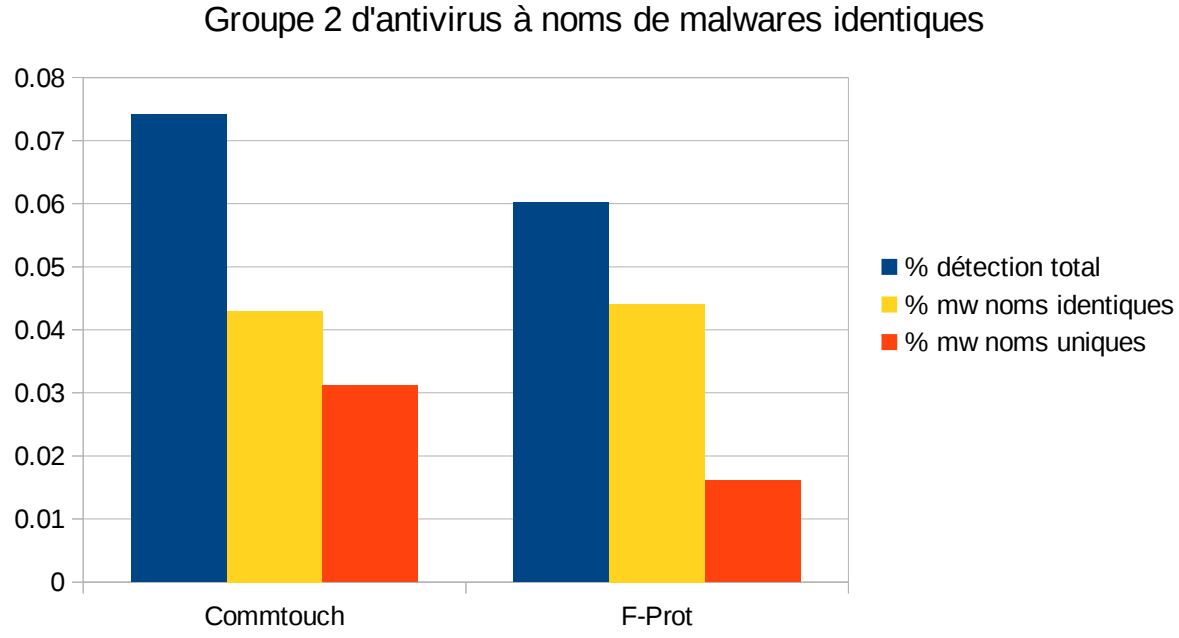


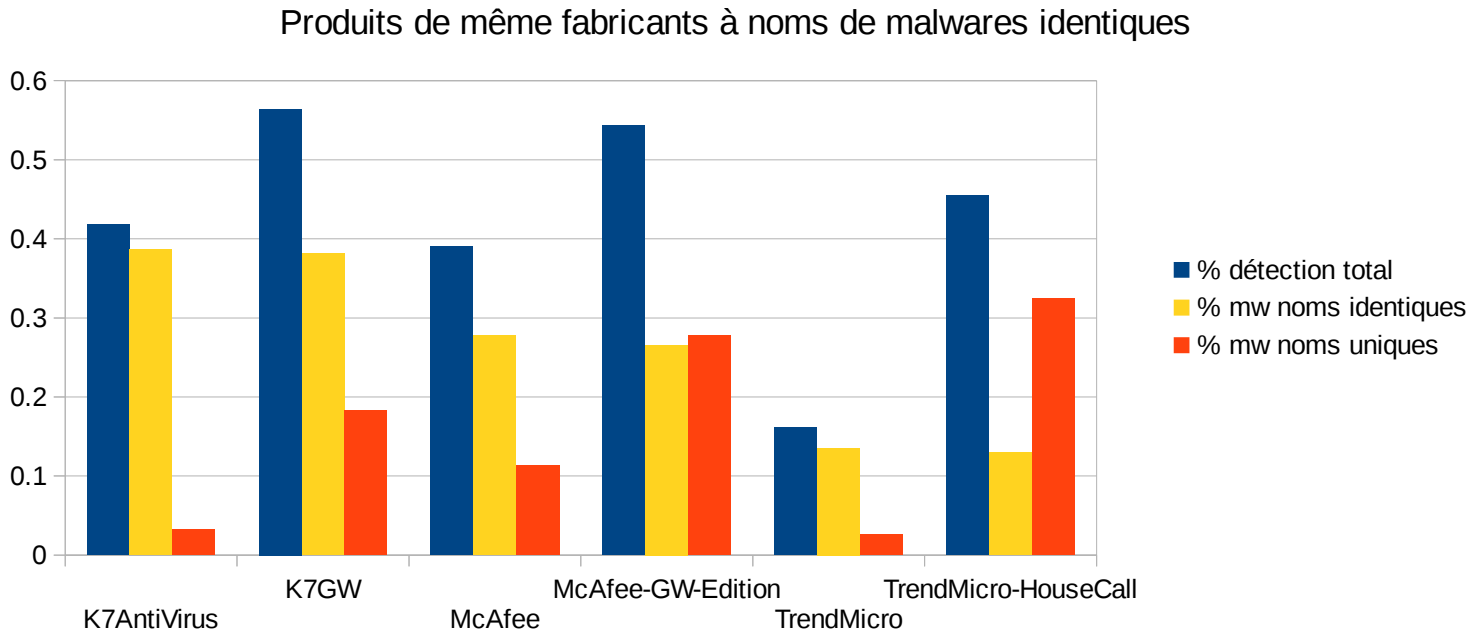
Nous voyons que plus de **60 % de ces noms complets de malwares** ne sont détectés qu'**une seule fois par un seul antivirus sur un seul fichier**. Ceci fait ressortir déjà le fait qu'il y a peu d'entente sur les noms de malwares entre les fabricants, et qu'il est très difficile de comparer leurs noms de malwares détectés.

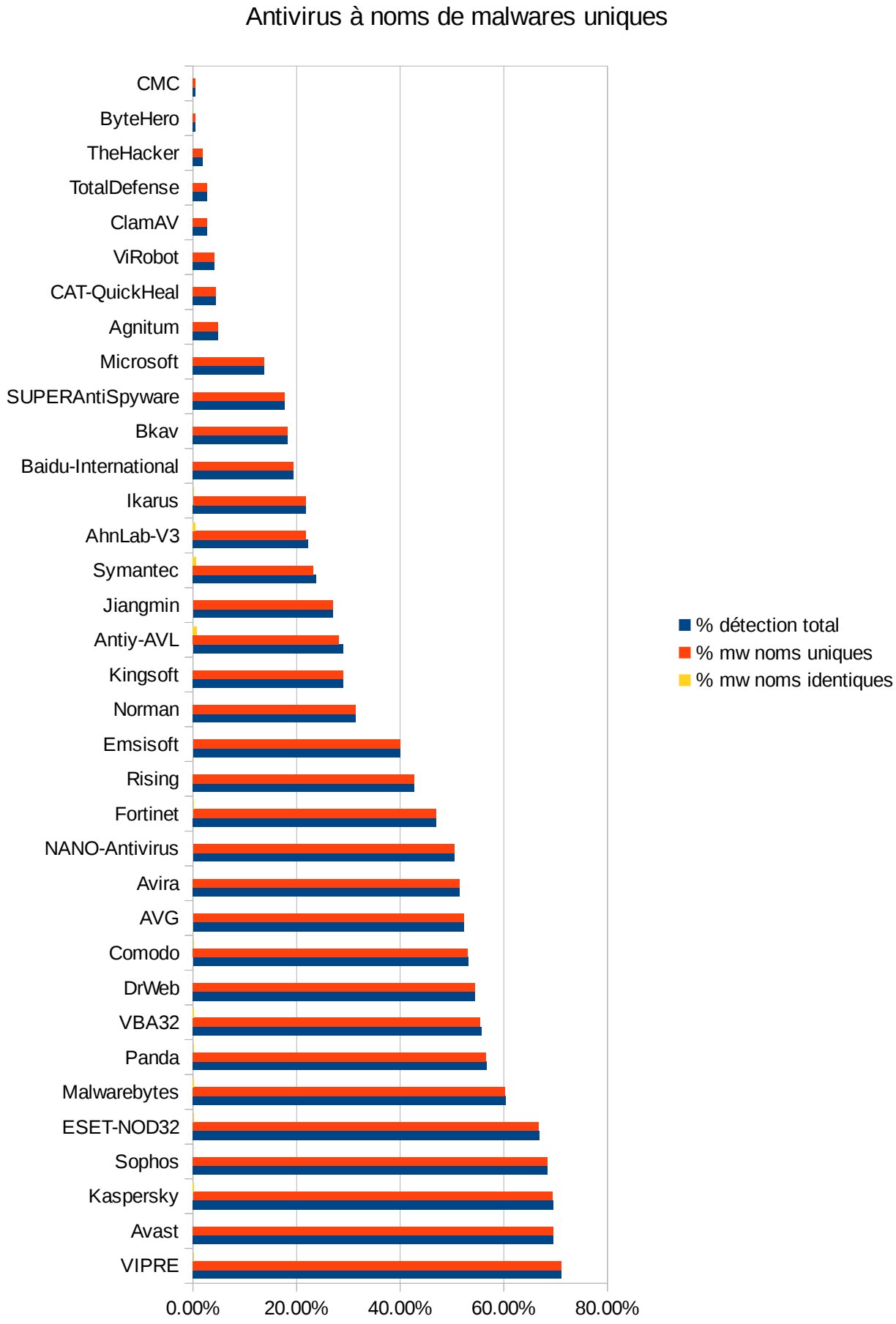
J'ai ensuite effectué une analyse plus en profondeur des différentes utilisations de ces noms complets par les différents fabricants. Il a notamment été possible de ressortir plusieurs groupes d'antivirus, utilisant de manière commune ces noms (soit des fabricants différents, ou des produits différents d'un même fabricant). Les différents graphiques résument ces groupes. On retrouve dans le premier groupe une partie de fabricants utilisant un moteur commun (BitDefender). Mais tous ne sont pas présents. **Donc certains fabricants utilisant un moteur commun utilisent aussi une nomenclature commune, mais pas tous**. On trouve dans le deuxième groupe deux fabricants qui

semblent utiliser une nomenclature commune sans pour autant avoir semble-t-il un moteur commun. Le troisième graphique montre l'utilisation des noms identiques pour les produits différents d'un même fabricant. Le dernier graphique est la liste de l'utilisation unique de noms par tous les autres antivirus.



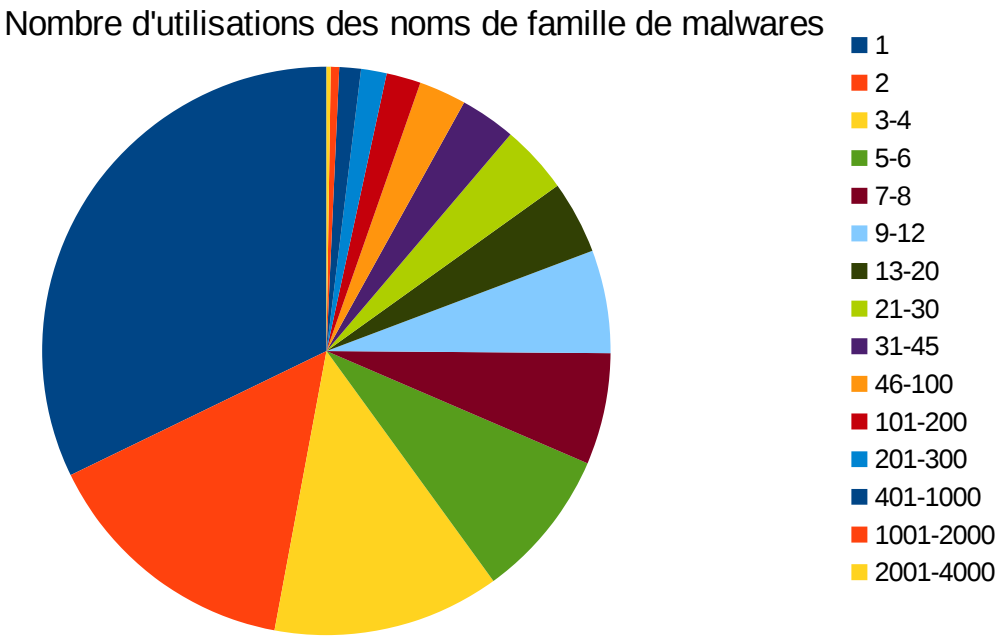



Produits de même fabricants à noms de malwares identiques
0.6
0.5
0.4
0.3
0.2
0.1
0
K7AntiVirus
K7GW
McAfee
McAfee-GW-Edition
TrendMicro
TrendMicro-HouseCall
% détection total
% mw noms identiques
% mw noms uniques


Antivirus à noms de malwares uniques
CMC
ByteHero
TheHacker
TotalDefense
ClamAV
ViRobot
CAT-QuickHeal
Agnitum
Microsoft
SUPERAntiSpyware
Bkav
Baidu-International
Ikarus
AhnLab-V3
Symantec
Jiangmin
Antiy-AVL
Kingsoft
Norman
Emsisoft
Rising
Fortinet
NANO-Antivirus
Avira
AVG
Comodo
DrWeb
VBA32
Panda
Malwarebytes
ESET-NOD32
Sophos
Kaspersky
Avast
VIPRE
0.00%
20.00%
40.00%
60.00%
80.00%
% détection total
% mw noms uniques
% mw noms identiques

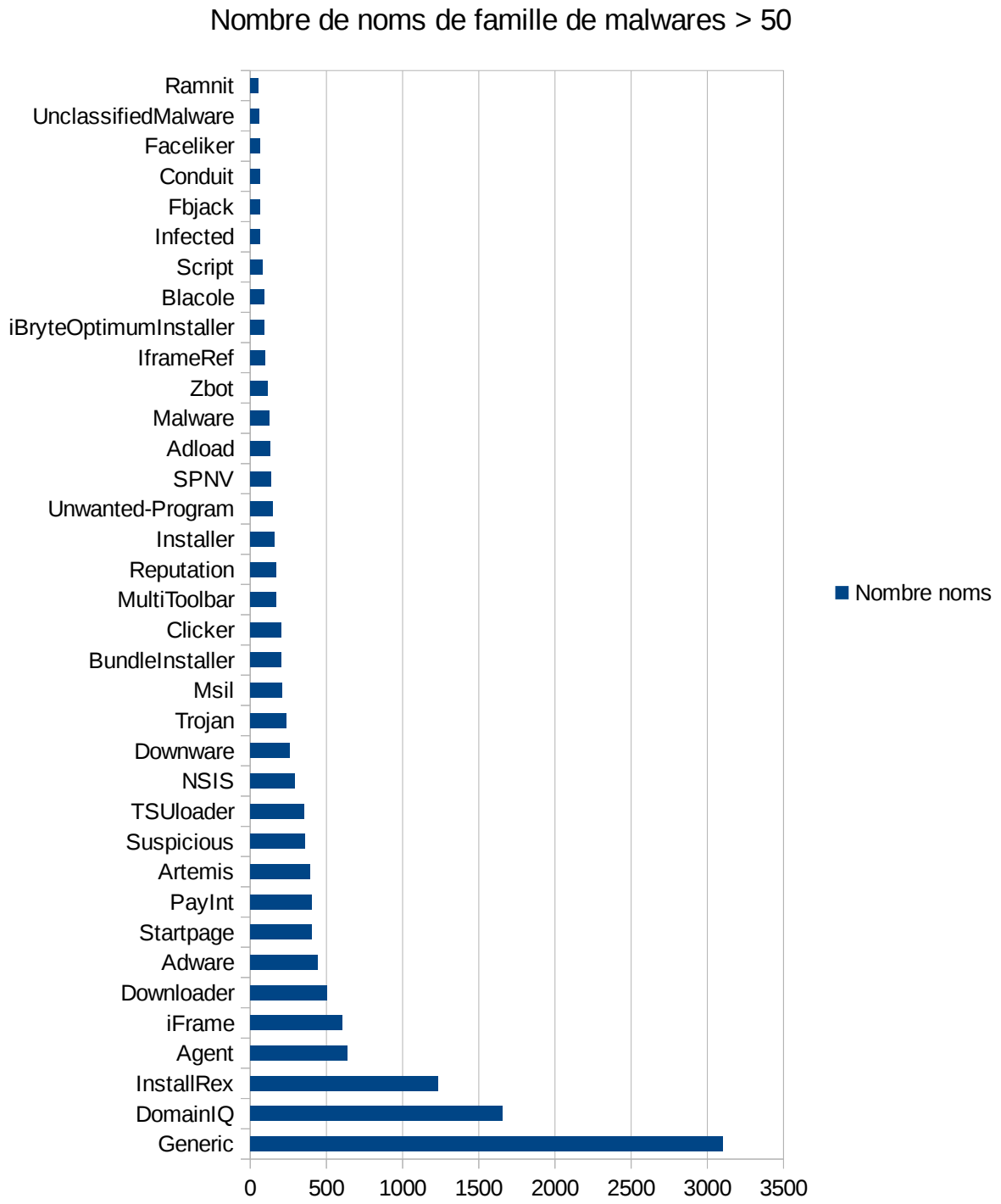
Noms de famille des malwares

La classification des types de mots clés m'a permis de faire ressortir les noms de familles des malwares (le nom usuel, court, utilisé pour identifier le malware). Dans le même ordre d'idée que pour l'analyse des noms complets, j'ai pu voir ici si le fait de réduire l'analyse sur le simple nom de famille permettait de mettre en évidence plus d'identifications ou nommages communs par les antivirus. Les proportions du nombre d'utilisation des noms de famille est résumé dans le graphique suivant.

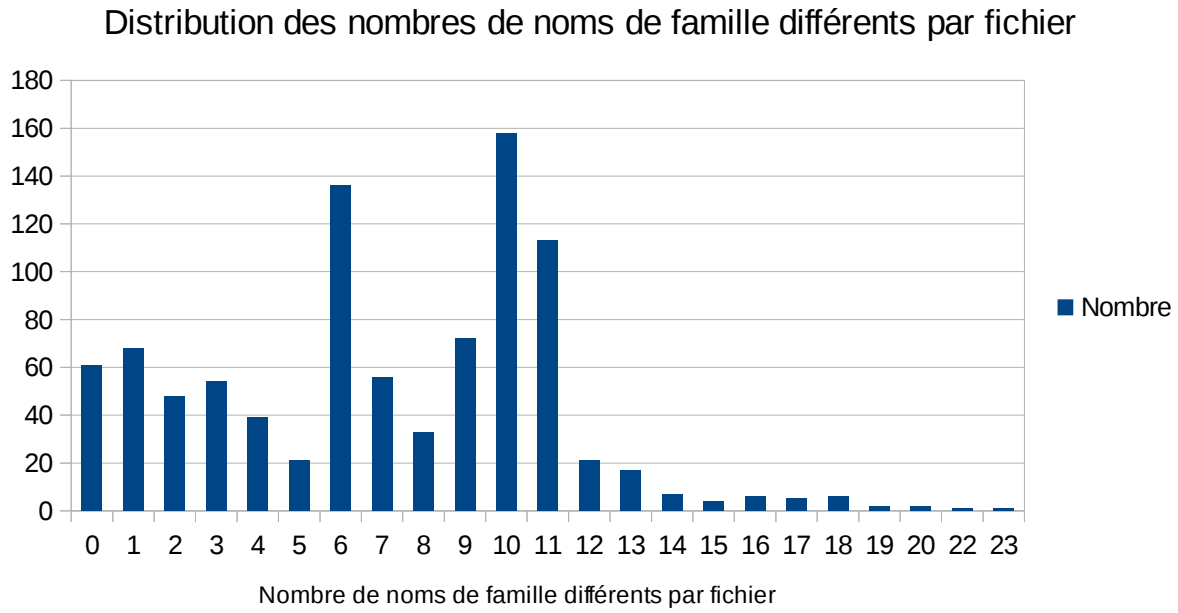


On voit que **32 % des noms de famille de malwares** n'ont été détectés que **par un seul antivirus sur un seul fichier**. On voit que même si on réduit l'analyse non plus sur le nom complet mais uniquement sur le nom de famille cette proportion reste importante. Il y a donc très peu d'échange au niveau de l'identification des malwares entre les fabricants d'antivirus.

Pour les noms de famille utilisés plus de 50 fois, voici la liste de ces noms classés par importance. Le nom le plus utilisé est un nom « générique » utilisé par les fabricants pour nommer les détections de nouveaux malwares qui n'ont pas encore été analysés en profondeur et classés (voir chapitre sur les types de plateformes/langages). Les noms suivants sont les noms des malwares les plus populaires à l'époque de la première date des définitions d'antivirus.



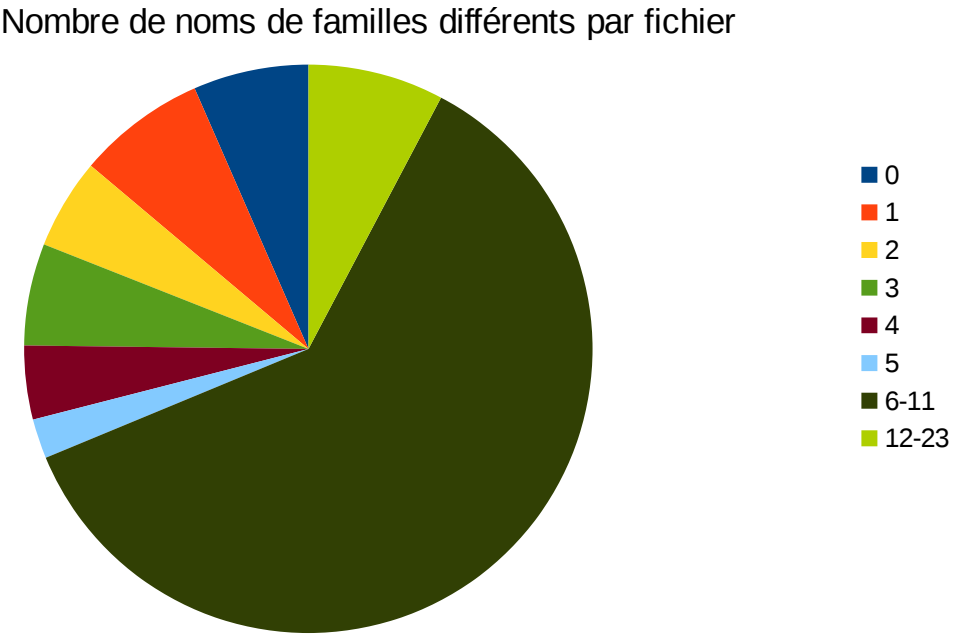
Un graphique du décompte du nombre de noms de familles différent par fichier est présenté ci-dessous.



Différents pics permettent de ressorti de nouveau certains malwares populaires.

Position :	**Nom du malware dominant :**
Pic 1	HfsIframe
Pic 6-7	InstallRex
Pic 8-11	DomainIQ

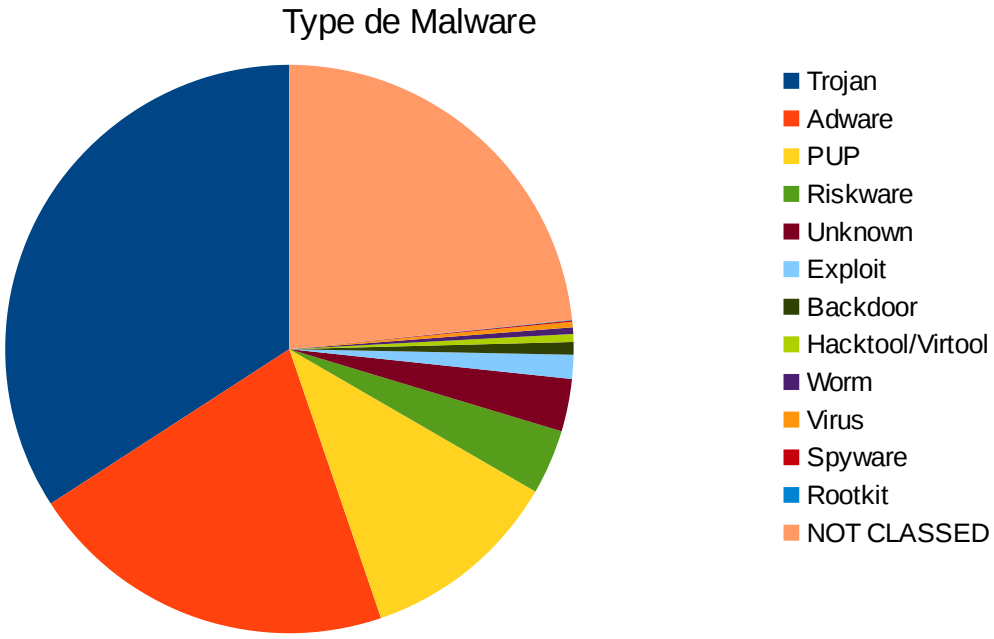
La présentation sur un graphique par secteur permet de ressortir la forte tranche entre 6 à 11 noms de familles différents par fichiers avec **une importance de plus de 60 %**. Ceci montre un grand nombre de noms différents utilisé par les fabricants d'antivirus avec peu d'échange entre eux sur ces noms ou leur détection.



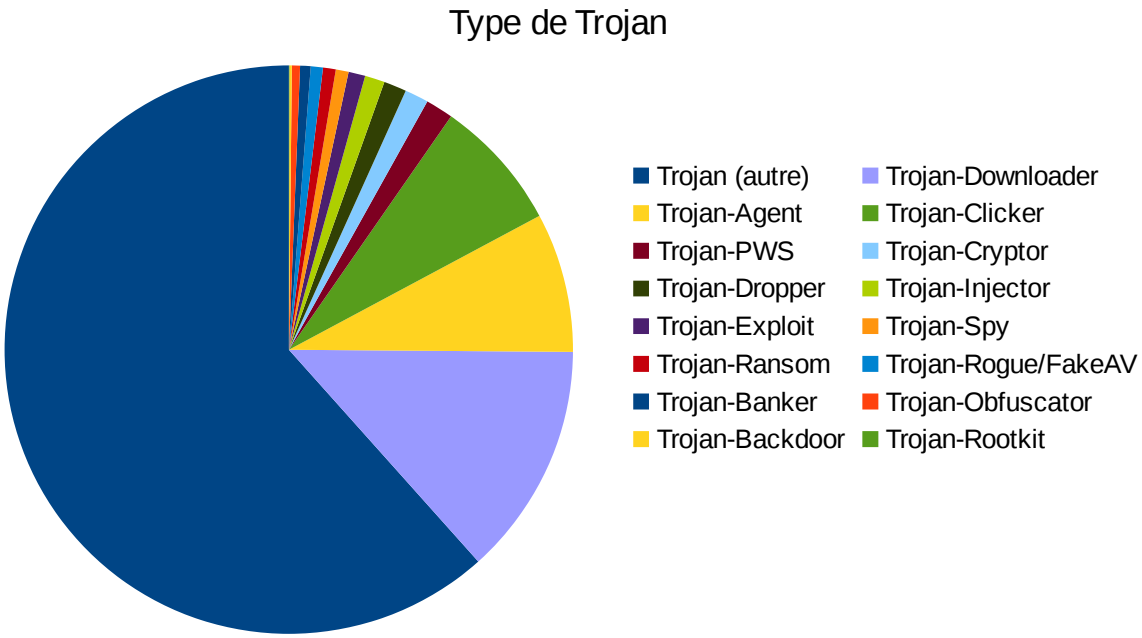
Types de malwares

L'étape suivante que j'ai menée a été de trier et classer parmi les mots clés composants les noms de malwares, afin de pouvoir les regrouper par catégories et comptabiliser leurs types de malware, plateformes/langages, etc. Pas moins de **1271 groupes d'expression régulières** ont été nécessaires pour trier parmi tes mots clés, qui ne sont pas utilisés de la même manière dans les divers conventions de nommage des fabricants d'antivirus. Durant ce traitement j'ai pu notamment remarquer que les fabricants d'antivirus utilisaient non seulement des noms de familles différents pour un même malware, mais classaient aussi des mêmes noms de famille dans des types de malwares différents, ou même utilisant des plateforme/langage différents. Une telle **disparité** non seulement dans la **convention de nommage**, mais aussi dans la **dénomination** du nom de famille principal, et en plus dans la **classification** dans un type de malware ou même une plateforme/langage, rend **toute comparaison extrêmement difficile et complexe**.

Durant ce travail le premier objectif a été de ressortir les types de malware identifiés par les fabricants d'antivirus. Pour la typologie des malwares, je me suis inspiré de la dénomination et de la classification donnée par Kaspersky (cf. : http://www.kaspersky.com/internet-security-center/threats/malware-classifications). Le premier graphique résume les proportions de types de malwares identifiés par les fabricants.



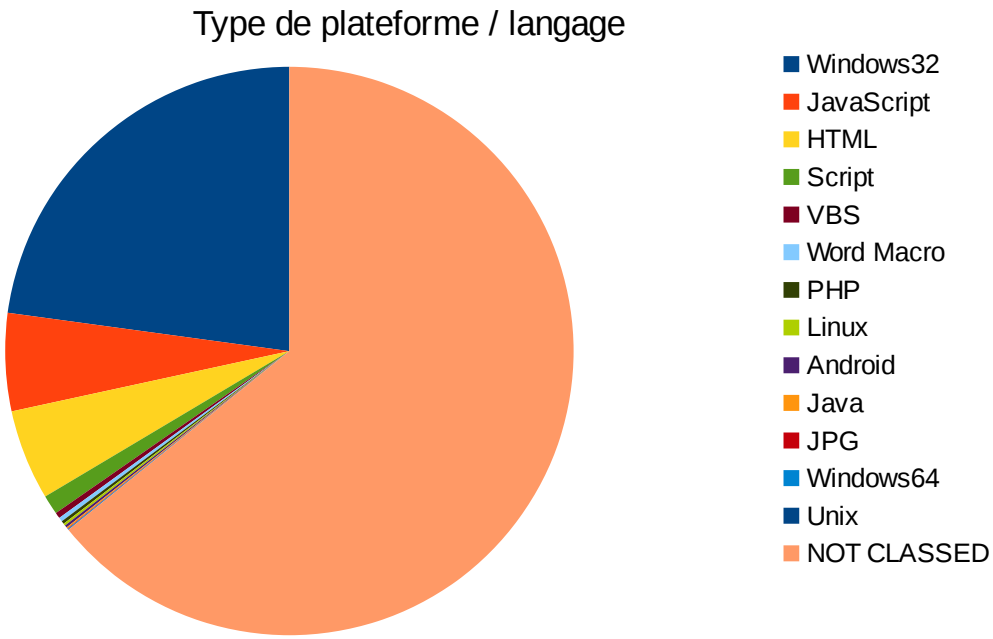
La première **catégorie la plus importante, des « Trojans » (chevaux de Troie), représente 34 %** des types identifiés. Mais les **Adware et autre PUP** (Potentially Unwanted Program), avec successivement **21 % et 11 %**, sont des catégories de gravités moindres mais importantes en quantité équivalant aux chevaux de Troie. En fonction des précisions des mots clés il a été possible pour cette première portion de chevaux de Troie, d'isoler le type de cheval de Troie du malware (voir la classification Kaspersky pour plus de précisions, et les divers pages descriptives des types par les divers fabricants). Le graphique suivant résume ces proportions de type de Trojan.



Les types les plus importants, **« Downloader », « Clicker », « Agent »** sont tous des chevaux de Troie qui permettent souvent d'infecter la machine avec d'autres malwares. On retrouve donc ici le but premier du cheval de Troie, qui est de s'implanter chez une victime avec une d'autres menaces qui suivent la première brèche.

Type de plateforme / langage

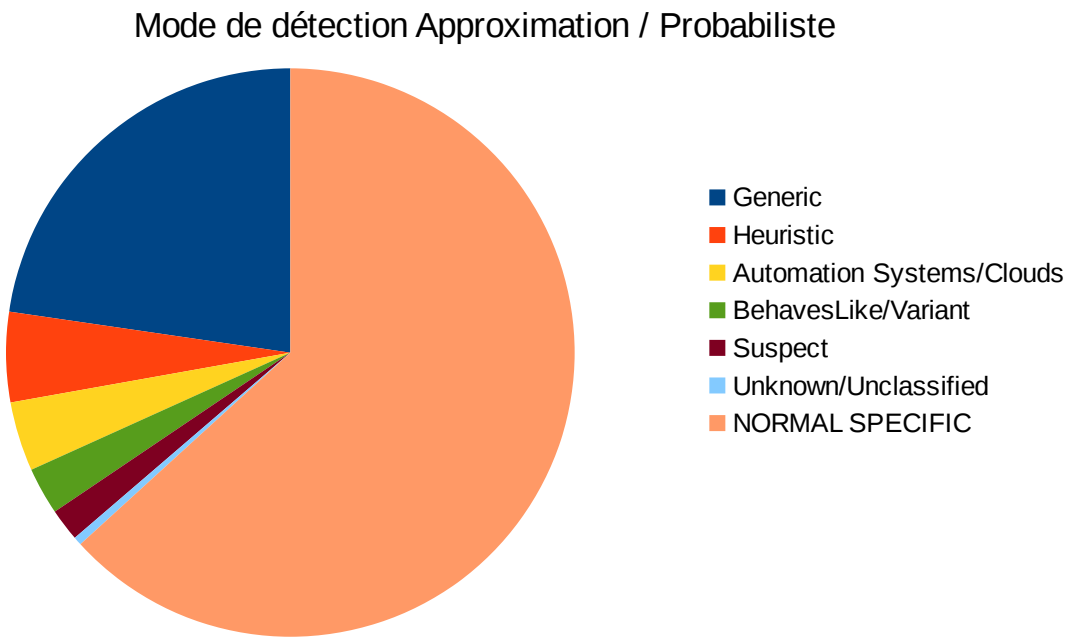
Le tri des mots clés a permis de ressortir aussi la **plateforme** sur laquelle de malware sévit, ou le **langage de programmation** utilisé par le malware. Le graphique suivant résume les proportions de ces informations.



On retrouve l'importance des **exécutables Windows** et des éléments constituant les **documents web** (**JavaScript** et **HTML**, à voir que les documents HTML incluent aussi les documents d'erreurs retournés lors de la récupération des malwares).

Mode de détection

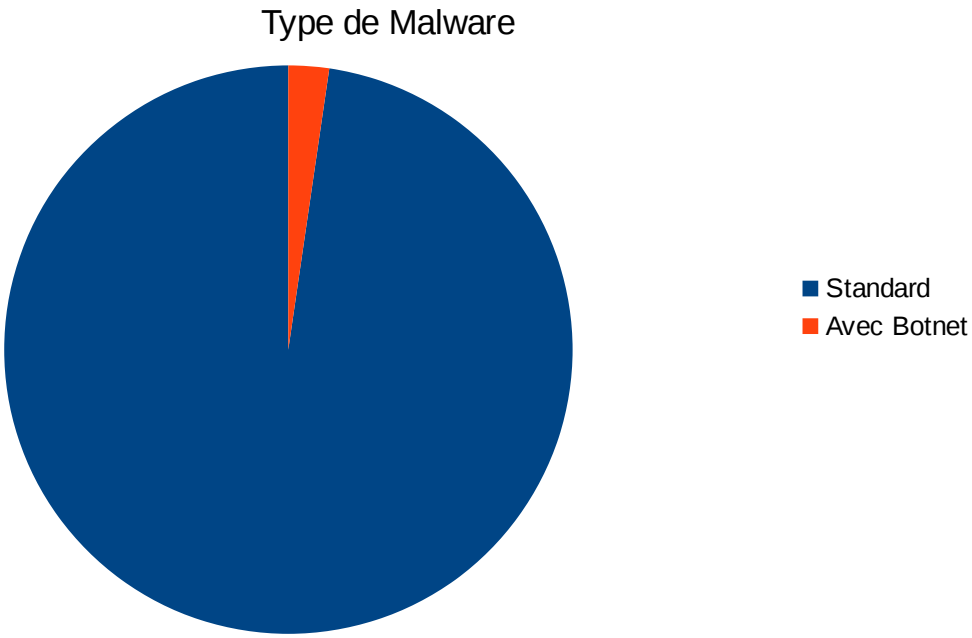
Certains mots clés permettent aussi d'identifier certains modes de fonctionnement utilisés par les antivirus pour la détection des malwares. Certains antivirus utilisent des **réseaux de détections** automatisés (dont réseaux « clouds »), qui permettent d'identifier automatiquement des nouvelles menaces de malwares avant qu'elles ne soient analysées plus intensément. Les fabricants utilisent aussi des algorithmes **heuristiques** pour identifier certains types de fonctionnements malveillants des exécutables. Ces types de détections sont plus incertains et sont donc plus une approximation avec une certaine probabilité d'être un malware. Le graphique suivant résume les malwares dont les mots clés dévoilent ces modes de fonctionnements.

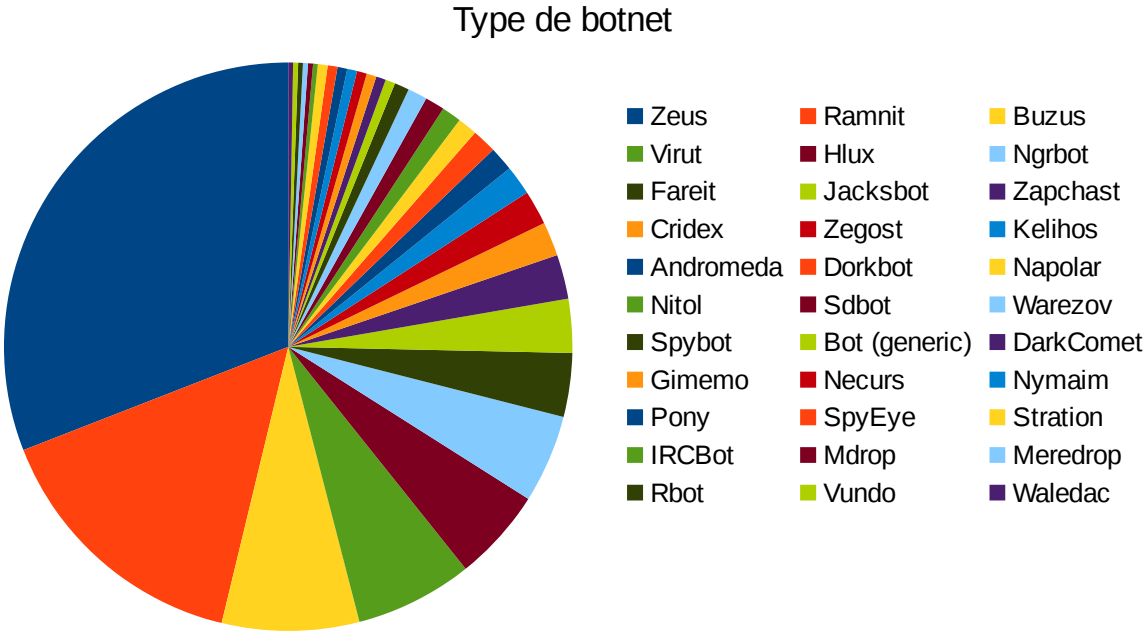


Les mots clés « **génériques** », avec une **proportion de 22 %** sont utilisés dans le premier nommage des malwares après les détections automatisées. Les mots clés « **heuristiques** » sont une deuxième part **avec 5 %** ainsi que les mots clés liés à des **systèmes d'automatisation et réseaux « clouds »** représente **presque 4 %.** Toutes les parts **non « normal spécifique »** représentent plus de 30 % des détections. On peut voir ici que le mécanisme de détection d'un malware est souvent complexe et demande parfois plusieurs étapes, qui sont mis en évidence par cet élément.

Botnets

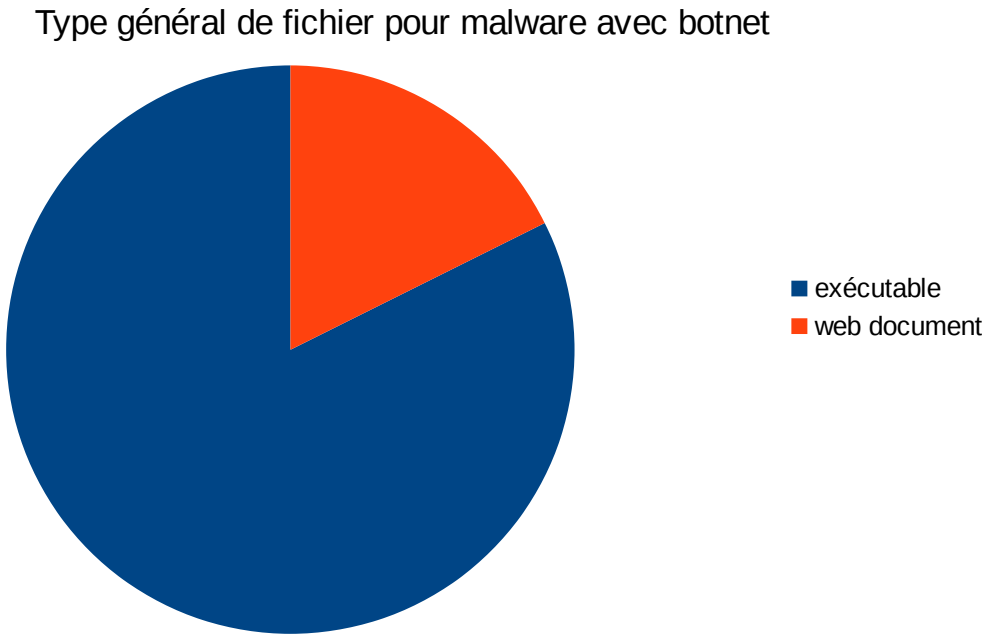
Le site botnets.fr recense les malwares ayant des fonctionnements « botnet » (parfois appelé « réseaux de zombies »). Certains malwares connectent l'ordinateur infecté à un réseau commun contrôlé par les créateurs du malware. Ceux-ci peuvent ensuite utiliser ce réseau via certaines commandes pour contrôler massivement tous les ordinateurs infectés, afin de leur faire faire des tâches spécifiques (attaques massives de sites web, envoi de campagnes de spam, etc.). Via les noms listés par le site botnet.fr et les noms de famille des malwares identifiés par les antivirus, il m'a été possible de comptabiliser la proportion de malwares ayant une fonction « botnet » ainsi que l'importance des divers botnets identifiés à travers les antivirus. Les graphiques suivants résument ces éléments.

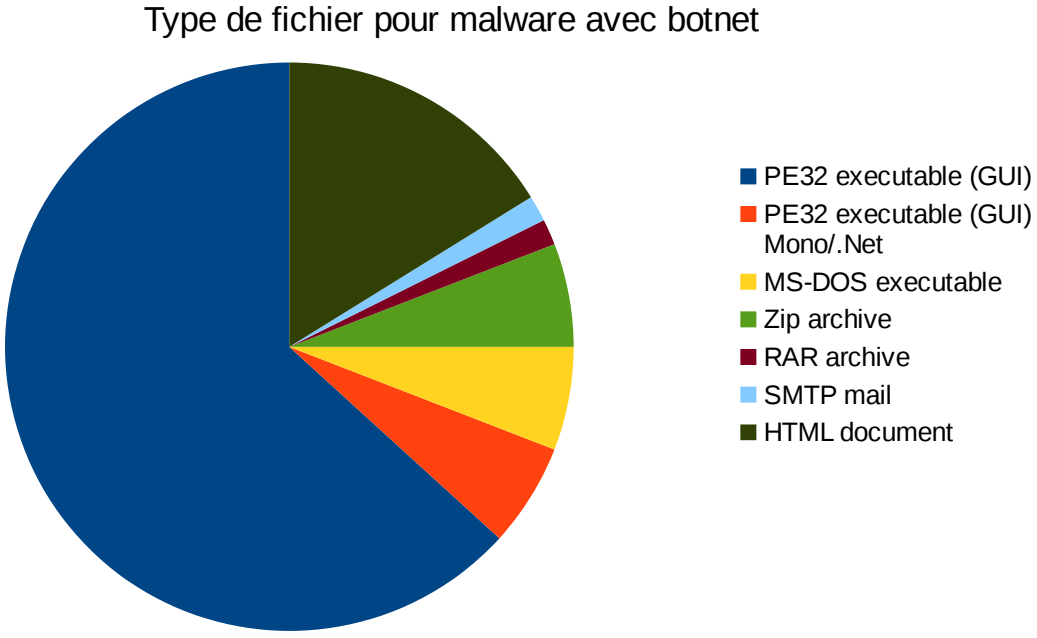




La première part est le cheval de Troie **Zeus** est un malware très populaire qui se connecte à des botnets très vastes, ici avec une **proportion de 30 % des malwares avec botnet**. Le deuxième **Ramnit** représente une **proportion de 15 % des malwares avec botnet**.

L'analyse du type de fichier lié à la détection d'un malware « botnet » permet de voir quels types de fichiers sont utilisés pour connecter des ordinateurs infectés à des botnets.

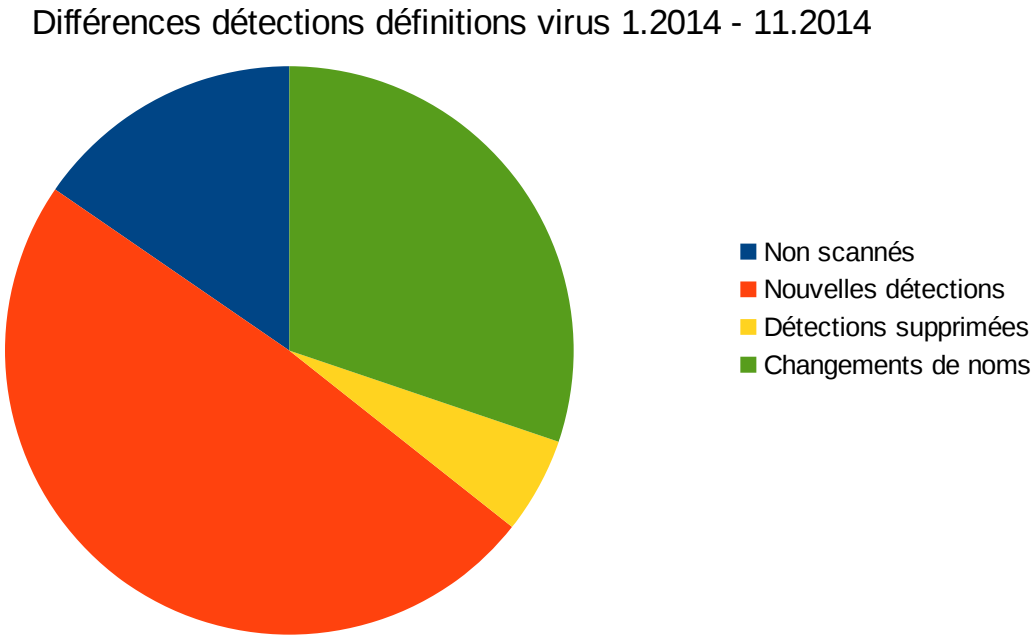




On retrouve une forte prédominance des exécutables, mais pas seulement. **Les pages web permettent aussi d'infecter des machines pour les relier à des botnets**.

Mise à jour des définitions de virus

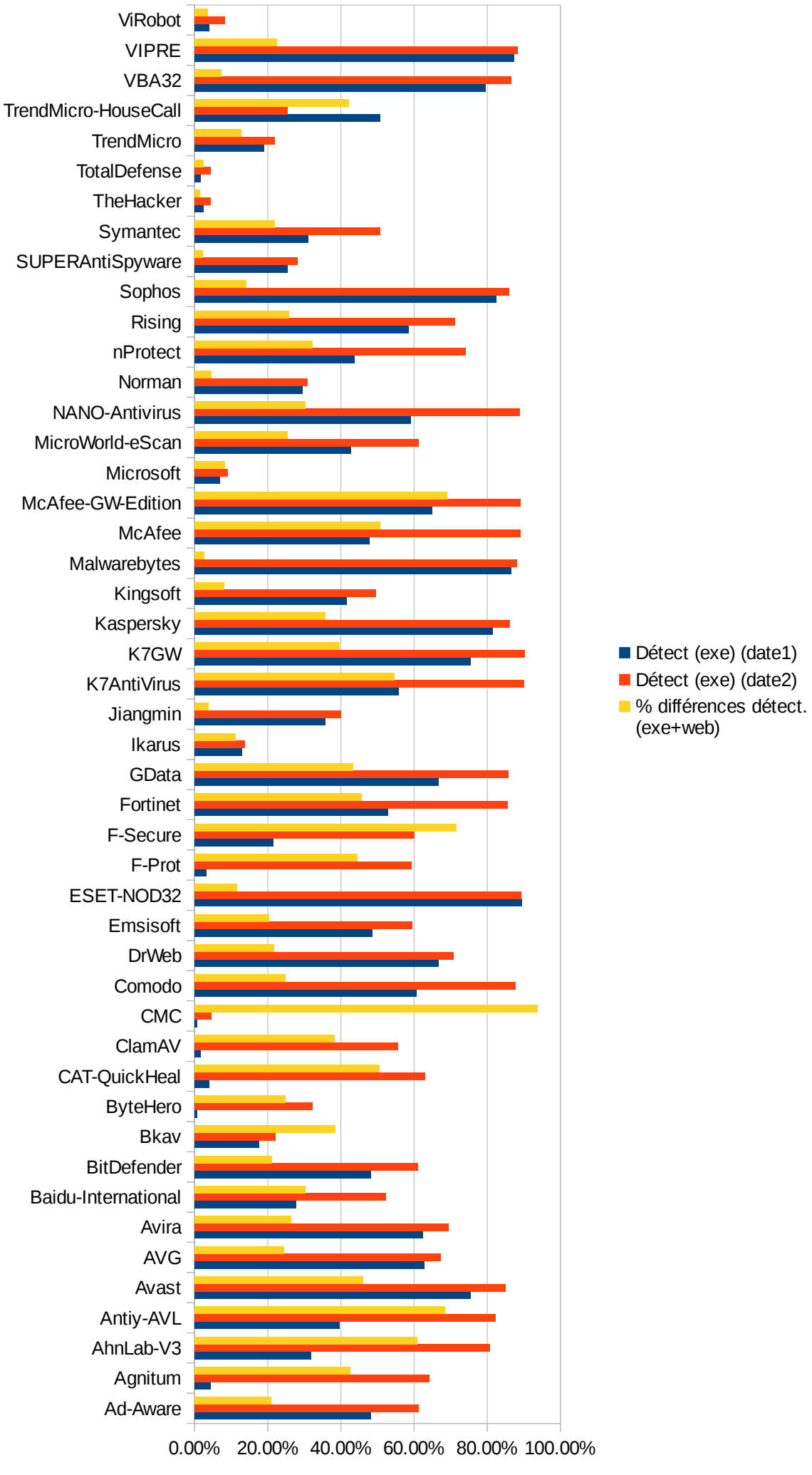
La dernière étape que j'ai effectuée était d'analyser l'évolution au cours du temps des définitions de virus des fabricants et voir comment la détection des malwares et le nommage de ceux-ci pouvaient éventuellement changer avec les mises à jour. L'analyse a été de comparer les données de détections fichier par fichier pour les deux dates : janvier 2014 et novembre 2014 (voir chapitre « Méthodologie »). Entre ces deux date, il y a eu un peu moins de **30 % de différences** dans les détections pour tout antivirus. Le premier graphique résume les proportions de ces changements de détections.

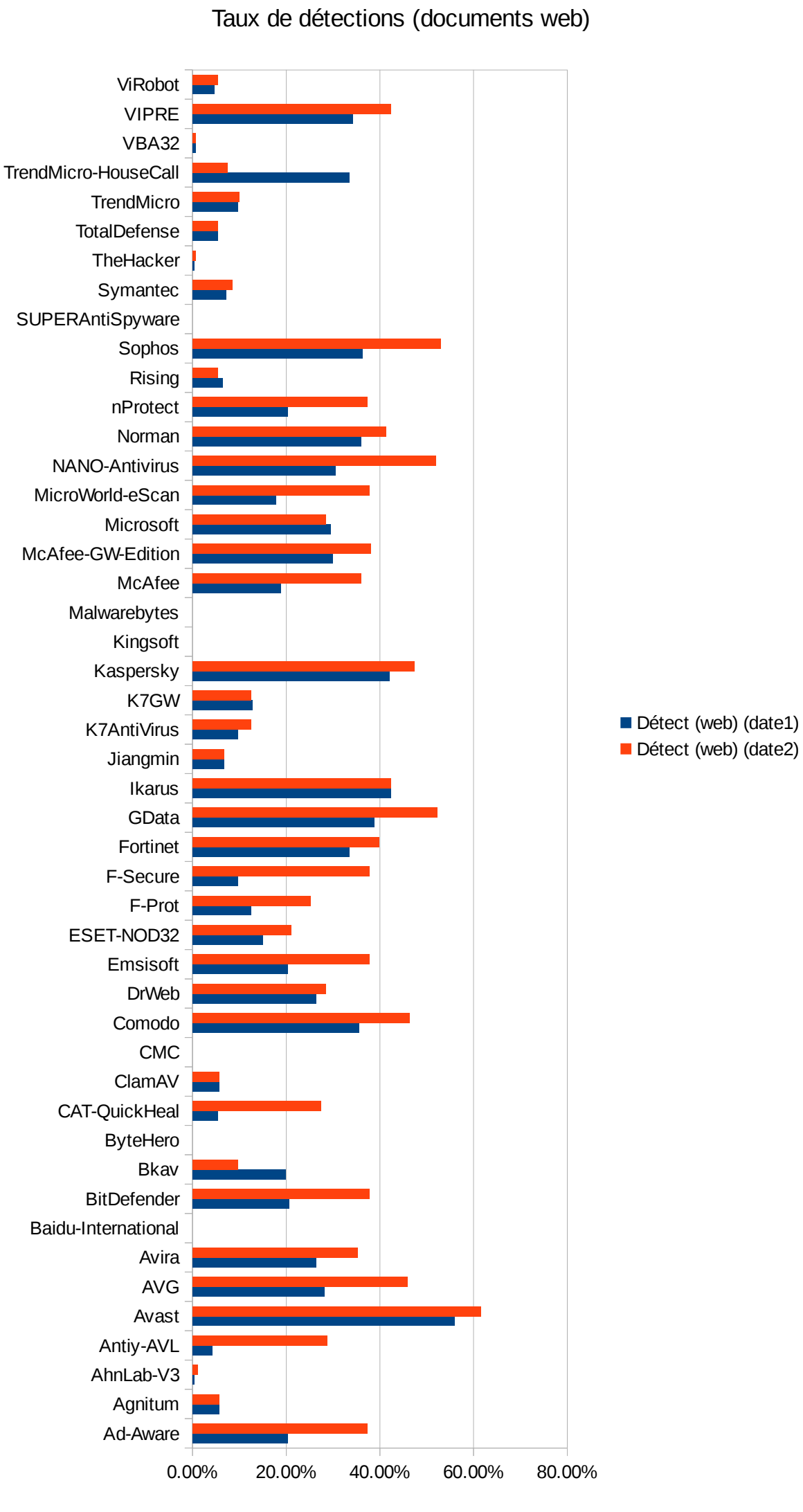


La part la plus importante, soit **près de 50 % (sur la part des changements), représente les nouvelles détections**. La proportion **des changements de noms, avec 30 %**, montre le fonctionnement décrit précédemment. Certaines détections automatisées sont utilisées par certains antivirus, souvent classés avec des noms « génériques », qui sont ensuite nommés plus précisément quand l'analyse en profondeur du malware a été effectuée. La part des « non scannés » représente un problème lors de la détection par l'antivirus soit à la date1 soit à la date2, ceci amenant une erreur qui empêche d'avoir les données et donc de pouvoir faire une comparaison.

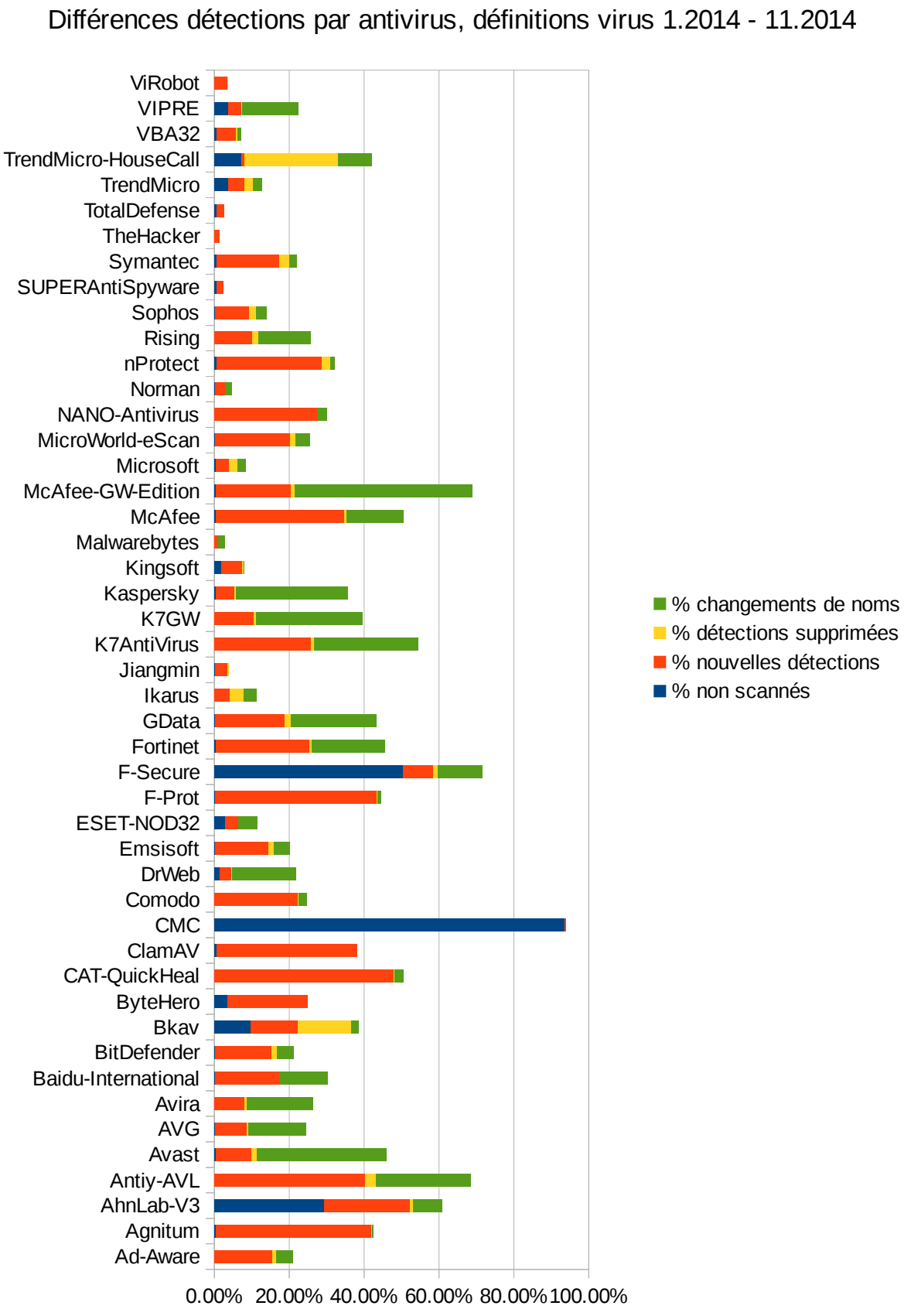
Le graphique suivant montre les taux de détections des exécutables aux deux dates, ainsi que le pourcentage de toutes les différences (non scannés, nouvelles détections, détections supprimées, changements de noms, pour les exécutables et les documents web) entre les deux dates pour chacun des antivirus. On voit que beaucoup d'antivirus ont un taux de détection qui devient proche à la deuxième date. Bien qu'il ne soit pas possible de distinguer les « faux-positifs » dans ces détections, ceci montre qu'avec le temps la plupart des antivirus détectent les exécutables de manière similaire. Le graphique suivant montre les taux de détections des documents web aux deux dates. On peut voir que, contrairement aux exécutables, la disparité des taux entre les antivirus pour les documents web reste beaucoup plus présente.

Taux de détections (exe), % différences détections défintions 1.2014 - 11.2014




Taux de détections (documents web)
ViRobot
VIPRE
VBA32
TrendMicro-HouseCall
TrendMicro
TotalDefense
TheHacker
Symantec
SUPERAntiSpyware
Sophos
Rising
nProtect
Norman
NANO-Antivirus
MicroWorld-eScan
Microsoft
McAfee-GW-Edition
McAfee
Malwarebytes
Kingsoft
Kaspersky
K7GW
K7AntiVirus
Jiangmin
Ikarus
GData
Fortinet
F-Secure
F-Prot
ESET-NOD32
Emsisoft
DrWeb
Comodo
CMC
ClamAV
CAT-QuickHeal
ByteHero
Bkav
BitDefender
Baidu-International
Avira
AVG
Avast
Antiy-AVL
AhnLab-V3
Agnitum
Ad-Aware
0.00%
20.00%
40.00%
60.00%
80.00%
Détect (web) (date1)
Détect (web) (date2)

Le graphique suivant montre le détail des différences des détections (non scannés, nouvelles détections, détections supprimées, changements de noms, pour les exécutables et documents web) entre janvier 2014 et novembre 2014 par antivirus. Le pourcentage est calculé par rapport au nombre total de fichiers.



On voit qu'il y a beaucoup de différences. Certains antivirus utilisent plus les changements de noms, donc des systèmes de détections automatisés avec noms génériques corrigés par la suite lors d'une analyse plus poussée. Certains antivirus ont beaucoup de nouvelles détections entre les deux dates, et d'autres au contraire très peu, le taux de détection de la première date étant très proche de celui de la seconde.

Conclusion

Cette analyse montre un écosystème des antivirus qui évolue globalement et essaye de s'adapter à l'évolution de l'écosystème adverse des malwares qu'il essaye de combattre. Même si on voit certains regroupements qui peuvent exister au niveau des moteurs d'antivirus, ou d'un certain échange au niveau des noms, globalement il y a encore très peu d'échanges entre les fabricants. Cet élément, s'il peut être un avantage commercial pour se démarquer de concurrents, est certainement **un inconvénient au niveau de la qualité de détection**, mais aussi **au détriment de la réactivité des antivirus** par rapport aux nouvelles menaces extrêmement agiles des malwares qu'ils combattent. Certains fabricants adoptent des stratégies d'automatisation via les algorithmes heuristiques ou les réseaux « clouds ». Par contre chacun travaillant majoritairement de son côté ils bénéficient **peu des avantages de groupe** que certains organismes de la nature possèdent en adoptant des stratégies de groupe plus poussées. Avec encore trop peu de mise en commun de nommages des malwares, toute stratégie de groupe devient très difficile.

Tout l'ensemble de l'environnement de ces deux écosystèmes est très changeant. Il est donc important de garder à l'esprit que **toute interprétation est fortement relative** à la période d'analyse, à la source des données analysées, ainsi qu'aux configurations des outils de détections. Garder une vue macroscopique est important mais difficile. De plus **une analyse continue** sur le long terme pour prendre en considération cette forte volatilité est **complexe à mener**.