info·SIRCO
ENQUÊTE ET PROTECTION
Septembre 2011

LES LIMITES DE LA RECHERCHE PAR MOTS-CLES POUR LA PREUVE ELECTRONIQUE

--Droit et sécurité--

La preuve électronique est un changement important dans l'univers de l'admissibilité de la preuve en Cour. Considérant que 93% de toute l'information générée dans les entreprises l'est sous forme digitale[1], la recherche de preuve doit être fiable pour être admise alors que la Cour a longtemps fonctionné sur le format papier.

La technique de recherche par mots-clés afin de trouver des preuves informatiques éprouve de fortes difficultés quant à la production d'une preuve fiable et admissible selon les critères judiciaires. En effet, elle peut produire des erreurs dans son exécution, mais surtout, elle n'est pas infaillible. De nombreuses méthodes consistant à cacher les documents et les fichiers voulus rendent alors la recherche par mots-clés infructueuse.

La nature des ordonnances

Un aspect n'est pas à négliger lors de la recherche de preuve informatique : les ordonnances attribuées pour ces recherches. En voici quelques exemples :

- **Injonction Mareva** : Généralement liée au cas de fraude, cette injonction permet de suspendre toute action du défendeur afin d'éviter la destruction de preuve.
- **Ordonnance Anton Piller** : Autorise les représentants de la partie demanderesse à pénétrer dans les locaux d'affaires ou autres lieux de la partie défenderesse pour effectuer des recherches sur le matériel visé par l'ordonnance.
- **Mandat de perquisition-Loi sur les faillites** : Permets de récupérer les actifs ou les documents ayant appartenu à la faillie qui ont été détournés ou volés précédemment à la faillite.

[1] Recherche menée par l'Université de Californie aux États-Unis en 1999.

Erreur de type 1 : « under inclusion »

La recherche par mots-clés implique la détermination de certains termes qui vont être recherchés par le système à travers les différentes bases de données et documents. Cependant, la requête donnera uniquement les documents ou fichiers qui disposeront du terme exact. Ainsi, bon nombre de documents pertinents peuvent rester inconnus puisqu'on ne peut pas déterminer *a priori* tous les termes exacts qu'il faudrait inclure dans la recherche pour être sûr d'obtenir tous les documents pertinents. À titre d'exemple, si vous faîtes une recherche avec le mot-clé « voiture », vous passerez à côté de tous les documents comportant les marques (BMW, Audi, etc.) ainsi que d'autres termes similaires (véhicule, automobile, etc.). De plus, et notamment dans les messages, les personnes se parlent sous forme abrégée ou mal orthographiée que la requête ne pourra trouver.

Erreur de type 2 : « over-inclusion »

À l'inverse, en ajoutant de nombreux mots-clés pour éviter la sous inclusion, la recherche donnera trop de résultats positifs, mais qui se révèleront être non pertinents après une analyse longue et coûteuse. Les documents recherchés seront donc noyés dans l'ensemble des documents et fichiers relevés par les requêtes. Par ailleurs, les termes spécifiés sont relevés indifféremment de leur contexte et certains mots-clés notamment les acronymes n'auront peut-être rien en commun avec la recherche voulue. Par exemple, l'acronyme AAF peut vouloir dire Army Air Forces ou bien American Advertising Federation et ils n'ont rien à voir en commun.

Les logiciels de camouflage et de dissimulation

Certains sites, forum sur Internet donnent des astuces pour pouvoir cacher des documents ou des dossiers complets. Les individus qui exposent ces conseils ne s'en cachent d'ailleurs même pas et évoquent entre eux les moyens pour cacher de la pornographie ou des documents sensibles. Des logiciels gratuits et performants tels que *Folder Hidden*, *Folder Lock*, *Easy file locker*, *Folder hider*, *MysecretFolder* permettent de créer des aires privées

cachées, cryptées et sécurisées avec pour certains des compléments tels que le nettoyage de l'historique, un écraseur de données ou encore un mode de navigation cachée.

Un enquêteur chevronné, utilisant les logiciels appropriés, peu démasquer la plupart de ces données cachées sur un disque dur. Aussi, la décryptions est possible, en utilisant la méthode « brute force cracking », qui brise la plus part des encryptions. La seule variable est le temps ! Une encryptions peut être brisée en 1 jour, ou bien en 3 mois.

La transformation des documents

- **La traduction** : Une des techniques les plus simples à exécuter consiste à prendre un document rédigé sur Microsoft Office Word, utiliser la traduction automatique du programme, sauvegarder la version traduite et effacer toutes les versions précédentes. Un logiciel de suppression (exemple : CCleaner) permettra d'effacer définitivement les documents.
- **Le cryptage et la stéganographie**: L'encryption d'un document permet de coder le document de façon à ce que personne ne puisse lire son contenu. Cependant, cette technique laisse des traces. Les sites et forums d'astuce préfèrent alors conseiller de transformer un document en image, c'est la stéganographie. Par ce procédé, les pixels du texte sont codés, cryptés et dilués dans les pixels de l'image et ainsi les données « sensibles » n'attirent pas l'attention et la recherche par mots-clés est inutile. Le logiciel libre *StregHide* permet de cacher un fichier ou un message texte dans une image (BMP ou JPG) ou un fichier son (WMA) et de le crypter.
- **La modification du nom** : Le nom du dossier ou du fichier peut être tout simplement par un autre terme et donc la recherche par mot-clé ne permettra pas de trouver.

La recherche par mots-clés ayant révélé ses faiblesses, le système judiciaire commence à se questionner sur la validité d'une telle méthode de recherche. Cette méthode utilisée seule ne rencontre pas les standards légaux pour une preuve judiciaire électronique.