WOJCIECH SMOL

J'ai vos (méta)données !

Degré de difficulté

Les métadonnées constituent une sorte d'ADN des documents chiffrés. Regardez comment les cybercriminels sont capables d'utiliser les informations invisibles présentes dans les fichiers partagés publiquement.

CET ARTICLE EXPLIQUE

La notion des métadonnées,

Comment les cybercriminels peuvent utiliser les métadonnées,

La notion des outils employés par les crackers pour rechercher et traiter les métadonnées,

Quelques cas authentiques d'une utilisation surprenante de métadonnées,

Les bonnes pratiques dans l'administration de la sécurité des (méta)informations.

CE QU'IL FAUT SAVOIR

Connaître les questions de base relatives aux formats de fichiers les plus populaires,

Connaître les types élémentaires d'attaques dont l'objectif consiste à collecter des informations,

Connaître les opérateurs avancés utilisés dans le navigateur Google.

Dans la réalité numérique d'aujourd'hui, la plupart d'organisations génèrent, stockent et archivent leurs données sous forme numérique. Les systèmes informatiques intégrés de la classe ERP (en anglais *Enterprise Resource Planning*) permettent d'éviter d'enregistrer et de stocker des centaines de milliers des feuilles de papier. Les aspects positifs des technologies modernes ne sont pas toutefois à l'abri des risques relatifs à la sécurité d'informations, inconnus auparavant.

Ces risques sont liés à une caractéristique d'information numérique, à savoir ses attributs supplémentaires, invisibles au premier abord. Les informations supplémentaires qui caractérisent l'information principale, telles que la date de création ou de la dernière modification, s'appellent les métadonnées (en anglais *metadata*). Ce sont des données sur les données – c'est une description la plus brève de ces structures. Elles constituent un problème supplémentaire dans le domaine de sécurité des informations, quasiment inconnu dans le cas de données traditionnelles. Les métadonnées sont souvent ignorées dans les procédures de sécurité informatique alors que je démontrerai qu'elles peuvent provoquer une fuite importante de donnés dans les piles. Cette situation résulte probablement du fait qu'elles sont en apparence *invisibles* et qu'un problème similaire n'existe pas dans le cas de données traditionnelles. Ces points ne justifient pas tout de même les personnes responsables de la sécurité d'informations. Tous ceux qui pensent que la protection appropriée des données principales résout complètement le problème de sécurité des métadonnées se trompent aussi.

Il ne faut pas compter dans cette question sur l'ignorance des cybercriminels. Un cracker expérimenté sait sûrement manipuler les métainformations, parmi d'autre choses. Il doit non seulement utiliser les métadonnées pour collecter des informations pour attaquer mais aussi effacer les métadonnées générées lors de son activité criminelle (les métadonnées constituent l'un des objets de base qui intéressent l'informatique légale).

(Méta)données

Je n'ai pas utilisé l'orthographe *(méta)données* par hasard. Il n'est pas simple en effet de distinguer les données et les métadonnées. Nous sommes incapables d'indiquer des différences en analysant la nature de deux types d'informations. Il s'agit tout simplement des informations sous forme numérique. C'est le contexte, dans lequel nous envisageons l'information concrète, qui décide en pratique des différences. À titre d'exemple, les paroles d'une chanson enregistrée dans un fichier texte constituent des données. Si toutefois les mêmes paroles sont inclues au fichier son avec l'enregistrement de la chanson, les mêmes données deviennent des métadonnées. Il est donc impossible de distinguer ces deux points sans avoir d'informations supplémentaires sur

Figure 1. *Excursion virtuelle autour d'une maison où une photo numérique a été prise : tout cela est possible grâce aux métadonnées contenues dans le fichier !*

le contexte de l'information analysée. La tâche principale des métadonnées consiste à fournir des informations permettant d'interpréter correctement et d'utiliser l'information principale. À titre d'exemple, une suite de caractères *75015* n'apporte pratiquement aucune information utile. Si les métadonnées *code postal en France* l'accompagnent toutefois, cela permettra de les utiliser correctement comme les données d'adresse. Une autre utilisation importante de métadonnées, qu'il faut évoquer, consiste à accélérer et à permettre une recherche d'informations principales à plusieurs critères (d'après de nombreux attributs qui caractérisent les données).

Nous pouvons considérer que les métadonnées accompagnent actuellement quasiment toutes les données numériques. Des exemples de structures de données, contenant des métainformations auxquelles il faut faire attention sont les suivants :

- fichiers graphiques – comme le format JPG particulièrement populaire,
- documents électroniques – tels que DOC et DOCX,
- documents au format universel PDF.

Tous les trois types susmentionnés constituent un risque réel important pour la sécurité d'informations en raison de métadonnées qu'ils utilisent. C'est un risque aussi bien pour les grandes organisations que pour les particuliers. Peu de personnes s'en rendent compte toutefois.

Une photo vaut plus que mille mots

Un proverbe connu dit qu'une image vaut plus que mille mots. Dans le cas des photos numériques, cette phrase prend une signification toute particulière.

La plupart d'appareils numériques créent des fichiers graphiques par défaut au format JPG. Ce n'est pas tout : ces appareils enregistrent le plus souvent aussi les métadonnées dans le fichier au format *Exif* (en anglais *Exchangeable Image File Format*). Sans entrer dans les détails de la spécification Exif, il faut savoir que les métabalises décrivent dans ce standard notamment :

- le nom de l'appareil avec lequel la photo a été prise,
- les paramètres de l'appareil, tels que durée d'exposition, valeur de l'écran, sensibilité de la matrice, etc.,
- la date de la prise de la photo,
- la résolution en pixels,
- la miniature de l'image,
- les coordonnées géographiques exactes de l'endroit où la photo a été prise.

Le problème est lié au fait que au moment de partager les photos prises, les particuliers et les grandes organisations font rarement attention aux métadonnées publiées avec !

La plupart d'attributs susmentionnés, enregistrés au cours de la prise de photos ne nécessitent aucune explication, nous nous arrêterons un instant aux deux derniers paramètres, à savoir la miniature de la photo et les données de géolocalisation. La miniature de la photo est enregistrée avec d'autres paramètres pendant la prise de la photo. L'utilisateur de l'appareil traite très souvent les photos par la suite dans les programmes graphiques spécialisés en modifiant les couleurs, en coupant un fragment, en supprimant les éléments inutiles, etc. Si le programme utilisé par la personne pour ce but ne supporte pas le format Exif, la photo sera le plus probablement modifiée mais ses métadonnées ne le seront pas. L'utilisateur obtiendra en résultat une photo modifiée contenant la miniature de l'image originale. Quelles peuvent en être les conséquences ? Prenons l'exemple de Catherine Schwartz, présentatrice américaine d'une chaîne télé TechTV. En juin 2003, elle a mis dans son blog plusieurs photos qui la représentent seule. Les photos elles mêmes n'étaient pas particulières, elles montraient la présentatrice en train de fumer une cigarette. La photo la montrait à partir des épaules. Les internautes ont découvert rapidement que les photos contiennent des métadonnées très intéressantes. La miniature Exif s'est avérée plus intéressante que la photo elle-même car la photo originale montrait la présentatrice à partir de hanches toute nue. Le programme utilisé pour le traitement de photos, Photoshop, n'a pas actualisé les métadonnées de la photo.

Cela semble surprenant au premier abord mais les appareils photo modernes (par exemple, Nikon Coolpix P6000) et les téléphones qui permettent

de prendre des photos (par exemple, iPhone) sont capables d'enregistrer automatiquement les coordonnées géographiques de l'endroit où la photo a été prise dans les fichiers graphiques. Vous connaissez probablement l'histoire d'un utilisateur d'iPhone (connu sur Internet comme Nephew chan) qui a présenté une photo de sa tente en bain sur un forum Internet public. Comme ces photos contenaient les coordonnées géographiques ajoutées authentiquement par le téléphone, un autre utilisateur du forum a retrouvé l'admirateur et a commencé à le faire chanter en demandant d'autres photos. Finalement, toute l'histoire a été mise au jour et tous les protagonistes ainsi que la communauté Internet ont appris les détails de cette histoire exceptionnelle.

Les exemples présentés démontrent clairement que partager ses propres photos sans réfléchir peut avoir des conséquences néfastes. Comment un cybercriminel peut collecter, analyser et utiliser à ses propres fins les métadonnées contenues dans les fichiers graphiques ?

La première étape consiste bien évidemment à trouver des photos susceptibles de contenir des métadonnées intéressantes. Trouver des données principales n'est pas le sujet clé de cet article, je mentionnerais donc seulement que pour trouver des photos appartenant à une organisation ou une personne particulière, il suffit de bien parcourir Internet. À l'ère de la révolution numérique, lorsque quasiment toutes les organisations et les particuliers ne peuvent pas se passer d'un site Web, d'un blog photo ou d'un e-commerce, Itrouver des photos numériques liées à une institution ou une personne qui nous intéresse consiste le plus souvent à former une requête appropriée dans l'un des moteurs de recherche. Une fois les fichiers intéressants trouvés, utiliser les métadonnées y contenues n'est qu'un jeu d'enfant.

Regardons comment les métadonnées collectées par l'intrus dans l'une des histoires racontées ont été utilisées. Afin de trouver une photo originale mises sur Internet par Nephew chan, il suffit de faire une requête dans le moteur de recherche Google. Nous retrouvons rapidement la photo utilisée par l'intrus, elle est accessible à l'adresse suivante : *http://images.encyclopediadramatica.com/images/0/01/Nephew-owned.jpg*. Ensuite, il suffit d'analyser les métadonnées Exif, contenues dans le fichier. La méthode la plus simple pour ce faire consiste à installer un complément appelé *Exif Viewer* dans le navigateur Firefox. Une fois le complément installé et le navigateur redémarré, il suffit d'afficher la photo en faisant un clic droit dessus, de sélectionner l'option *View Image Exif Data*. La fenêtre du complément *Exif Viewer* s'affichera. Elle contient une série d'informations Exif. L'image contient les données suivantes (liste réduite, en raison de lisibilité) :

- Camera Make = Apple,
- Camera Model = iPhone,
- GPS Latitude Reference = N,
- GPS Latitude = 38/1,3550/100,0/1 [degrees, minutes, seconds] ===› 38° 35.5° ,
- GPS Longitude Reference = W,
- GPS Longitude = 90/1,2657/100,0/1 [degrees, minutes, seconds] ===› 90° 26.57° .

Exif Viewer génère également le fichier KML (en anglais *Keyhole Markup Language*), permettant d'afficher immédiatement l'emplacement indiqué par les données GPS dans le programme Google Earth. Après avoir trouvé la photo qui nous intéresse, nous sommes capables d'afficher la carte satellite de l'endroit où elle a été prise ! De plus, grâce au service Google Street View, proposant les vues panoramiques de certaines parties du monde (aujourd'hui principalement les États-Unis), nous pouvons faire une excursion virtuelle autour de la maison (Figure 1), où la photo a été prise ! Est-ce la magie ? Non, il s'agit tout simplement de profiter des métadonnées partagées de manière irréfléchie.

Nous essayerons à présent trouver des métainformations intéressantes cachées dans les photos présentes sur la toile. En parcourant le site Web du Président de Pologne, j'ai décidé de vérifier si les métadonnées des photos y présentes sont préparées de manière professionnelle et si elles ne cachent pas d'informations supplémentaires. J'ai trouvé des photos simples en apparence à l'adresse *http://www.prezydent.pl/x.download?id=29526128*. Elles présentent Lech Kaczyński en réunion avec le pape actuel. L'analyse des métadonnées du fichier JPG, à l'aide du complément *Exif Viewer*, a révélé plusieurs détails intéressants. Avant tout, la vue de la photo originale (Figure 2) révèle que la Première Dame a été effacée de la photo placée sur le site. Est-ce que le Président a honte de son épouse et a ordonné de l'effacer d'une partie de photos publiées sur le site Web officiel ? D'autres métadonnées qui ne devraient pas se trouver forcement sur la version finale de la photo : *By-line = Jacek Turczyk* et *Originating Program = FotoWare FotoStation*. La photo nous donne les coordonnées personnelles du photographe qui l'a prise et le type du logiciel (spécialisé et cher) qu'il a utilisé. En analysant d'autres photos de ce site, nous serons capables de trouver d'autres métadonnées, une partie de photos ne contient en revanche aucune donnée Exif. Les personnes responsables du site *www.prezydent.pl* n'ont pas encore défini une politique correcte de gestion des métadonnées. J'ajoute encore qu'il est difficile de trouver une photo dépourvue de métadonnées intéressantes sur le site

Figure 2. *La première dame a été supprimée sur une partie de photos présentes sur le site Web du Président de Pologne*

Web *www.premier.gov.pl* (premier ministre polonais).

Publier les photos contenant des métadonnées peut faire apparaître des risques supplémentaires, invisibles au premier abord. Remarquons qu'en analysant les métadonnées, nous obtenons les informations directement ou indirectement sur le logiciel spécifique utilisé par les photographes ou les personnes qui modifient les photos. À titre d'exemple, si la photo est prise avec l'appareil iPhone, l'auteur possède le plus probablement le logiciel Itunes sur son ordinateur. Si la photo est prise avec l'appareil Canon EOS 400D, l'ordinateur de l'auteur est équipé le plus probablement d'un logiciel fourni par le fabricant. Si les données Exif montrent la dernière modification effectuée au moyen du paquet Adobe Photoshop, l'auteur dispose le plus probablement de ce logiciel. Les métadonnées contiennent souvent les informations qui précisent la version concrète du logiciel. Un cracker rusé peut utiliser ce type d'informations pour choisir un exploit correct (ciblé à un type concret de logiciel), ce qui lui permettra d'effectuer une attaque efficace sur les ordinateurs de l'auteur des photos. Ce type d'informations (type et version du logiciel) peut être utilisé aussi dans les attaques du type *spear phishing*. *Spear phishing* constitue une sorte de phishing ciblé (en anglais *spear* – lance). À titre d'exemple, si nous savons que l'auteur de la photo utilise un logiciel spécialisé *FotoWare FotoStation Pro*, l'intrus faisant semblant d'être un représentant de la société FotoWare peut envoyer un message spécialement préparé à sa victime. En se servant des symboles et des structures graphiques utilisées par la société *FotoWare* (ces informations se trouvent sur le site officiel du fabricant), l'intrus peut préparer un faux message envoyé par ce fabricant spécialement aux utilisateurs enregistrés du paquet *FotoStation Pro*. Ce message contiendrait un complément critique (qui en réalité est un cheval de Troie, un virus, etc.). du logiciel et qui recommanderait de l'installer immédiatement. Connaissant en plus les coordonnées de la personne attaquée (à titre d'exemple, grâce à l'entrée *By-line* = dans les métadonnées de la photo), il peut adresser ce message en utilisant le prénom et le nom découverts pour augmenter sa crédibilité. Le phishing de ce type, précisément ciblé, a plus de chance de réussir car il est plus crédible que les centaines de milliers de messages généraux adressés à tout le monde et non à une personne précise.

Pour terminer nos réflexions sur les métadonnées contenues dans les fichiers graphiques, il faut mentionner deux services qui peuvent aider les intrus à trouver les photos contenant des métadonnées intéressantes. Si un intrus trouve une photo publiée par une personne ou une institution qu'il recherche et cette photo est dépourvue de métadonnées, il existe une autre astuce. Pour trouver d'autres occurrences de la même photo ou d'une photo similaire sur Internet, le cracker peut se servir d'une recherche d'images inverse (en anglais *reverse image search*). Le service *tineye.com* propose de rechercher toutes les occurrences d'une photo donnée (indiquée par une adresse URL ou téléchargée depuis un disque local) sur Internet. L'intrus espère ainsi de trouver une autre occurrence de la même photo avec des métadonnées intéressantes. Le service *Wayback Machine* peut servir aux recherches similaires. Ce service permet de parcourir les versions d'archives de n'importe quel site Web. Même si une institution publie aujourd'hui des photos dépourvues de métadonnées, ces moyens de protection n'étaient pas en vigueur dans le passé et la version d'archive du site pourrait contenir des photos avec des métadonnées.

Comme vous pouvez voir sur les exemples démontrés, si vous publiez des photos sur Internet, réfléchissez ce que vous partagez en le faisant. En publiant une photo compromettante, n'oubliez pas que cacher un visage ou découper un fragment pourrait être insuffisant. Vérifiez donc si les métadonnées ne contiennent pas de miniature de la photo originale !

Non seulement les photos

Générer les métadonnées importantes du point de vue de la sécurité d'informations n'est pas uniquement le domaine de fichiers graphiques. Les métainformations sont intégrées dans de nombreux formats différents de fichiers. Il faut faire particulièrement attention aux fichiers au format PDF et Microsoft Office car ils sont les plus populaires.

L'histoire de Dennis Rader peut nous apprendre que les métadonnées contenues dans un fichier DOC sont extrêmement importantes. Ce meurtrier en série, qui avait tué 10 personnes aux États-Unis dans les années 1974 – 1991, était particulièrement connu de sa cruauté et de la passion pour correspondre avec la police et les médias. En 2005, il a décidé d'envoyer un message sous forme électronique et l'a fait sur une disquette. Les policiers ont analysé les métadonnées contenues dans le fichier DOC, ont trouvé

```
<</Subtype/XML/Length 3980/Type/Metadata>>stream
<?xpacket begin="d»¿" id="W5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe XMP Core 4.0-c316 44.253921, Sun Oct 01 2006 17:14:39">
   <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
      <rdf:Description rdf:about=""
            xmlns:pdf="http://ns.adobe.com/pdf/1.3/">
         <pdf:Producer>Acrobat Distiller 8.1.0 (Windows)</pdf:Producer>
      </rdf:Description>
      <rdf:Description rdf:about=""
            xmlns:pdfx="http://ns.adobe.com/pdfx/1.3/">
         <pdfx:SourceModified>D:20080413160751</pdfx:SourceModified>
      </rdf:Description>
      <rdf:Description rdf:about=""
            xmlns:xap="http://ns.adobe.com/xap/1.0/">
         <xap:CreatorTool>Acrobat PDFMaker 8.1 for Word</xap:CreatorTool>
         <xap:ModifyDate>2008-04-13T12:09:37-04:00</xap:ModifyDate>
         <xap:CreateDate>2008-04-13T12:08:34-04:00</xap:CreateDate>
         <xap:MetadataDate>2008-04-13T12:09:37-04:00</xap:MetadataDate>
      </rdf:Description>
      <rdf:Description rdf:about=""
            xmlns:xapMM="http://ns.adobe.com/xap/1.0/mm/">
         <xapMM:DocumentID>uuid:206e98f1-b7ff-431f-8dcf-bfe81531fc82</xapMM:DocumentID>
         <xapMM:InstanceID>uuid:f8cc4281-2136-4545-82b0-35635c2d7ad4</xapMM:InstanceID>
         <xapMM:subject>
            <rdf:Seq>
               <rdf:li>2</rdf:li>
            </rdf:Seq>
         </xapMM:subject>
      </rdf:Description>
      <rdf:Description rdf:about=""
            xmlns:dc="http://purl.org/dc/elements/1.1/">
         <dc:format>application/pdf</dc:format>
         <dc:creator>
            <rdf:Seq>
               <rdf:li>Ruhnka, Bagby</rdf:li>
            </rdf:Seq>
```

Figure *3. Métadonnées texte contenus dans le fichier PDF*

le nom de l'église à laquelle Rader était lié (*Christ Lutheran Church*) et les données de l'utilisateur qui avait modifié le fichier en dernier (Dennis). Ces informations ont bien évidemment suffit pour trouver et arrêter le meurtrier en série.

L'histoire moins sobre de David L. Shith est également connue. David, concepteur du virus connu Melissa, a été retrouvé grâce aux données GUID (en anglais *Globally Unique Identifier*) contenues dans plusieurs fichiers DOC. Ces informations ont permis d'arrêter Shith et de le condamner à 20 mois de prison.

La plus simple méthode pour trouver des métadonnées dans n'importe quel fichier, même celui dont le format est inconnu, consiste à l'ouvrir dans un éditeur de texte. En général, à côté des suites de caractères complètement illisibles se trouve une série de lignes au format XML contenant des entrées lisibles et faciles à interpréter. À titre d'exemple, si nous ouvrons un fichier PDF (Figure 3) dans l'éditeur de texte Notepad++, nous trouverons plusieurs informations intéressantes. L'entrée *‹xap:CreatorTool›Acrobat PDFMaker 8.1 for Word‹/xap:CreatorTool›* suggère clairement que l'auteur a utilisé le programme *PDFMaker 8.1* pour générer le fichier. Dans le cas de fichiers PDF, nous n'avons pas besoin de regarder le fichier texte, ce qui n'est pas très agréable. Il suffit d'ouvrir le fichier susmentionné dans le navigateur *Adobe Reader* et de sélectionner ensuite l'option *propriétés* dans le menu *fichier*. Vous verrez alors s'afficher une fenêtre (Figure 4) présentant plusieurs métainformations contenues dans le fichier, notamment :

- auteur : "Ruhnka, Bagby",
- application : Acrobat PDFMaker 8.1 for Word,
- concepteur PDF : Acrobat Distiller 8.1.0 (Windows),
- version PDF : 1.6 (Acrobat 7.x).

Ce type de données (type et version du logiciel utilisé par l'auteur) peut servir à l'intrus pour choisir un exploit efficace dans les attaques du type *spear phishing* dont nous avons parlé auparavant. Remarquons aussi que la ligne *Auteur : "Ruhnka, Bagby"* peut contenir des logins potentiels (car ce sont des noms que l'auteur utilise lors du travail sur l'ordinateur) que le cracker pourra ensuite tester pour obtenir un accès aux services utilisés par l'auteur en prenant son identité. Dans ce cas-là, servir *Ruhnka* et *Bagby* sont des noms de deux auteurs du document.

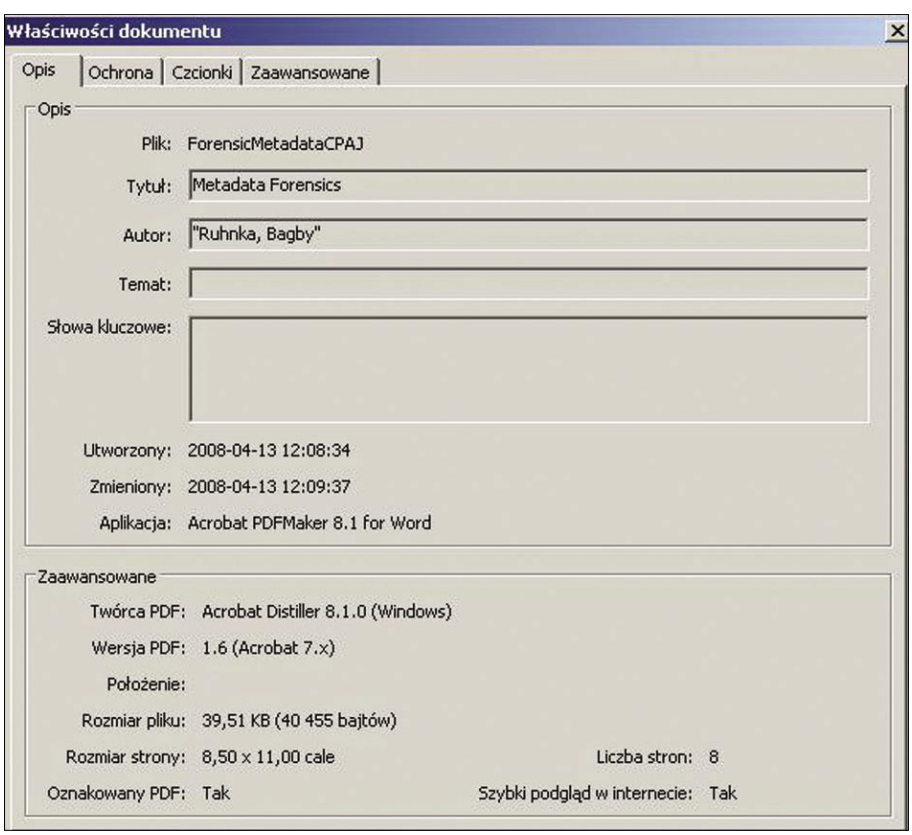
Les informations similaires peuvent être trouvées suite à l'analyse des métadonnées contenues dans les fichiers DOC (et autres types de fichiers générés par le paquet Microsoft Office). De même que les fichiers PDF, il est possible d'ouvrir les fichiers générés par MS Word dans un éditeur texte et de trouver les métadonnées contenues dans le fichier. Ce n'est pas très pratique. Il est possible d'obtenir ces données sous une forme plus claire en affichant les propriétés du fichier DOC sous Windows et en passant à l'onglet *Conclusion*. Dans le fichier test que j'ai téléchargé depuis le site *www.abw.gov.pl* (site de l'Agence de la sécurité interne) j'ai trouvé des métadonnées suivantes :

- auteur m_wilczek,
- enregistré la dernière fois par war009262.

Est-ce que les données de ce type peuvent constituer un risque ? Cela ne peut pas être exclu. Elles peuvent aider en quelque sorte à déterminer les données personnelles des auteurs de ce document ou à déterminer les logins d'utilisateur qu'ils utilisent. Cela prouve également que l'Agence de la sécurité interne ne fait pas particulièrement attention au problème de sécurité des métadonnées.

Microsoft lui-même avertit sur son site que les métadonnées contenues dans les fichiers Office peuvent révéler les informations suivantes :

- prénom, nom, initiaux,
- nom de l'entreprise ou de l'organisation,
- nom de l'ordinateur,
- nom du serveur de fichiers ou du disque où le document a été enregistré,
- données relatives aux objets OLE utilisés dans le document,
- données personnelles des personnes qui ont édité auparavant le fichier,
- données relatives à la version du document,



- informations concernant le modèle du document utilisé,
- commentaires.

Nous pouvons donc constater qu'un particulier ou une institution qui publie sur Internet ses propres fichiers MS Office risque beaucoup. Un petit pour cent de ces fichiers est publié par les auteurs qui se rendent compte que ces fichiers peuvent fournir de nombreuses informations aux curieux.

Pour terminer, je voudrais présenter un logiciel conçu spécialement pour la collecte et le traitement de nombreuses métadonnées, appartenant à une organisation concrète. Il s'agit de MetaGoofil. C'est un vrai outil complexe permettant de collecter des métadonnées depuis toute sorte de documents publiés sur les sites Internet d'une organisation indiquée. Le fonctionnement de l'application est assez simple. MetaGoofil recherche (dans le moteur de recherche Google) les fichiers contenant des métainformations (par exemple, site:*domena.com* filetype: pdf) dans le domaine et les types de fichiers donnés. Ensuite, les fichiers trouvés sont téléchargés sur le disque local et les métadonnées y présentes sont collectées et filtrées à l'aide de la bibliothèque *libextractor* (pour tester les énormes fonctionnalités de cette bibliothèque, rendez-vous sur le site Web *http://gnunet.org/libextractor/demo.php3?xlang=English*). Les résultats globalisés sont enregistrés sur le disque sous forme du fichier HTML. Afin de bénéficier des fonctionnalités du programme MetaGoofil, il suffit de le lancer avec les paramètres suivants : *./metagoofil.py -d domena.com -f all -l 100 -o domena.html -t temp*. Voici la signification de tous les paramètres d'appel :

- *-d domena.com*: nom du domaine à analyser,
- *-f all*: type de formats de fichiers pris en compte (all signifie l'analyse de tous les types de fichiers supportés),
- *-l 100*: limite du nombre des résultats traités,
- *-o domena.html*: nom du fichier HTML résultat,
- *-t temp*: répertoire contenant les fichiers téléchargés pour une analyse.

Sur le Net

- *http://www.cert.org/* – Computer Emergency Response Team,
- *http://www.remote-exploit.org/backtrack.html* – BackTrack,
- *http://exif.org/specifications.html* – spécification Exif,
- *http://en.wikipedia.org/wiki/Catherine_Schwartz* – Exif thumbnail story,
- *http://encyclopediadramatica.com/User:Darkanaku/Nephew_chan* – Nephew chan story,
- *http://www.microsoft.com/poland/athome/security/email/spear_phishing.mspx* – Qu'est-ce une attaque spear phishing ?,
- *http://en.wikipedia.org/wiki/Dennis_Rader* – Dennis Rader story,
- *http://en.wikipedia.org/wiki/Melissa_virus* – David L. Smith story,
- *http://office.microsoft.com/en-us/help/HA010776461033.aspx* – Metadata in MS Office,
- *http://www.edge-security.com/metagoofil.php* – Metadata analyzer, information gathering tool,
- *http://gnunet.org/libextractor/demo.php3?xlang=English* – libExtractor – Online Demo,
- *http://www.irongeek.com/* – Irongeek.

MetaGoofil est capable de collecter les informations intéressantes comme :

- logins potentiels des utilisateurs utilisés dans l'organisation donnée,
- chemins aux ressources fichiers (cela permet de reconnaître les systèmes d'exploitation utilisés, les noms de réseau et les noms des montages partagés) dans lesquels les fichiers analysés étaient édités,
- adresses MAC (d'après les identifiants GUID des fichiers Office où se trouve l'adresse physique de l'hôte courant) des ordinateurs sur lesquels les fichiers ont été édités.

Je ne répéterai pas comment les informations de ce type peuvent être utilisées. Comme vous pouvez le constater, les fonctionnalités du script (MetaGoofil est en réalité un script écrit en langage Python) sont énormes. L'intrus à la recherche d'informations sur une organisation donnée n'a plus à chercher et à analyser individuellement chaque fichier publié sur le site. Le programme n'omettra aucun détail et affichera toutes les informations obtenues sous forme d'une page HTML claire. L'intrus n'a qu'à utiliser ces informations pour planifier une attaque efficace.

Conclusion

Les protections adéquate de métadonnées peuvent constituer un sujet d'un article à part. Une analyse générale de la question démontre que cela ne pose aucun problème. Des applications gratuites existent qui permettent de supprimer les données Exif des fichiers JPG. Microsoft propose des compléments gratuits pour le paquet Office, permettant de supprimer les métadonnées dans les fichiers créés aux formats les plus populaires. Internet propose également de nombreux programmes indépendants, capables de gérer plusieurs formats de métadonnées. Enfin, il est possible d'éviter de nombreuses fuites d'informations tout simplement en ne partageant pas les fichiers DOC sur le réseau car ils ne s'y prêtent pas. MS Word est un programme conçu pour éditer les fichiers et ils devraient être utilisés à ces fins. Ce n'est sûrement pas un format conçu pour les publications, en particulier sur la Toile !

Pourquoi donc les particuliers et les institutions sérieuses (bureau du président de Pologne, Agence de sécurité interne) partagent des milliers de métainformations sur leurs sites ? Dans le cas des particuliers, nous pouvons expliquer cette situation par manque de connaissances et inconscience de risques. Mais comment expliquer les opérations de l'Agence de sécurité interne ?

Toutes les institutions doivent réfléchir sérieusement à la mise en place d'une politique cohérente de gestion des métadonnées. Les métadonnées sont peut-être critiques dans de nombreux cas et il faut donc les prendre en considération dans les procédures intégrées de sécurité informatique.

À propos de l'auteur
L'auteur est diplômé de la faculté d'Automatique, d'Electronique et d'Informatique de l'Ecole Polytechnique de Silésie à Gliwice. Il se spécialise dans les bases de données, les réseaux et les systèmes informatiques. Il travaille comme administrateur réseau et systèmes informatiques dans la société Mostostal Zabrze Holding S.A. Contact avec l'auteur : *wojciech.smol@mz.pl*.