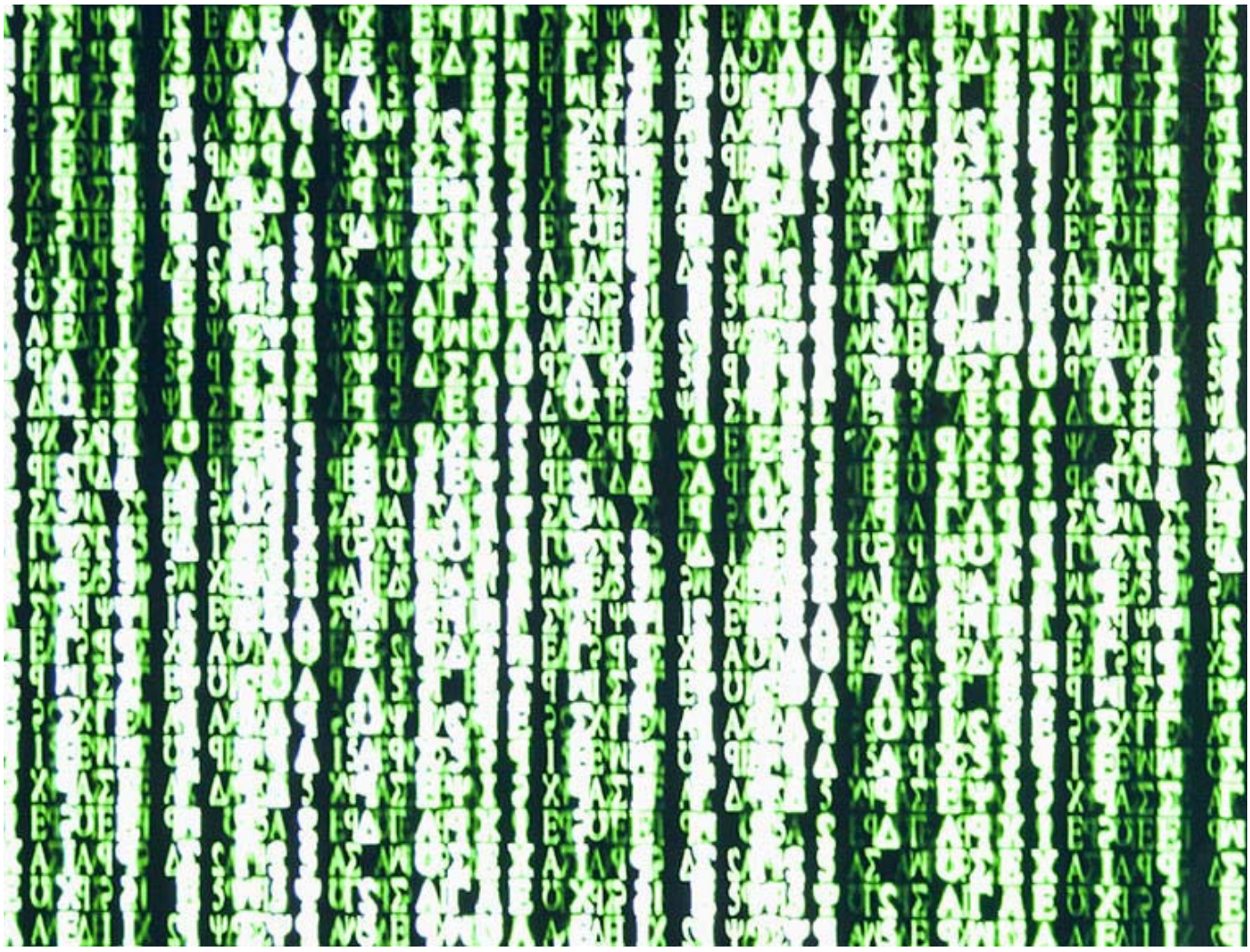
CNRS LE JOURNAL

Publié sur *CNRS Le journal* (https://lejournal.cnrs.fr)

Numérique
Informatique

Virus et malwares : les chercheurs contre-attaquent

07.03.2016, par Charline Zeitoun



© DEUX/CORBIS

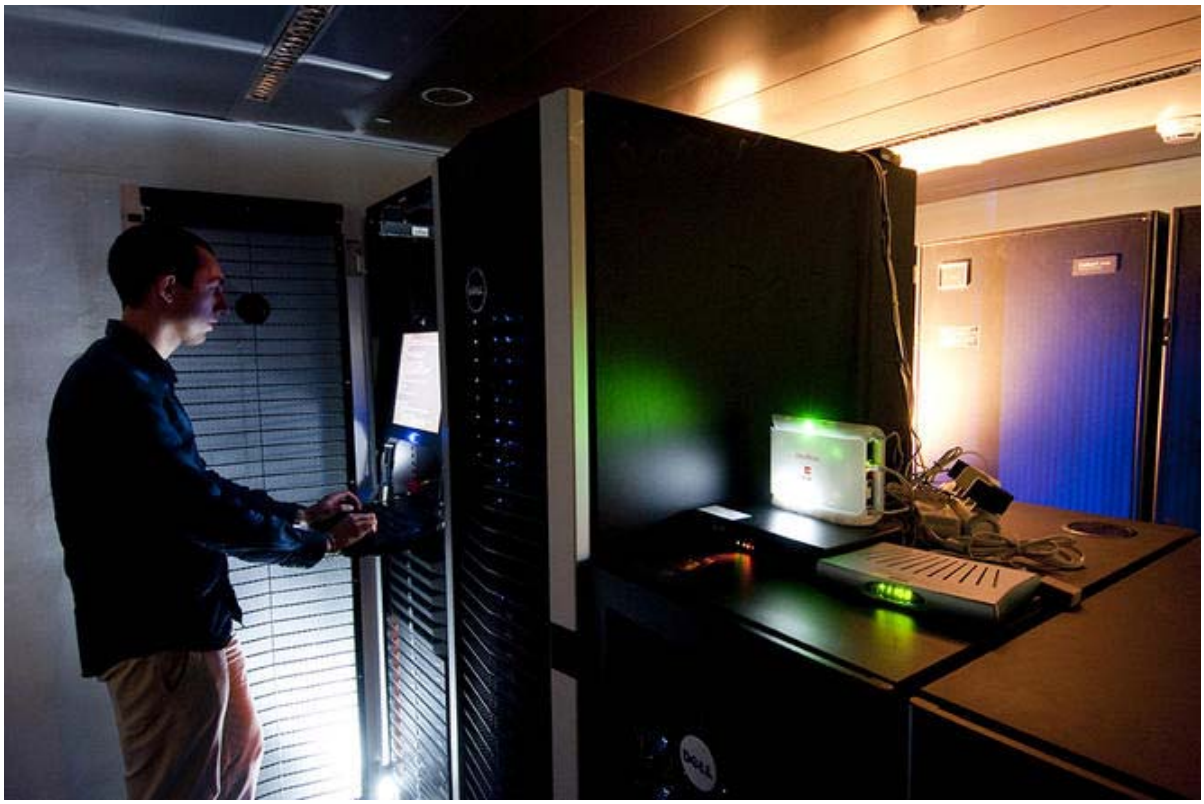
Grâce à leur collection de 6 millions de malwares, les chercheurs du Laboratoire de haute sécurité ont mis au point un anti-virus d'un nouveau genre, déjà utile à la gendarmerie et bientôt disponible pour les entreprises. Visite de la première plateforme de recherche française dédiée à la sécurité informatique.

Portes blindées, sas, caméra de surveillance et reconnaissance biométrique de l'œil : le Laboratoire de haute sécurité (LHS) du Loria[1], à Nancy, est une forteresse où sont confinés six millions de virus informatiques *[Programme capable d'infecter un autre logiciel d'ordinateur en le modifiant de façon à ce qu'il puisse à son tour se reproduire. Non malveillants à l'origine, les virus sont aujourd'hui souvent porteurs de code malveillant et peuvent perturber plus ou moins gravement le fonctionnement de l'ordinateur infecté. Ils se répandent via les réseaux informatiques ou bien via des dispositifs périphériques comme les clés USB, etc.]*. Une collection des pires « méchants » de la planète Web

attrapés sur la toile par les chercheurs du LHS. Ces malwares *[Logiciel clairement malveillant, développé dans le but de nuire à un système informatique, sans le consentement de l'utilisateur dont l'ordinateur est infecté.]* piratent nos données, détruisent nos logiciels ou nos disques durs, voire forcent nos ordinateurs à déverser des torrents de spams *[Courrier électronique indésirable. Il s'agit en général d'envois en grande quantité effectués à des fins publicitaires.]* pour paralyser les serveurs d'un concurrent ou un site jugé ennemi. Intérêt de collectionner ces super-vilains ? Pouvoir les analyser en profondeur. Puis concevoir des outils qui permettent d'en détecter les « mutants », des variantes issues de la même souche mais légèrement modifiées, et qui n'ont pas encore fait suffisamment de dégâts pour être répertoriés et intégrés aux anti-virus du commerce. *« Ces logiciels, eux, ne détectent en général que les virus qu'ils connaissent déjà*, commente Jean-Yves Marion, directeur du LHS. *C'est pourquoi leurs concepteurs suivent attentivement les résultats de la recherche pour améliorer leurs programmes. »* D'autant qu'il y a aujourd'hui beaucoup plus d'attaques qu'il y a dix ou vingt ans. Et que le temps des geeks des années 1980, qui craquaient les systèmes pour la beauté du geste, est bien loin…

Provoquer les attaques en exhibant des ordinateurs vulnérables

« L'amateurisme n'est plus de mise, reprend le chercheur. *La plupart des attaques ont des visées lucratives ou d'espionnage et sont le fait de groupes criminels ou d'organisations gouvernementales qui mettent au point des virus dont la conception demande des mois de travail. »* En novembre 2014, l'éditeur d'antivirus Symantec révélait ainsi l'existence de Regin. Depuis au moins six ans, ce malware subtilisait des mots de passe et réalisait des captures d'écran dans le réseau informatique du siège de l'Union européenne, mais aussi dans des centres de recherche, des compagnies aériennes et des réseaux de communication de plusieurs pays d'Europe, ainsi qu'en Russie, en Arabie Saoudite, au Mexique, etc., via ordinateurs et smartphone GSM. Regin est si hors norme, si complexe (il aurait nécessité un an de travail à quatre personnes à temps plein) que, selon les experts, il ne pouvait provenir que d'une organisation gouvernementale[2].



Les chercheurs du LHS utilisent un télescope virtuel, connecté à Internet via des lignes ADSL classiques, afin de simuler la présence d'ordinateurs vulnérables et de susciter des attaques qu'ils peuvent ensuite capturer.

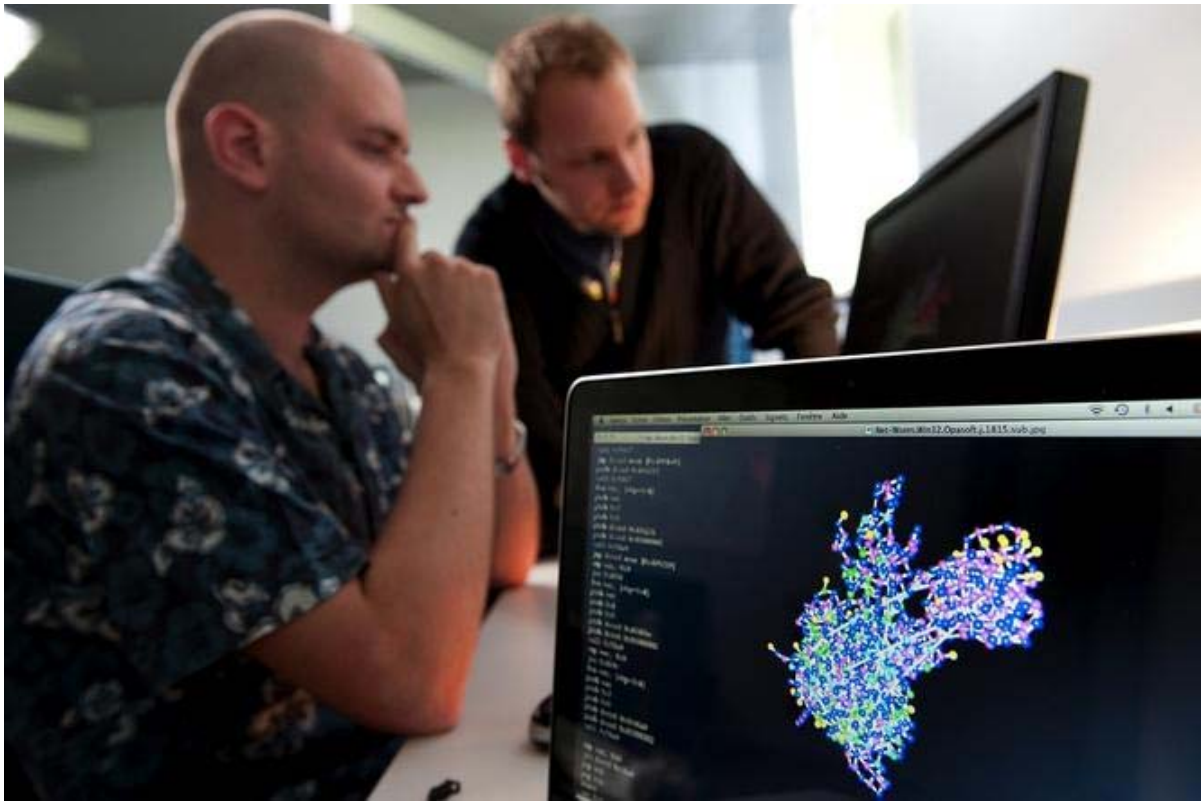
© KAKSONEN/INRIA

La plupart des attaques sont le fait de groupes criminels ou d'organisations gouvernementales.

Dans ce contexte de cyber-espionnage et de cyber-criminalité, l'attrape-virus du LHS « ramasse » tout ce qui traîne sur le Web afin d'enrichir sa collection. *« C'est un télescope virtuel, développé par l'équipe Madynes* [3], explique Jean-Yves Marion. *Connecté à Internet via des lignes ADSL classiques, cet outil permet de simuler la présence de centaines d'ordinateurs vulnérables, de façon à susciter le maximum d'attaques que l'on va ensuite capturer : c'est la technique du "pot de miel" pour attirer l'ours et l'attraper une fois qu'il y a mis la patte »*, explique le chercheur.

Mais comment exhiber en ligne de faux ordinateurs imprudents ? Cela revient grosso modo à envoyer des messages de signalisation. Quand, au carrefour de ses nœuds, le protocole d'Internet demande périodiquement aux machines connectées « toi en face, qui es-tu ? », le télescope virtuel produit ainsi de fausses réponses du type : *« Je suis un mac et je me promène sans anti-virus »* ou bien *« Je suis un PC et je fonctionne avec tel système d'exploitation »*, système d'exploitation qu'on aura bien sûr pris le soin de choisir dans sa version la moins aboutie et la plus riche de failles…

Une fois capturés, les malwares passent à la moulinette du logiciel Gorille, l'arme secrète du laboratoire lorrain. *« Comme avec les anti-virus du commerce, la méthode consiste à chercher une signature qui est propre au malware et qui permettra de l'identifier »*, explique Jean-Yves Marion. Mais la grande différence porte sur cette fameuse signature : *« On ne se restreint pas à la traque de quelques lignes de code ou de chaînes de caractères typiques d'un code malveillant déjà identifié,* poursuit l'informaticien. *Pour nous, la signature d'un malware, c'est sa structure complète. »* Bref, un portrait-robot de la « silhouette » générale du malfaiteur, au lieu d'un indice du type *« il portait une montre "Bollex" »*. Résultat : on peut retrouver le suspect s'il change de montre, ou plutôt si les concepteurs du malware l'ont légèrement modifié pour en faire un mutant.



Sur l'écran : graphe 3D d'un malware. Portrait-robot de la « silhouette » générale du malware, ce graphe permet d'identifier ce dernier même après légère modification.

© KAKSONEN/INRIA

Un nouvel anti-virus adapté aux entreprises

« Pour extraire la structure d'un programme malveillant, il faut regarder la liste des instructions qui le composent, en code assembleur (c'est le langage de la machine) », explique Fabrice Sabatier, de l'équipe Carbone du LHS. En fonction de la nature de ces instructions (effectuer un calcul, répéter telle action, demander à l'utilisateur de rentrer une donnée, etc.), celles-ci sont représentées par une forme géométrique. Ces formes géométriques sont ensuite reliées par des nœuds qui symbolisent les « sauts conditionnels » : ce sont les fameux *« if-then-else »* qui donnent l'ordre d'exécuter telle action ou bien telle autre en fonction d'une condition. Ces représentations sont assez classiques en informatique. *« Mais un programme est constitué de millions de nœuds ! »*, reprend le chercheur. Tout l'art de la méthode réside ensuite dans des règles de simplification, en supprimant tel nœud et non tel autre, ou tel paquet d'instructions jugé peu caractéristique, afin d'obtenir un dessin ou « graphe » de quelques milliers à quelques centaines de milliers de nœuds seulement ***(voir la vidéo plus bas)***.

Nous avons aidé la gendarmerie à identifier les souches de différentes attaques virales, de type ransomware, qui venaient de la même source.

« C'est ainsi que, grâce à nos graphes, visualisables en 3D et en couleur, nous pouvons comparer n'importe quel programme via sa structure globale, ou certains de ses "morceaux", avec les échantillons de notre collection. Et, s'il y a de gros points communs, la présomption d'avoir

affaire à un malware sera d'autant plus forte », résume Jean-Yves Marion. *« Notre méthode est trop complexe pour tourner sur les PC du grand public, mais nous avons aidé la gendarmerie à identifier les souches de différentes attaques virales de type Ransomware [Logiciel malveillant qui prend en otage des données personnelles en les chiffrant, ou bloque l'accès d'un utilisateur à une machine. Il demande ensuite à son propriétaire d'envoyer de l'argent (une rançon) en échange de la clé qui permettra de déchiffrer les données ou de débrider l'accès de utilisateur.] qui venaient de la même source »*, commente Fabrice Sabatier. La technique devrait aussi bientôt servir aux entreprises, le Loria monte une start-up pour cela[4]. L'intérêt pour elles est particulièrement grand car la méthode, fondée sur une analyse en profondeur des malwares, permet également de remonter jusqu'aux buts de ces derniers : en étudiant les différentes fonctionnalités de ces logiciels, on peut découvrir ce qu'ils voulaient voler ou détruire.

Graphe 3D d'un malware

01:14

CNRS

Graphe 3D d'un malware [11] par CNRS [12]
Dans cette vidéo : un graphe 3D est élaboré pas à pas à partir du code d'un malware.

Identifier les différentes fonctionnalités, noyées dans les milliers de lignes de code, est d'ailleurs un thème-clé de la recherche fondamentale en informatique. Parce qu'il n'existe pas d'algorithme capable d'affirmer à coup sûr que deux programmes font la même chose (c'est la question de l'indécidabilité, démonstration mathématique que l'on doit à Alan Turing [13]). *« Par ailleurs,* poursuit Jean-Yves Marion, *analyser les virus nous amène à des questions tout aussi fondamentales du type : en fin de compte, qu'est-ce qu'un programme malveillant ? »* Comment le distinguer d'un programme ordinaire qui agirait de façon pas toujours justifiable ? Qu'est-ce qui le différentie de votre application favorite de jeu en ligne, si cette dernière se met à vous géolocaliser ? Ou à transmettre vos données à un tiers, comme en a justement été accusé un fameux jeu d'oiseaux il y a deux ans[5] ? Côté chercheurs, il faut au final parvenir à définir dans quels contextes certaines fonctionnalités sont suspectes et dans quels autres elles ne le sont pas…

Au cœur du code des malwares mutants

Et pour l'avenir, faut-il craindre de nouvelles générations de virus, de nature et de structure révolutionnaires ? *« Les nouveautés en virologie informatique tiennent surtout dans la façon de protéger les malwares en les "déguisant" »*, répond Jean-Yves Marion. Il y a les mutants, évoqués plus hauts, mais aussi le fait de crypter le code du virus ou encore de le ziper *[Action de compression de données ou de codage de source. Cette opération consiste à transformer une suite de bits en une suite de bits plus courte pouvant restituer les mêmes informations, et ce grâce à un algorithme particulier.]* et re-ziper, jusqu'à des centaines de fois, dans un autre programme, afin de le cacher et de le soustraire à l'analyse de l'anti-virus. Question dissimulation, il y a pire encore : certains malwares s'auto-modifient au fur et à mesure de leur exécution sur un PC, avec des fonctionnalités cachées qui se déclenchent par vagues, qui peuvent s'effacer en cours de route ou ne jamais s'activer !

Les nouveautés en virologie informatique tiennent surtout dans la façon de protéger les malwares en les déguisant.

« Depuis cinq ou six ans, il y a une véritable ingénierie de la protection des virus informatiques. Nous avions fait un test avec un de nos

échantillons : nous l'avons protégé avec un logiciel du commerce (utilisé dans le cadre de la protection des fichiers pour les droits d'auteur), et il est passé à travers les mailles du filet de nombreux anti-virus qui connaissaient pourtant l'original », explique Jean-Yves Marion. Tandis qu'au LSH, on peut exécuter un virus : en l'exécutant, on le dé-zipe autant de fois que nécessaire et on observe ses vagues d'auto-modification, ce qui permet d'accéder à ses parties cachées, au plus profond de son code. *« Nous pouvons réaliser des expériences en toute sécurité sur notre réseau, sans craindre la contagion vers d'autres machines, car c'est un cluster confiné, déconnecté du monde extérieur, qui permet de simuler des centaines de machines virtuelles »*, souligne l'informaticien.

Garantir la sécurité numérique des citoyens

Bien sûr, Google et les autres Gafa *[Acronyme formé avec les initiales de Google, Apple, Facebook et Amazon, géants du Net.]* veillent aussi au grain. Le moteur de recherche américain possède d'ailleurs une collection de malwares encore plus impressionnante que celle du LHS avec ses 300 ou 400 millions d'échantillons, même si les doublons, mutants et autres, y foisonnent et gonflent les effectifs. *« En marge de ces acteurs privés, la recherche publique joue un rôle capital, notamment dans le domaine de la virologie, trop peu développé en France et en Europe »*, insiste le directeur du LHS, première plateforme de recherche académique française dédiée à la sécurité informatique.

« Parallèlement aux actions mises en place par des sociétés guidées par des intérêts commerciaux, la recherche publique doit proposer des solutions pour garantir la sécurité numérique des citoyens, chercher les failles des outils du commerce et en informer le grand public qui est ensuite libre de les utiliser ou non », souligne le chercheur. Ce processus salutaire pousse aussi parfois les concepteurs à proposer rapidement des patchs correcteurs. Un exemple édifiant a fait trembler la toile en juin dernier : une démonstration sur Logjam, réalisée par des chercheurs du Loria, avait ainsi mis en évidence une faille majeure dans le protocole https [14] qui sécurise les connexions Internet. Sans compter que le monde tout connecté que nous nous préparons offrira de plus en plus de cibles aux hackers. *« Avant d'être développés à grande échelle, le pacemaker branché sur Wi-Fi, la voiture autonome, le bracelet qui prend votre pouls ou le vote électronique [15] réclameront d'apporter un certain nombre de garanties »*, avertit Jean-Yves Marion. À qui choisirons-nous d'en confier la responsabilité ?

À lire aussi :
« Logjam, la faille qui met Internet à nu » [14]
« Le vote électronique, pour quelles élections ? » [15]

Notes

1. Laboratoire lorrain de recherche en informatique et ses applications (CNRS/Inria/Univ. de Lorraine).
2. The Intercept, magazine anglais d'investigation, pointe les États-Unis et l'Angleterre du doigt : https://theintercept.com/2014/11/24/secret-regin-malware-belgacom-nsa-gchq/ [16] et https://theintercept.com/2014/12/13/belgacom-hack-gchq-inside-story/ [17]
3. Managing dynamic networks and services/Supervision des réseaux & services dynamiques, http://www.loria.fr/la-recherche/equipes/madynes [18]
4. Cette start-up, Simorfo, est créée en collaboration avec la société Tracip.
5. Selon le *New York Times* du 27 janvier 2014, le jeu « Angry Birds » aurait servi à la NSA, agence américaine de renseignement, et à son homologue britannique, le GCHQ, pour collecter certaines données de ses utilisateurs.