Exonet Techniques de recueil de traces et de données personnelles

Description du thème

Propriétés	Description
Intitulé long	Techniques de recueil de traces et de données personnelles
Formation concernée	Terminale STMG Système d'information de gestion (SIG)
Matière	Système d'information de gestion (SIG)
Présentation	L'exercice permet de découvrir quelques unes des traces qu'une navigation sur internet peut laisser. D'abord en se plaçant du point de vue de l'internaute (navigateur, cookies, données conservées par le serveur, etc.) puis en se plaçant du point de vue opposé (celui du serveur web : journaux, analyse, etc.). Le thème de l'usurpation d'identité est également abordé par l'analyse d'une tentative de hameçonnage (*phishing*) par courrier électronique.
Notions	**Thème** « Communiquer pour collaborer » **Question de gestion** « En quoi les systèmes d'information transforment-ils les échanges entre les acteurs des organisations ? » **Notion** « Traces numériques » **Finalité** « Repérer les techniques de recueil de traces et de données personnelles et les possibilités de leur exploitation bienveillante ou non »
Outils	Navigateur HTTP, ligne de commandes (commande nslookup)
Mots-clés	Traces, Cookies, Géolocalisation, Hameçonnage, Données personnelles
Auteur(es)	Olivier Korn avec les relectures attentives d'Alexandra Davant et de Gaëlle Castel
Version	v 1.0
Date de publication	Juin 2014

PREMIÈRE PARTIE : Introduction à la notion de traces numériques

a) Découverte de la CNIL

Afin de pouvoir répondre précisément aux interrogations qui suivent, vous devrez d'abord prendre le temps de suivre une démonstration sur le site de la **Commission nationale de l'informatique et des libertés (**Cnil : http://www.cnil.fr/).

Travail à faire :

Trouvez un lien sur la page d'accueil permettant d'en savoir plus sur la Cnil (« à propos » ou « qui sommes-nous »)… Quel est le texte du lien ? Quelle est l'adresse vers laquelle il mène ?	

La Cnil semble-t-elle être un organisme digne de confiance ?	
Comment pouvez-vous l'affirmer et comment s'en assurer ? La page repérée plus haut suffit-elle à l'affirmer ?	

b) Démarrage de l'expérience

La démonstration se trouve dans la rubrique « Vos droits » et plus précisément dans la sous-rubrique « Vos traces », dont l'adresse est http://www.cnil.fr/vos-droits/vos-traces/. À chaque étape, vous devrez effectuer quelques recherches pour répondre aux questions qui suivent.

Démarrez l'expérience (http://www.cnil.fr/vos-droits/vos-traces/experience/). Remarquez qu'en plus du cadre central qui présente les différentes étapes de l'expérience au fur et à mesure de son déroulement, un ou plusieurs cadres situés à droite apporte(nt) des précisions supplémentaires si nécessaire (par exemple pour répondre aux questions qui suivent).

Travail à faire, étape « Votre ordinateur » :

Déterminez l'adresse IP de votre poste de travail.	
Cette adresse est-elle identique à celle identifiée par le site de la Cnil ? Pourquoi ?	
Le nom d'hôte a été obtenu par une requête DNS inverse. Ce nom d'hôte permet-il d'apprendre quelque chose sur vous et/ou sur votre localisation ?	
Indiquez comment votre localisation peut, dans certains cas, être déterminée à partir de votre adresse IP.	

Travail à faire, étape « Votre navigateur » :

Les informations recueillies sur votre système sont-elles exactes ?	
Ces informations ont été recueillies par le biais d'un logiciel écrit en JavaScript. En observant la nature des informations collectées, dites si le logiciel JavaScript s'est exécuté sur votre poste de travail (logiciel local) ou sur le serveur de la Cnil (logiciel distant). Justifiez votre réponse.	

Travail à faire, étape « Les cookies » :

Remplissez le formulaire et passez à l'étape suivante mais **ne supprimez pas le *cookie* tout de suite**.

Déterminez un moyen d'accéder aux cookies stockés par votre navigateur. Notez-le ici. Attention, selon le navigateur il faudra passer par l'exploration du disque dur ou par un outil intégré (par exemple celui servant à supprimer des *cookies*).	

Vérifiez le contenu du *cookie*. Comment les paramètres saisis sont-ils stockés ?	
À quoi correspond le chemin ?	
Y a-t-il une date d'expiration ?	

Poursuivez ensuite l'étape « Les cookies » jusqu'à son terme.

Travail à faire, étape « Les recherches » :

Effectuez des recherches en suivant les instructions.

Quelles sont les durées de conservation maximales recommandées par les autorités européennes aux moteurs de recherche ?	
Avez-vous déjà constaté par vous-même que certains sites vous proposaient des publicités ciblées ? Sur quels sites ? Quel type de produit ?	

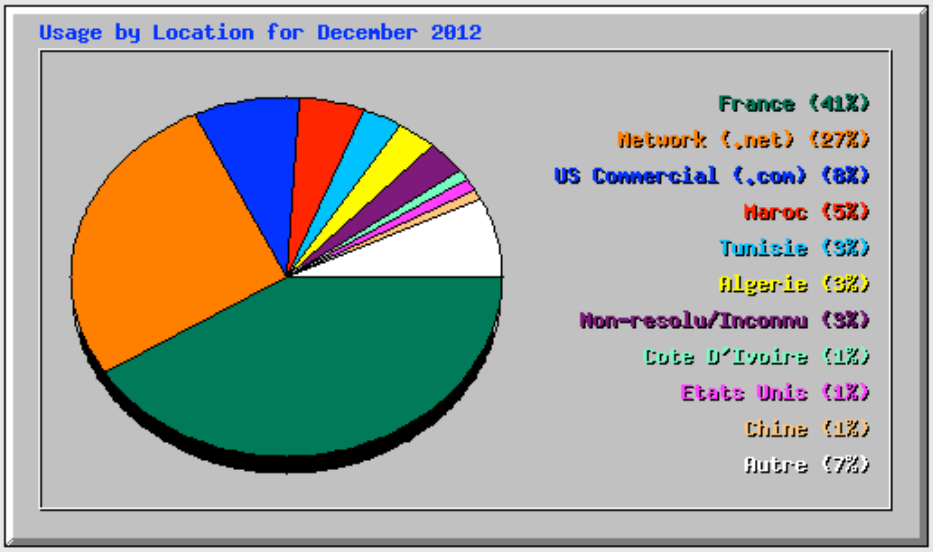
c) Suite de l'expérience

Poursuivez l'expérience avec « Les échanges sur internet » puis « Les journaux de connexion », dans l'espace « Vos droits / Vos traces » du site de la Cnil.

DEUXIÈME PARTIE : Mise en application

a) Exploitation de journaux de connexion

L'administrateur du réseau Certa vous confie l'illustration suivante, extraite des statistiques de visites du site *web* dont il a la charge.



Top 100 sur un total de 131 Total Locations

#	Hits		Fichiers		kB F		kB In		kB Out		Location
1	318148	40,98%	258835	43,71%	19007957	30,83%	10515	2,77%	165291	0,38%	France
2	211056	27,19%	174549	29,48%	14648022	23,76%	26706	7,03%	141715	0,33%	Network (.net)
3	60071	7,74%	48109	8,12%	4212712	6,83%	3614	0,95%	102436	0,24%	US Commercial (.com)
4	35378	4,56%	27639	4,67%	3909190	6,34%	3413	0,90%	29062	0,07%	Maroc
5	25011	3,22%	20034	3,38%	2630248	4,27%	8326	2,19%	19850	0,05%	Tunisie
6	24960	3,22%	20230	3,42%	5746258	9,32%	9646	2,54%	39615	0,09%	Algerie
7	24439	3,15%	20306	3,43%	3028635	4,91%	6996	1,84%	26361	0,06%	Non-resolu/Inconnu
8	7400	0,95%	5921	1,00%	1310502	2,13%	449	0,12%	57744	0,13%	Cote D'Ivoire
9	6259	0,81%	4579	0,77%	202231	0,33%	661	0,17%	4214	0,01%	Etats Unis
10	5907	0,76%	5307	0,90%	409605	0,66%	151	0,04%	1061	0,00%	Chine
11	5722	0,74%	4326	0,73%	357617	0,58%	953	0,25%	1292	0,00%	Canada
12	5254	0,68%	4249	0,72%	306851	0,50%	608	0,16%	2752	0,01%	Belgique

Dans le journal de connexion du serveur HTTP, aucune mention d'un quelconque pays n'est pourtant visible. La septième ligne du « top 100 » laisse penser à l'administrateur que les pays sont trouvés par « résolution DNS inverse ». Des outils en ligne de commande comme « nslookup » permettent d'effectuer ce genre de résolution. Des outils en ligne le permettent également.

Travail à faire :

Effectuez la résolution DNS inverse des adresses IP 93.67.8.9 et 37.8.164.51. Quels sont les pays ainsi découverts ?	
La résolution DNS inverse de l'adresse IP 88.7.60.2 permet-	

elle de déterminer le pays d'origine de cette adresse ?	

Certains éditeurs vendent des bases de données de géolocalisation à partir des adresses IP. Vous pouvez tester cela par exemple à l'adresse http://www.maxmind.com/fr/home.

Travail à faire :

Vérifiez les pays des trois adresses IP testées précédemment.	

L'utilitaire de statistiques présente également ce tableau :

Top 20 sur un total de 6762 groupes de mots-clés			
#	Hits		Mots-clés
1	724	5,85%	certa
2	347	2,80%	reseau certa
3	166	1,34%	réseau certa
4	158	1,28%	certa dijon
5	112	0,90%	reseaucerta
6	90	0,73%	dépendance fonctionnelle
7	89	0,72%	bts sio
8	88	0,71%	pgi
9	63	0,51%	amortissement constant
10	63	0,51%	simulateur reseau
11	62	0,50%	pfeg
12	56	0,45%	cump
13	54	0,44%	budget flexible
14	49	0,40%	simulateur réseau
15	48	0,39%	référentiel bts sio
16	35	0,28%	referentiel bts sio
17	34	0,27%	amortissement in fine
18	34	0,27%	calcul de la van
19	33	0,27%	calculer un ratio
20	32	0,26%	pfeg ressources

Travail à faire :

Observez la ligne 2 de l'extrait de journal des connexions fourni en **annexe 1** et déterminez quel élément sur la ligne permet de remplir ce	

genre de tableau.	
Observez le même élément sur les autres lignes du journal des connexions et énoncez-en quelques usages possibles.	

http://www.ecommercemag.fr/E-commerce/Article/L-e-pub-a-l-ere-de-l-ultra-ciblage-32270-1.htm
http://fr.wikipedia.org/wiki/Ciblage_comportemental
http://fr.wikipedia.org/wiki/Trace_num%C3%A9rique

b) Exploitation des adresses de courrier électronique

Un courrier vient de vous parvenir, sur votre adresse de courrier (votrecompte@votrefai.fr), de la part de update@facebookmail.com.

Le contenu du message est le suivant :

> Bonjour,
>
> Vous avez un message personnel sur Facebook de la part d'un ami.
> Pour le lire, veuillez consulter la pièce jointe.
>
> L'équipe Facebook.

Le message est accompagné d'une pièce jointe nommée « Message Facebook.zip ».

Ce message est suspect à plus d'un titre…

Travail à faire :

Donnez les raisons permettant	

de penser que ce message est frauduleux.	

La plupart des logiciels de courrier électronique permettent de visualiser l'ensemble du courrier tel qu'il a été reçu, y compris avec ses parties « cachées », utiles essentiellement à l'acheminement et au classement du message. Certains logiciels appellent cela « voir le code source du message ».

L'**annexe 2** présente le code source du message en question. Vous analyserez ce code source pour confirmer votre suspicion.

Le serveur de courrier de votre fournisseur d'accès (mail.votrefai.fr) a enregistré l'adresse IP de la machine qui lui a confié le message et le nom sous lequel elle s'est présentée.

Travail à faire :

Précisez quel est le nom du champ d'entête correspondant.	
Quel est le résultat d'une résolution DNS inverse sur l'adresse IP enregistrée dans ce champ ?	
Quel est le résultat d'une résolution DNS sur le nom de la machine présent dans ce champ ?	
Y a-t-il correspondance ? Que pouvez-vous en déduire ?	

Une analyse du contenu du fichier « Message Facebook.zip » permet de confirmer les craintes. Un virus s'y trouve. Il est très peu probable que le véritable propriétaire de l'adresse update@facebookmail.com soit véritablement à l'origine de ce message.

Pour comprendre comme cela est possible, vous consultez le fonctionnement résumé d'un échange entre un client et un serveur de messagerie (**annexe 3**).

Travail à faire :

Indiquez à quel moment de l'échange entre le client et le serveur l'adresse de l'émetteur est transmise.	
Observez quel est le contrôle exercé par le serveur sur l'adresse de l'émetteur suite à cette transmission.	

Annexe 1 – Extrait d'un journal des connexions

Il s'agit de connexions établies avec le serveur HTTP (*web*) du réseau Certa.

La commande HTTP « GET » est utilisée par les navigateurs pour télécharger un objet (page HTML, fichier image, etc.).

N° ligne	Contenu du journal
1	90.54.101.154 - - [16/Dec/2012:19:43:31 +0100] "GET /rss/ HTTP/1.1" 200 12208 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:17.0) Gecko/17.0 Thunderbird/17.0" 410 12648
2	62.35.143.187 - - [16/Dec/2012:19:43:33 +0100] "GET /sio/ HTTP/1.1" 200 17077 "http://www.google.fr/url?sa=t&rct=j&q=reseau%20certa&source=web&cd=2&ved=0CDoQjBAwAQ&url=http%3A%2F%2Fwww.reseaucerta.org%2Fsio%2F&ei=TxbOUP2SHsTE0QWiqIGIAQ&usg=AFQjCNGqhbtjjcBTlYsLtUtbJfnjbybzeg&bvm=bv.1355325884,d.d2k" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/534.57.2 (KHTML, like Gecko) Version/5.1.7 Safari/534.57.2" 552 17372
3	62.35.143.187 - - [16/Dec/2012:19:43:34 +0100] "GET /Styles/certa.css HTTP/1.1" 200 6832 "http://www.reseaucerta.org/sio/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/534.57.2 (KHTML, like Gecko) Version/5.1.7 Safari/534.57.2" 331 7124
4	62.35.143.187 - - [16/Dec/2012:19:43:35 +0100] "GET /Data/logoCerta.gif HTTP/1.1" 200 3710 "http://www.reseaucerta.org/sio/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/534.57.2 (KHTML, like Gecko) Version/5.1.7 Safari/534.57.2" 543 4002
5	92.143.96.106 - - [16/Dec/2012:19:44:08 +0100] "GET /sujets/ HTTP/1.1" 200 14222 "http://www.google.fr/url?sa=t&rct=j&q=bts%20sio%20%20annales%20et%20corriger&source=web&cd=4&ved=0CEkQFjAD&url=http%3A%2F%2Fwww.reseaucerta.org%2Fsujets%2F&ei=_hXOUKDbMqfV0QWtnIHABA&usg=AFQjCNFavf81R0iHLw9BBgbphzSrIN30Zg&sig2=qO5OHhQWAi94ClKRi-L2UA&bvm=bv.1355325884,d.d2k" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_8) AppleWebKit/534.57.2 (KHTML, like Gecko) Version/5.1.7 Safari/534.57.2" 867 14477
6	62.35.143.187 - - [16/Dec/2012:19:44:11 +0100] "GET /sio/edm/ HTTP/1.1" 200 18390 "-" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/534.57.2 (KHTML, like Gecko) Version/5.1.7 Safari/534.57.2" 551 18685
7	92.143.96.106 - - [16/Dec/2012:19:44:12 +0100] "GET /sujets/ecodroit.php HTTP/1.1" 200 20263 "http://www.reseaucerta.org/sujets/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_8) AppleWebKit/534.57.2 (KHTML, like Gecko) Version/5.1.7 Safari/534.57.2" 640 20518
8	92.143.96.106 - - [16/Dec/2012:19:44:17 +0100] "GET /docs/sujets/ed11n.zip HTTP/1.1" 200 213268 "http://www.reseaucerta.org/sujets/ecodroit.php" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_8) AppleWebKit/534.57.2 (KHTML, like Gecko) Version/5.1.7 Safari/534.57.2" 654 213570
9	157.55.35.114 - - [16/Dec/2012:19:44:22 +0100] "GET /docs/cotecours/contexteSIO_M2L.pdf? HTTP/1.1" 200 473490 "-" "Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)" 309 36500
10	92.143.96.106 - - [16/Dec/2012:19:44:31 +0100] "GET /docs/sujets/ed11nCorr.zip HTTP/1.1" 200 194485 "http://www.reseaucerta.org/sujets/ecodroit.php" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_8) AppleWebKit/534.57.2 (KHTML, like Gecko) Version/5.1.7 Safari/534.57.2" 658 194787
11	62.35.143.187 - - [16/Dec/2012:19:44:34 +0100] "GET /btsig/index.htm HTTP/1.1" 200 15550 "http://www.reseaucerta.org/sio/maths/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/534.57.2 (KHTML, like Gecko) Version/5.1.7 Safari/534.57.2" 606 15845

Annexe 2 – Code source d'un courrier électronique que l'on croit être frauduleux

```
Return-Path: <update@facebookmail.com>
Delivered-To: votrecompte
Received: from facebookmail.com (unknown [41.230.217.24])
       by mail.votrefai.fr with ESMTP id B221E100495
       for <votrecompte@votrefai.fr>; Fri, 28 Dec 2012 17:16:39 +0100 (CET)
From: update@facebookmail.com
To: votrecompte@votrefai.fr
Subject: Vous avez un nouveau message sur Facebook !
Date: Fri, 28 Dec 2012 17:18:23 +0100
MIME-Version: 1.0
Content-Type: multipart/mixed;
       boundary="----=_NextPart_000_0005_525F3EBF.ED4225E1"
X-Priority: 3
X-MSMail-Priority: Normal

This is a multi-part message in MIME format.

------=_NextPart_000_0005_525F3EBF.ED4225E1
Content-Type: text/plain;
       charset="UTF-8"
Content-Transfer-Encoding: 8bit

Bonjour,

Vous avez un message personnel sur Facebook de la part d'un ami.
Pour le lire, veuillez consulter la pièce jointe.

L'équipe Facebook.


------=_NextPart_000_0005_525F3EBF.ED4225E1
Content-Type: application/octet-stream;
       name="Message Facebook.zip"
Content-Transfer-Encoding: base64
Content-Disposition: attachment;
       filename="Message Facebook.zip"

UEsDBBQAAgAIAICAnEHlyQZP0p0HAADGCQAMABEAZG9jdW1lbnQuZXhlVVQNAAcBAAAAAQAAAAEA
AADMfXtA1MX2+GcfwIKruyq+X1Rr+Q7ESkUThSV8rwK+tSzZ1HwFn/VRoEsLNz982rLSstdN0ro9
       [...] etc. [...]
wHklb2zkjfO+RHULHy1+xV8gyktfYvsB4739/FjK1zdLedvqV0pwlRXlVD7q8Z6K3/hUfKhSuiKB
JO+WvnEYxZpqsXrtaonfVq1YtUpS8vwKqfMfNZUTV7hwBbUQRedHA3ybFUjfsq/T/qGc5tL6ZeIf
av/h7zj+P1BLAQIXCxQAAgAIAICAnEHlyQZP0p0HAADGCQAMAAkAAAAAAAAAAAAgAAAAABkb2N1
bWVudC5leGVVVAUABwEAAABQSwUGAAAAAAEAAQBDAAAADZ4HAAAAAAAA

------=_NextPart_000_0005_525F3EBF.ED4225E1--
```

Exemple de champ d'entête :

Subject: Vous avez un nouveau message sur Facebook !

Le nom du champ est le mot « Subject ».
Le contenu du champ est la phrase « Vous avez un nouveau message sur Facebook ! »
Le nom et le contenu d'un champ sont séparés par un double-point et un espace.

Un champ peut être détaillé sur plusieurs lignes. Dans ce cas, les lignes supplémentaires sont décalées vers la droite (à l'aide du caractère de tabulation).

Chaque champ d'entête a une utilité différente. Dans notre exemple, le champ « Subject » sert à indiquer au destinataire quel est l'objet (ou le sujet) du message.

Annexe 3 – Fonctionnement résumé d'un échange entre un client et un serveur de messagerie (protocole SMTP)

La colonne C/S indique qui envoie le message : C est le client et S est le serveur.
Le client établit d'abord une connexion avec le serveur sur son port TCP numéro 25. Une fois la connexion établie, c'est le serveur qui « parle » en premier…

C/S	***Requête/Réponse***	***Commentaire***
S	220 *nom_serveur* ESMTP *commentaires*	Le serveur précise le protocole SMTP qu'il connaît (ici : ESMTP).
C	HELO *nom_client*	Ou EHLO pour passer en mode ESMTP (Extended SMTP).
S	250 *nom_*serveur	Le serveur se présente.
C	MAIL FROM:<*adresse_expéditeur*>	L'adresse doit respecter la norme (*nom_compte@nom_domaine*). Attention : pas d'espaces autour du double-point (même si certains serveurs SMTP ont une tolérance).
S	250 Ok	La réponse commence par « 2 » : « tout va bien ».
C	RCPT TO:<*adresse_destinataire*>	Même remarques que pour « MAIL FROM ».
S	250 Ok	
C	DATA	On annonce au serveur qu'on veut lui transférer le message lui-même.
S	354 End data with <CR><LF>.<CR><LF>	Les réponses qui commencent par « 3 » exigent une action complémentaire du client. Ici, il s'agit d'envoyer le message lui-même.
C	From: <*adresse_expéditeur*> To: <*adresse_destinataire*> Subject: *objet_du_message* *Texte du message sur une ou plusieurs lignes* .	Ceci est juste un exemple. Notez la ligne vide entre les entêtes de message et le corps du message. On finit le message par un point entouré de retours-chariot, ainsi que la réponse précédente du serveur le demandait.
S	250 Ok: queued as *numero_enregistrement*	
C	QUIT	
S	221 Bye	Le serveur ferme ensuite la connexion TCP.